

OPTIMIZING FEATURE REDUCTION AND SELECTION TECHNIQUES FOR SURFACE ELECTROMYOGRAPHY



Author

Affaf Khan

Registration Number

326899

Supervisor

Prof. Dr. Asim Waris

**DEPARTMENT OF BIOMEDICAL ENGINEERING & SCIENCES
SCHOOL OF MECHANICAL AND MANUFACTURING ENGINEERING
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY
ISLAMABAD**

JUNE 2023

Optimizing Feature Reduction and Selection Techniques for Surface Electromyography

Author

Affaf Khan

Registration Number

00000326899

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Biomedical Engineering

Thesis Supervisor

Prof. Dr. Asim Waris

Thesis Supervisor's Signature: _____



DEPARTMENT OF BIOMEDICAL ENGINEERING & SCIENCES
SCHOOL OF MECHANICAL AND MANUFACTURING ENGINEERING
NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

ISLAMABAD

JUNE 2023

THESIS ACCEPTANCE CERTIFICATE

Certified that the final copy of the MS thesis written by Ms. **Affaf Khan** (Registration No. **326899**), of SMME (School of Mechanical and Manufacturing Engineering) has been vetted by the undersigned, found complete in all respects as per NUST Statutes / Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for the award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the school have also been incorporated in the said thesis.

Signature: _____

Name of Supervisor: Dr. Asim Waris

Date: 09/06/23

Signature (HOD): _____

Date: 14-6-23

Signature (Dean/Principal): _____

Date: 14-6-2023

National University of Sciences & Technology
MASTER THESIS WORK

We hereby recommend that the dissertation be prepared under our supervision by: (Student Name & Registration No.) **Affaf Khan (326899)**

Titled: **Optimizing feature reduction and selection techniques for surface electromyography** be accepted in partial fulfillment of the requirements for the MS Biomedical Engineering degree award.

Examination Committee Members

1. Name: Dr. Syed Umer Gillani

Signature: 

2. Name: Dr. Adeeb Shehzad

Signature: 

Supervisor's Name: Dr. Asim Waris

Signature: 

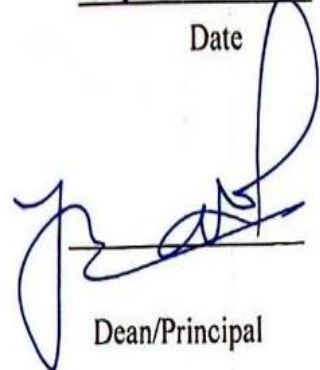
Date: 09-06-2023


Head of Department

14-6-23
Date

COUNTERSIGNED

Date: 14-6-2023


Dean/Principal

Declaration

I certify that this research work, “Optimizing feature reduction and selection techniques for surface electromyography” is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources has been properly acknowledged/referred to.



Affaf Khan

2020-NUST-MS-BMES-00000326899

Proposed Certificate for Plagiarism

It is certified that Ph.D./M.Phil. /MS Thesis Titled **Optimizing Feature Reduction and Selection Technique for Surface Electromyography** by **Affaf Khan (326899)** has been examined by us.

We undertake the following:

The thesis has significant new work/knowledge compared to already been published or is under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph, or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.

The work presented is original and the author's own work (i.e., there is no plagiarism). No ideas, processes, results, or words of others have been presented as Author's own work.

There is no fabrication of data or results which have been compiled/analyzed.

There is no falsification by manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.

The thesis has been checked using TURNITIN (copy of originality report attached) and found within limits as per the HEC plagiarism Policy and instructions issued from time to time.

Name & Signature of Supervisor

Dr. Asim Waris

Signature



Dr. Asim Waris
Assistant Professor
School of Mechanical &
Manufacturing Engineering
(SMME) NUST, Islamabad,

Plagiarism Certificate (Turnitin Report)

This thesis has been checked for plagiarism. The Turnitin report endorsed by the supervisor is attached.

Signature of Student

Affaf Khan

Registration Number

00000326899

A handwritten signature in blue ink, appearing to read 'S. M. Khan', is written over the registration number.

Signature of Supervisor

Copyright Statement

- Copyright in the text of this thesis rests with the student author. Copies (by any process) either in full or of extracts, may be made only per the author's instructions and lodged in the Library of NUST School of Mechanical & Manufacturing Engineering (SMME). Details may be obtained by the Librarian. This page must form part of any such copies made. Further copies (by any process) may not be made without the author's permission (in writing).
- The ownership of any intellectual property rights which may be described in this thesis is vested in the NUST School of Mechanical & Manufacturing Engineering, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the SMME, which will prescribe the terms and conditions of any such agreement.
- Further information on the conditions under which disclosures and exploitation may take place is available from the Library of NUST School of Mechanical & Manufacturing Engineering, Islamabad.

Dedication

I dedicate my thesis to my loving husband and
supportive family.

ACKNOWLEDGMENTS

All praises to ALLAH ALMIGHTY for blessing me with this opportunity and fulfilling my dreams.

I extend my gratitude to all my instructors throughout my MS degree who have taught and guided me; their hard work has brought me to the conclusion of my degree. I am immensely grateful to my research supervisor *Professor Dr. Asim Waris*, whose support and guidance helped me during this research project. His unending passion for research has provided innumerable students with a constructive environment to conduct their research studies. I would like to thank him for being the best advisor and mentor. It was an honor to work under his guidance.

I am thankful to my guidance and examination committee members: *Dr. Omer Gillani* and *Dr. Adeeb Shehzad* from SMME, NUST, for their intuitive comments and encouragement. I would also like to acknowledge SMME lab engineers and assistants for their technical support.

I am deeply indebted to my (Late) Father, *Dr. Tariq Nawaz Babar*, for his blessings that will always remain with me throughout my life. I owe my mother, *Dr. Nabila Tariq*, a debt of gratitude for all her assistance, encouragement, and motivation. Despite all the difficulties and family pressures, she has always believed in me and strived hard to make all my dreams a reality. My husband, *Engr Muhammad Ali Hassan*, and his love, support, and belief were the things that compelled me to conclude this chapter of my life. I owe them this and every other success in my life. I would also like to express my heartfelt gratitude to all my family members for their everlasting love and care. They are the best family one could ask for.

Furthermore, I would like to thank my classmates, *Faria Hassan, Adnan Habib, Uzma Shafiq, and Major Ajudarullah* for creating such a positive learning environment in the lab. Finally, special thanks to my friends *Mahnoor Naeem, Laiba Fatima, and Maham Nayab* for their care and moral support in bringing me out of every difficult situation during this project. I am also thankful to my seniors and juniors for their scientific discussion and guidance.

Thank you all.

Affaf Khan

Abstract

A biological signal called an EMG monitors the electric impulses produced in muscles at various points during their contraction, which are indicative of neuromuscular processes. Surface electromyography's (sEMG) extraction of features is a substantial procedure to obtain the valuable information that is obscure in the data and to exclude redundant components and intrusions. Feature vector selection is crucial for effective EMG signal classification. However, several redundant characteristics have been found in studies of the categorization of EMG signals when utilizing a particular feature set. This research has discussed the properties and characteristics of different frequency and time domain features. Many features of the time domain are unnecessary and redundant and can be categorized based on the information and mathematical qualities, according to the scatter plots of different features, arithmetical analysis, or different classifiers. Contrarily, the statistical characteristics of the EMG spectral density of power are used to determine all frequency domain properties. The EMG recognition system's performance at the class separability level is inappropriate. This research also suggests using inefficient features for classifications in applications involving the classification of EMG signals.

Key Words: *sEMG, feature extraction, optimizing and feature reduction technique, classifiers.*

Table of Contents

THESIS ACCEPTANCE CERTIFICATE.....	ii
FORM TH-4.....	iii
Declaration	iv
Proposed Certificate for Plagiarism.....	v
Plagiarism Certificate (Turnitin Report).....	vi
Copyright Statement.....	vii
ACKNOWLEDGMENTS	ix
Abstract	x
Table of Contents	xi
List of Figures	xiii
CHAPTER 1.....	1
1.1 Electromyography	2
1.2 Background, scope, and Motivation	5
1.3 Data Acquisition for sEMG Signal.....	7
1.3.1 Surface EMG	7
1.3.2 Experimental Protocol.....	8
CHAPTER 2.....	11
2.1 Dimensionality reduction	12
2.2 Feature reduction techniques.....	13
2.3 Linear Methods	13
2.3.1 Principal Component Analysis (PCA).....	14
2.3.2 Factor Analysis	17
2.3.3 Linear Discriminant Analysis	18
2.3.4 Truncated singular value decomposition method.....	19
2.4 Non-Linear Methods	20
2.4.1 Kernel PCA.....	20

2.4.2 Multidimensional scaling	21
2.4.3 Isometric mapping (ISOMAP)	22
CHAPTER 3.....	23
3.1 Setting and Pre-processing for sEMG features	24
3.1.2 Pre-processing.....	25
3.2 Feature Extraction.....	26
3.3 Feature Selection.....	30
3.4 PCA implementation.....	32
3.5 Implementation of machine learning algorithm	32
CHAPTER 4.....	34
4.1 Results.....	35
4.1.1 Scatter Plots	35
4.1.2 Cross validation means accuracy of machine learning algorithms	37
4.2 Discussion	43
4.3 Statistical Analysis	43
CHAPTER 5.....	45
5.1 Conclusion.....	46
References	47

List of Figures

Figure 1 EMG data acquisition	2
Figure 2 Types of EMG electrodes. a. Surface Electrode, b. Needles Electrodes, c. Indwelling Electrode	3
Figure 3 Positions of the electrodes is as follows: A. sEMG electrodes is placed on the muscles of the finger extensors (A.1 Uniformly placed electrodes; A.2 Reserve electrode); B. sEMG electrodes is sited on the muscles of the finger flexors (B.1 Uniformly positioned electrodes; B.2 Free electrode); and C. The sensors are situated on the arm (C.1 Evenly placed electrodes; C.2 Spare).	8
Figure 4 Basic finger movements (flexions and extensions).	9
Figure 5 Preprocessing of the sEMG signal.....	10
Figure 6 Dimensionality reduction	12
Figure 7 Dimensionality reduction techniques.....	13
Figure 8 Linear and Non-linear data distribution	14
Figure 9 Correlation among two variables. (A) positive correlation ($r = 1$), (B) no correlation ($r = 0$) and (C) negative correlation ($r = -1$)	16
Figure 10 Depiction of the principal components space. The original time series is shown in vector shape in the two-dimensional PC spaces (red dots), and the yellow arrow show the first waveform connected to the data point	17
Figure 11 Separation of two variables based on LDA.....	19
Figure 12 PCA (left) and kernel PCA (right).....	21
Figure 13 Visualization of MDS and NMDS with the original datasets	22
Figure 14 Flow of sEMG signal for PCA and classifier implementation.....	25
Figure 15 Difference between feature extraction and feature selection	32

List of tables

Table 1 Mathematical formulas of extracted features.	30
Table 2 ONE-WAY ANOVA (P-VALUES)	44

CHAPTER 1

INTRODUCTION

1.1 Electromyography

When muscles contract, electric signals are generated in the tissues that indicate neuromuscular activities. This biological signal is called electromyography. The central nervous system primarily controls the muscle's contraction and relaxation. EMG gauges a muscle's response, or the amount of electricity produced in response to a nerve stimulating it. The examination looks for anomalies in neuromuscular function. One or more tiny needles (often referred to as electrodes) are placed into the muscle at some time throughout the examination. On an oscilloscope, which is a screen that displays electrical power as waves, the electrical signal detected by an electrode is then shown visually. To hear the response, a signal amplifier is necessary. The electric movement of muscles during relaxation, a moderate contraction, and a powerful contraction is measured by EMG. Muscle tissue doesn't often generate electric signals while at rest. On the insertion of electrode, a short period of motion is visible on the oscilloscope and eventually it fades away.

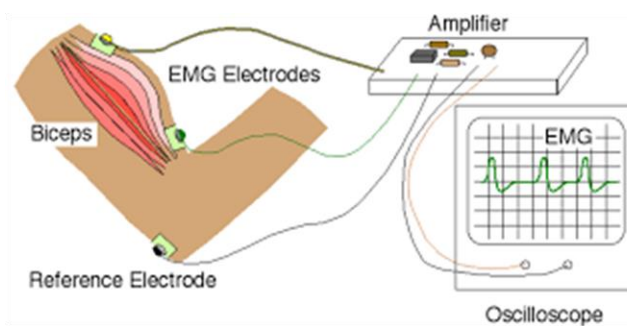


Figure 1 EMG data acquisition

As shown in above diagram, the EMG data is acquired via different mechanisms and prototypes. The process of acquiring data includes a subject, whose EMG signal must be examined, electrodes, amplifier, and an oscilloscope to visualize the signal. Now a days, following electrodes have been commonly used to extract EMG signal:

Surface Electrodes:

Non-invasive surface electrodes are positioned above the targeted muscles on the skin's surface. Surface electrodes can be further classified into three categories:

- **Disposable Adhesive Electrodes:** Such electrodes feature an adhesive backing and conductive adhesive gel to adhere to the skin. They are appropriate for one-time or brief term uses since they are simple to apply and take off.

- **Solid Gel Electrodes:** Electrodes with a firm gel coating that clings to the skin are known as solid gel electrodes. They offer good signal quality and are reusable. Before using them, the conductive gel must be applied.
- **Dry Electrodes:** Conducting gel or adhesive is not needed for these dry electrodes. They make electrical interactions with the skin using techniques such as polymers with conductive properties or specialized materials. Dry electrodes can take the shape of sensor-equipped patches or electrode arrays. They are practical to use because they do not require applying gel or removing the glue. In contrast to gel-based electrodes, they could have a larger impedance and a somewhat worsened signal quality.

Needle Electrodes:

Needle electrodes are tiny, needle-like electrodes that are implanted right into the muscle tissue. They are sometimes referred to as subcutaneous electrodes or fine-wire electrodes. They are utilized to record individual muscle fibers with greater accuracy and precision. Usually reusable, needle electrodes need to be sterilized before each use.

Indwelling Electrodes:

Specialized electrodes known as indwelling electrodes are inserted within muscle tissue. They may deliver continuous EMG signals and are employed for long-term monitoring. Indwelling electrodes are mainly utilized for research projects or certain medicinal applications and need surgical insertion.

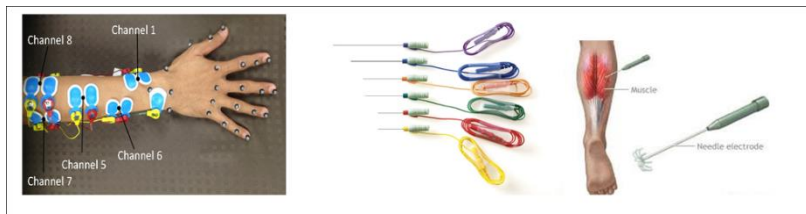


Figure 2 Types of EMG electrodes. a. Surface Electrode, b. Needles Electrodes, c. Indwelling Electrode

The selection of an EMG electrode is influenced by a number of parameters, including the reason for using it, the intended quality of the signal, ease of use for the patient, and the level of experience of the healthcare provider performing the EMG test.

The nervous system helps handle the complicated EMG signal, which depends on the muscles and tissues' morphological and physiological properties. EMG signal encounters noise as it goes through numerous tissues. Moreover, the EMG sensor groups together signals from multiple

motor units instantaneously, mainly when placed near the surface of the skin that could result in the interaction of various impulses. Depending on their contractile characteristics, structural makeup, and regulatory processes, the muscles of the skeleton, smooth muscles, and coronary muscles are the three distinct kinds of muscle tissues found in the human body. EMG is associated with the study of the muscles of skeleton that are found ascribed to the bones. The contraction of these muscles is under voluntary control and is initiated by neuron impulses to the muscles. Moreover, for contraction, muscular tissues are enriched with neurons. These neurons are known as "motor neurons". A single motor neuron can supply stimulation to various muscle fibers.

Numerous disciplines, including sports medicine, rehabilitation, ergonomics, biomechanics, and neurology, use sEMG in a variety of ways. In sports medicine, the sEMG can be applied to analyze patterns of muscle activation during various workouts or sporting events, revealing information about mobilization and coordination. sEMG helps track improvements during physical therapy and monitor the activity of muscles throughout rehabilitation. Additionally, sEMG is used by scientists and medical professionals to examine muscle diseases like muscular dystrophy as well as myopathy as well as to evaluate how well people with neurological problems like a stroke or injury to the spinal cord can use their muscles. It can support the diagnosis of neuromuscular illnesses and the tracking of disease development. Numerous methods are used in the evaluation of sEMG data, including timing measurements, assessments of muscle exhaustion, co-contraction analysis, and quantification of levels of muscle activity. These analyses can reveal important details about how muscles work, how well they execute, and any potential problems. It is important to note that sEMG differs from intramuscular electromyography (iEMG), which entails the insertion of needle electrodes into the muscle tissue. iEMG is an invasive treatment that is often only used for specialized clinical or research objectives, even though it produces recordings that are more exact. The neuromuscular system can be understood, and muscle function can be assessed using sEMG. It is frequently utilized in many applications relating to human mobility, rehabilitation, and medical diagnosis because of its non-invasiveness and adaptability.

The amplitude, frequency, signal structure, and morphology, as well as signal processing and analysis, are the properties and analysis of sEMG signals. An sEMG signal's amplitude represents the degree of muscle activity. It reflects the level of electrical activity produced by the motor units and is expressed in microvolts (V). Few examples of the variables that might affect the amplitude of the signal are force of the muscle, functioning motor units, and the distance between the electrode and muscle fibres. sEMG signals have a wide range of frequencies.

- **Low frequency (0–50 Hz):** this range represents the muscle's resting activity and includes inputs from slow-twitch muscle fibres and background noise. It is the lowest of the signal's three primary frequency bands.
- **Middle frequency (50-150 Hz):** This band, which is predominantly linked to fast-twitch muscle fibres, provides much of the pertinent data about muscle activation.
- **High frequency (above 150 Hz):** this range is frequently filtered out during data processing since it typically contains noise.

sEMG signal morphology and shape can reveal information about muscle activation patterns. The motor unit action potentials (MUAPs) comprise most of the signal and are a sequence of action potentials. Various factors, including the kind of muscle fibers engaged, the magnitude of the motor unit, and the intensity of muscular contraction can affect the time frame, amplitude, and form of MUAPs. Different signal processing and evaluation techniques are used to obtain helpful information from sEMG data. These methods include filtering to get rid of noise and artifacts, rectification to get the signal's true values, smoothing to get rid of oscillations, and extraction of features to characterize the signal's unique properties. RMS, MAV, and attributes like median frequency and spectral moments are frequently used features.

Biomedical engineering is increasingly requiring EMG signal identification with strong and improved techniques. Medical prognosis and biological programs are the main drivers of interest in EMG signal analysis. One of the possibilities is referred to as rehabilitation and management of motor disabilities. Analysis of neuromuscular issues is aided by the models and activation charge of (MUAPs) in electromyography signals. Nowadays, research and creative efforts are being done in this area, leading to the advancement of sophisticated algorithms, the upgrading of the existing methodologies, the improvement of detection techniques to reduce noise, and the acquisition of accurate EMG signals. Hardware execution have been finished for artificial hand control, grip recognition, and human-device interface. Research must be done to categorize the actual issues with evaluating EMG signals and to support the measures that are currently being used.

1.2 Background, scope, and Motivation

Traditional reviews of EMG signal data, especially in a medical environment, are largely informative and contemplated pinpointing the duration of EMG bursts as well as the times of onsets and peaks. They also aim to characterize the intensity of muscular initiation by specifying indices like the average of the EMG burst or the EMG peak amplitude. These investigates, while being relatively simple to perform and requiring little computational power, provide a brief explanation of the EMG firing pattern's potential complexity, particularly when analysing EMG waveforms from

several different muscles is required. Numerous quantitative analyses of the development of the EMG signals over the past ten years to address the shortcomings of the traditional ones, concurrent with the rapid increase in computational power made possible by personal computers [6]. The breakdown of the waveforms of EMG signal into a limited number of basic waveforms, which represent the significant aspects of the fundamental EMG starting patterns, is the common principle behind these analyses. The challenge of employing sEMG to control an advanced hand prosthesis (AHP) is most likely best stated in [10]. While a subject performs several types of movements of concern, surface EMG is recorded from the skin of the forearm.

Signals from surface electromyography (sEMG) are frequently employed in innumerable fields such as computer-human interaction, rehabilitation, and medical diagnosis. But the high-dimensional and chaotic nature of sEMG signals makes it difficult to classify and analyze them effectively. Through the reduction of dimensionality and the selection of pertinent features, feature extraction approaches are essential for optimizing the analysis of sEMG data. This thesis aims to investigate and enhance feature selection and reduction methods for sEMG signals, eventually improving the precision and effectiveness of signal processing approaches.

These studies are especially helpful when categorizing quantitatively EMG patterns obtained from different muscles or from similar muscles during various voluntary tasks. The low-dimensional space defined by the reduced number of basis waveforms obtained from these studies allows for the analysis of the dissemination of the earliest EMG initiation patterns. Either the frequency domain or the temporal domain can be used for waveform decomposition. The Fourier analysis illustrates instances of frequency domain waveform decomposition. [7], furthermore, by wavelet analysis [8, 9] in segments. Many techniques for decomposing EMG waveforms work in the temporal domain and are basically variations on factor analysis. [33]. These include non-negative matrix factorization (NNMF), truncated singular value decomposition (SVD), independent component analysis (ICA), and principal component analysis (PCA) [11–16]. The fact that PCA is supported by a wide range of available commercial statistical software packages or is relatively simple to program makes it stand out even though there is no appropriate factorization method, and each option should be carefully considered in light of the characteristics of the dataset to be studied. For these reasons, PCA, when compared to other techniques, has the best chance of being widely adopted for the analysis of large amounts of data of EMG signals produced from the recordings of multi-electrode in both examination and clinical surroundings. As a result, the purpose of this paper is to give a general overview of PCA used on surface EMG signals. It will first briefly go over the key methodological aspects prior to using a few instances from the movement control documents in

which PCA has been effectively applied to shed light on neural mechanisms that may oversee managing typical movements from our motor collection like arm pointing and gait [17].

Additionally, data science and AI professionals are now utilizing data science ideas to boost company ROI. The corporate landscape is evolving because of Datatobiz's data analytics services, including data mining, predictive analytics, and data visualization. It is possible to utilize the proper approaches and tools for EMG signal analysis, correctly comprehend the signal's aspects and characteristics, and maybe design hardware implementations for a variety of applications involving EMG signals.

1.3 Data Acquisition for sEMG Signal

1.3.1 Surface EMG

10 dynamic double differential OttoBock MyoBock 13E200 surface EMG electrodes³ are used to record muscle activity in surface EMG. These electrodes produce an amplified, bandpass-filtered, and rectified variant of the raw sEMG signal. The bandwidth of this signal is between 0 and 25 Hz, as opposed to the raw sEMG signal's range of 15 to 500 Hz. The electrodes include an amplification gauge that may be set anywhere between 2,000 and 100,000 times. The optimal configuration for the gauge of the amplifier is in the range or roughly 14,000 times, according to qualitative experiments [24]. These electrodes are frequently described as reliable as well as safe in literature. They are easier to get because they don't need to be cleaned or shaved, which makes the experimenter and subject more at ease. There have been several attempts to locate the appropriate forearm muscles in healthy subjects for electrode placement (see, for example, [23], [25]), based on anatomical considerations and fundamental ideas about the physiology of a muscle and the electrical impulse that it produces while it is actively contracted.

Despite this, primary studies on similar patterns for sEMG [26], [29], which have lately been deep-rooted in [27], [28], demonstrate that targeted electrode placement is not necessary because pattern recognition tools are able to reimburse for suboptimal positioning or benefit from muscle crosstalk. Eight regularly set apart electrodes are positioned behind the elbow at a set gap from the radio-humeral joint to take advantage of both techniques, and two additional electrodes are positioned in the extensor and flexor muscles (see Figure 1). When it comes to amputees, The position of the two specific electrodes must be verified in every situation to ensure that it makes sense.

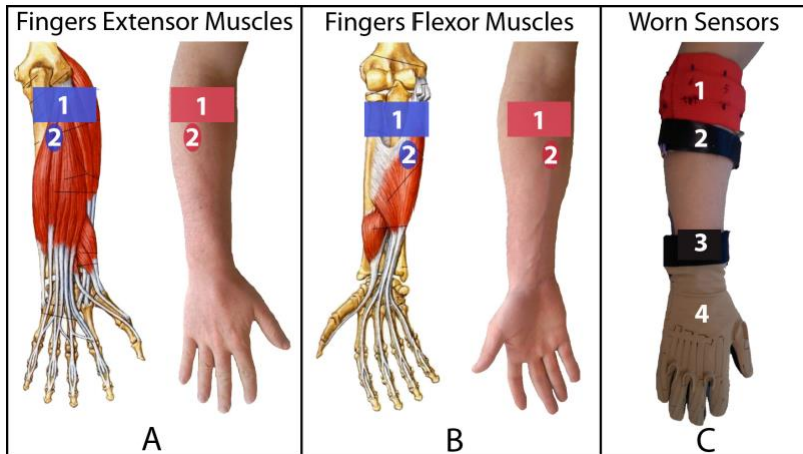


Figure 3 Positions of the electrodes is as follows: A. sEMG electrodes is placed on the muscles of the finger extensors (A.1 Uniformly placed electrodes; A.2 Reserve electrode); B. sEMG electrodes is sited on the muscles of the finger flexors (B.1 Uniformly positioned electrodes; B.2 Free electrode); and C. The sensors are situated on the arm (C.1 Evenly placed electrodes; C.2 Spare).

1.3.2 Experimental Protocol

Protocol that has been followed for the data acquisition is explained thoroughly in this segment. The Ethics Commission approved the experiment for the preliminary stages where most of the data was collected. Both the WMA (World Medical Association) and the safety of the related hardware are upheld. Before the investigation begins, each subject provided written informed consent, and their personal and clinical information as well i.e., age, gender, height, weight, fitness, dominance, and health status reported by them are all included in this study. Additionally, it is being considered to expand the clinical data record to include more details about the subjects' anatomical characteristics. Data are kept anonymously for a set period and are removable at the subject's request.

The debilities of the shoulder, hand, and arm process for practical movements [30] and lists of conventional hand posture and grasps, grasping arrangements and related research served as inspiration for the definition of a collection of wrist and hand actions of interest in the stimulus section. 12 fundamental finger flexions and extensions in all are acquired. These 12 movements are thumb adduction, thumb abduction, thumb flexion, thumb extension, index extension, index flexion, middle extension, middle flexion, ring extension, ring flexion, little finger extension, and little finger flexion. Figure 2 graphically shows each basic hand movement.

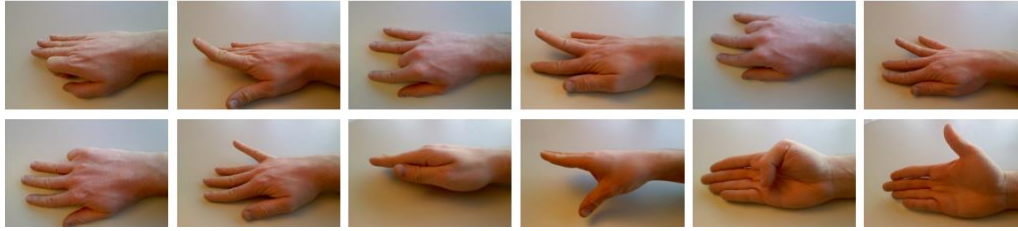


Figure 4 Basic finger movements (flexions and extensions).

The subject is positioned in front of the table with a big screen for data collection while sitting comfortably in an adjustable chair. The right hand is used to hold the inclinometer, dataglove, and sEMG electrodes. The individuals are requested to simply mimic the gestures shown in the short films as exactly as possible while they appear on the screen. The movement seen on the screen is asked of the subject bilaterally. (This process can be thought of as combining bilateral and imitation training [27]).

To familiarise themselves with the method, each subject first goes through a "training phase" in which they repeat three times each movement. Following the training phase, the subject is given a consecutive cycle of ten recurrences of every movement while data are being gathered. Five seconds are allotted for each movie, and three seconds can pass between each action. It is important to note that the sequences are not randomly generated on purpose to encourage repetitive, nearly unconscious movements. The exercise took 16 minutes to complete. Between the training process and the exercise, 5 minutes of rest are permitted to prevent muscle exhaustion and their impact on the sEMG signal. The experiment lasts roughly 50 to 60 minutes in total.

The electrodes of sEMG signal and inclinometer are attached to a typical data acquisition card selecting the signals at 100Hz; this allows the dataglove to stream data via a Bluetooth-channelled sequential port at a frequency of 25Hz for connectivity and data synchronization. A laptop running Windows XP and specially created computer program for this investigation is connected to the dataglove and the digital audio recording card. The data is entirely asynchronously stored with each sample being tagged with a precise timestamp to give the data the greatest flexibility in data processing. The HRT package [31], which is precisely the order of the microsecond, enforces accurate timestamping on Windows. The labels for every sample are then generated by the stimulus generator as a separate dataset operating at 25Hz, allowing every sample to be linked to a participant, exercise, and the indices of the movements.

To pre-process the data, the whole data is synchronized by interpolating linearly for all data up to the highest possible recording frequency (i.e., 100Hz). Then, by applying a 2nd order

butterworth filter with the “0” phase, the sEMG and Cyberglove signals are filtered at 1 Hz with lower cutoff frequencies. Divide every movement (including rest) into three segments of equal size, and only keep the data from the centre segment to exclude samples with confusing labels, or those collected throughout the period from the break and the real movement. The data in this centre section is then averaged to produce one sample per movement, ensuring computational viability. Observing the total ten samples (one for all repeat) for every combination of movement and subject, half samples are from the rest position. Figure 3 shows the preprocessing chain that was just described. Finally, the data for each participant is regularized so that every sEMG signal has a zero mean and a unit standard deviation.

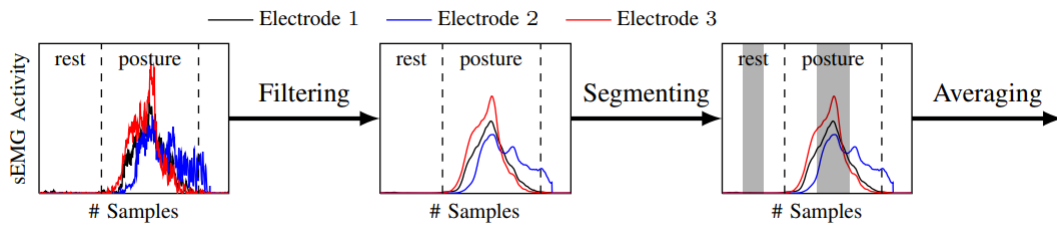


Figure 5 Preprocessing of the sEMG signal.

Following the preprocessing and feature extraction of the signals, the data is utilized to train the supervised method of machine learning. Later, the acquired connection joining the sEMG and the hand posture, their location, or effect is utilized to regulate the afterhyperpolarization (AHP). Even though sEMG is more closely connected to forces rather than locations, a substantial body of study has effectively allocated the plotting of sEMG to features of the hand, such as movement of the finger [33], [28], and acquisitive postures. Hand joints' low inertia, which enables the establishment of correspondences between isometric and isotonic hand muscle configuration and postures, is the key to success. (Effective attempts to map the sEMG to the entire arm position are being made).

CHAPTER 2

FEATURE REDUCTION

TECHNIQUES

2.1 Dimensionality reduction

Techniques for dimensionality reduction are essentially an element of the data's initial processing stage before training the model. It takes a lot of calculations and resources to use machine learning, in addition to the physical labor that goes by with it, analyzing the data with the help of variables. In such scenarios, dimensionality reduction methods are very helpful. Figure 4 illustrates the dimensionality reduction.

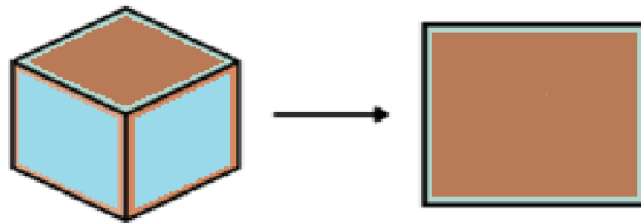


Figure 6 Dimensionality reduction

Using the dimensionality reduction technique, a high-dimensional dataset can be reduced in size without sacrificing any of the critical aspects of the original data. These dimensionality reduction techniques are essentially a type of data pre-processing that is carried out prior to model training. For instance, develop a model to predict the weather for the subsequent day using the existing climate conditions. Thousands of these aspects of the environment are too complex to analyze, including sunlight, moistness, cold, heat, and plenty more which could influence the existing conditions. Consequently, number of features can be reduced by identifying the features which has the higher degree of correlation and cluster them. Since we are aware of their strong correlation, we can combine moisture and rain into just one dependent feature in this instance. Using the dimensionality reduction technique, complex data can be condensed into a more manageable format without losing any of its original meaning. To learn about variations, trends, and patterns, techniques for machine learning and deep learning require massive volumes of data. Unfortunately, the dimensionality curse frequently arises since there are numerous features in such sizable data sets.

Large datasets usually contain sparse data as well. Features that are utilized in the process of training models but have little or no significance during testing perform poorly. We call this sparsity. Grouping similar features in the data is challenging due to these redundant features. Therefore, the dimensionality curse is fought via dimensionality reduction approaches. Subsequently, the performance of the models will be improved by eliminating the redundant data providing less room for assumption. It will

also reduce the computational cost, money and saves ample amount of time. Several algorithms for deep learning and machine learning do not function well on high-dimensional data, however, this issue can be fixed once the dimension is lowered. Since clustering of clean and non-sparse data is simpler and more reliable, these data will produce more statistically significant findings.

2.2 Feature reduction techniques

Algorithms for the dimensionality reduction are widely classified in two groups:

- Linear Methods
- Non-Linear Methods

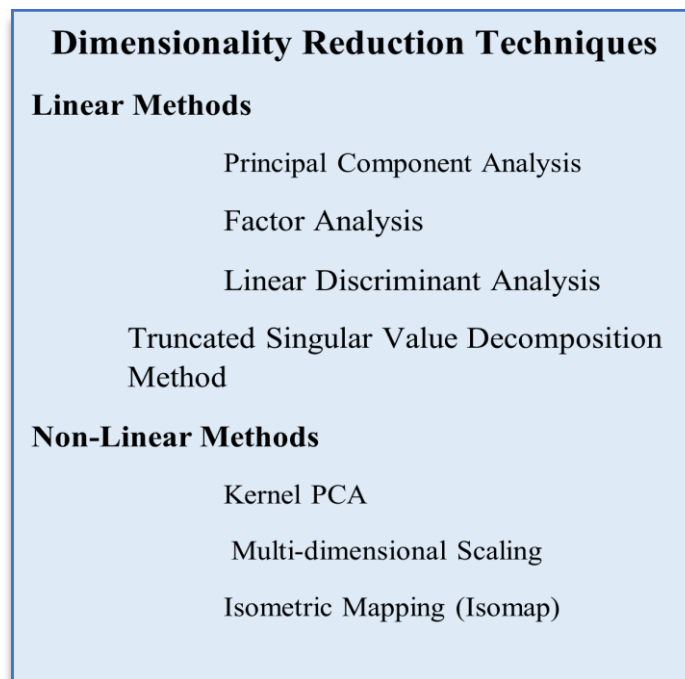


Figure 7 Dimensionality reduction techniques

2.3 Linear Methods

Given their straightforward symmetrical analyses and frequently appealing evaluation characteristics, linear dimensionality reduction techniques are a keystone of the assessment of data with high dimensions. Numerous valuable data features, including covariance, or dynamical structure, the correlation amongst sets of data, input-output connections, and the margin between data classes, are captured by these techniques. For more than a century, linear dimensionality reduction techniques have

been developed in machine learning, statistics, and applied sciences. These methods are now fundamental resources for analyzing noisy and data with high dimensions. Such techniques result in a linear low dimensional mapping of the actual high dimensional data while preserving some of the interesting features of the data. Consequently, linear reduction of dimensionality can be used to extract significant feature spaces, denoise or compress data, visualize, or explore the structure of the data, and so on. The field of linear dimension reduction approaches shows to be very complex [34] given the number of approaches used across a wide range of categories of data and fields.

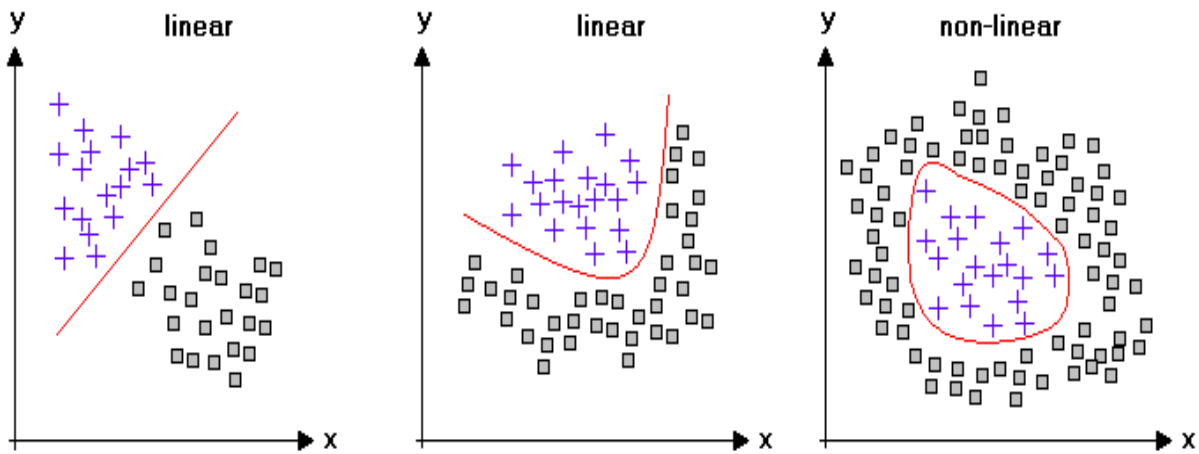


Figure 8 Linear and Non-linear data distribution

2.3.1 Principal Component Analysis (PCA)

PCA is one of the dimensionality reduction approaches used in data science. Consider a collection of 'p' variables that are correlated. This method breaks down the 'p' values in a fewer number of 'k' variables, which are often uncorrelated variables, such as ($k < p$). 'k' variables are these principal components that move in a way that is akin to the primary data set. To figure out which features are linked, PCA examines the relationship among the features that it combines. The produced dataset has less features that are directly linked with one another. In this manner, the model reduces associated features while also computing the greatest variance in the initial dataset. It locates the path of the obtained variance and then places them into the space with fewer dimensions, creating new components known as principal components.

These parts do a really good job of capturing the original features. As a result, it minimizes the reconstruction error whilst identifying the ideal components. By reducing the amount of data, machine learning algorithms can operate more quickly and effectively.

PCA explanation:

PCA forms the basis for multidimensional data analysis that employs projection methods. It is essential to define a multivariate data table as having a lesser number of variables to spot patterns, jumps, clusters, and outliers. This overview may reveal the relationships between data and its parameters in addition to those within the variables. Although PCA dates to Cauchy, Pearson is credited with developing it in figures. Pearson defined PCA as locating "lines and planes of closest fit to systems of points in space" [35]. PCA, an extremely universal tool, can analyze datasets containing a range of attributes, including integration, missing values, categorical information, and inaccurate measurements.

The goal is to find the most important information within the data and describe it as a group of summary indices called primary components. Principle component analysis, which is centered on statistical analysis, is a feature-extracting technique from a hefty collection of waveforms (particularly, the EMG signals) a selected group of waveforms known as the basic waveform or principal components which defines the prevalent elements presented in the data. In statistics, the PCs account for much of the variance in the dataset [10,18].

Covariance and Correlation:

Covariance as well as correlation are the foundation of PCA. It is possible to define the covariance, which measures the way that both variables (X, Y) change together, as follows:

$$\mathbf{cov}(X, Y) = H\{[X - H(X)][Y - H(Y)]\} \quad \mathbf{(i)}$$

where H represents the average. An evaluation of the direct relationship between the two variables, or the way one variable differs as a function of the other, correlation is linked to covariance (EQ. 2) [18]. It is characterized as:

$$\mathbf{cor}(X, Y) = \mathbf{cov}(X, Y) / [\mathbf{SD}(X)\mathbf{SD}(Y)] \quad \mathbf{(ii)}$$

where the standard deviation (SD) is, due to the SD elements at the denominator, it should be noted that the correlation despite the covariance, is only affected by the temporal relationship between the two variables and not by the amplitude of the oscillations in the two variables. The product-moment correlation coefficient of Pearson (r), which has a range of values between -1 and 1, measures the correlation between two variables.

When two variables are highly correlated (covariance), it means that their trends are comparable, whereas when they show a negative correlation (or covariance), their trends are distinct. The correlation coefficient

of Pearson is the same as 0 if there is no relationship between the two variables. Examples of pairs of variables that might be highly associated ($r = 1$), uncorrelated ($r = 0$), or oppositely correlated ($r = -1$) are shown in Figure 7.

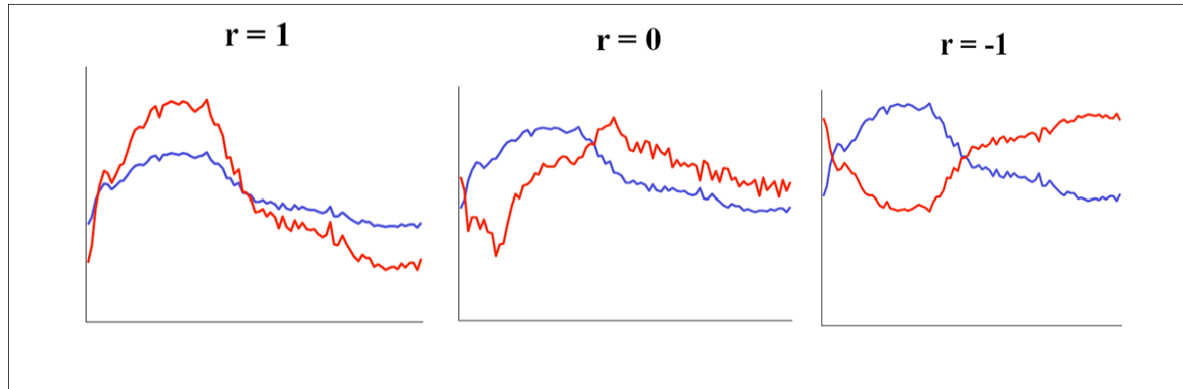


Figure 9 Correlation among two variables. (A) positive correlation ($r = 1$), (B) no correlation ($r = 0$) and (C) negative correlation ($r = -1$)

Computing the Principal Components:

Computing the covariance or matrix of correlation of the set of variables, the EMG signals is the first step in performing a PCA. According to the definitions for correlation and covariance, PCA retrieved from a matrix of covariance will likely be controlled using signals with higher scales. In contrast, the data obtained from a correlation matrix are merely impacted by the chronological relationships between the initial time series. A correlation matrix is the preferred option since, in practice, applying PCA to signals generated by EMG focuses primarily on finding shared temporal properties among EMG signals. The raw EMG signals need to be corrected, filtered at low frequencies with the cut-off frequency ranges between 10–20 Hz, and then normalized before a correlation (covariance) matrix can be computed (for instance, when step data is being processed, the EMG signals might be standardized to the portion of the step cycle). In summary, a correlation matrix is extracted via the PCA as follows:

- The orthogonal axes' orientations, or the eigenvectors, make up most of the variation in the dataset.
- The eigenvalues, or scalar element for eigenvectors, show the percentage of the overall variance that each eigenvector contributes to.
- Principal components, also known as score factors (PC), show the waveforms connected to each eigen value and eigen vector.
- The factor loadings (FL), or weighting coefficients, indicate the Pearson correlation coefficients between each unique EMG waveform and the principal components, allowing the primary EMG signal to be reconstructed by the biased average of PC.

$$\mathbf{EMG} = \sum_1^n \mathbf{pc} * \mathbf{fl} \quad \text{(iii)}$$

Where pc is the principal component and fl is factor loadings, EMG is the product of these two variables with limits ranging from 1 to n.

In general, depending on the quantity and complexity of the initial data set, the overall variance of up to 80-90% can be accounted for two to five PCs. The PCs retrieved using this method are ranked according to the proportion of variation explained. This results in a striking reduction in the dataset's dimensionality. In effect, by employing the weighted coefficients as vector coordinates, the PCs with low order provide a lower dimensional area over which the initial set of signals from the EMG could be exemplified.

EMG signals with comparable temporal components will typically cluster together in the PC space, whereas signals with distinct patterns of time will be spread out. One example of this PC space representation is shown in Figure 8. More than 80% of the data's variance of 25 original waveforms was accounted for by the first two PCs, which were then used to depict the original waveforms as vectors in 2D space bounded by these two PC waveforms. Take note of how severely weighted the first PC is on the original waveform seen in the inset.

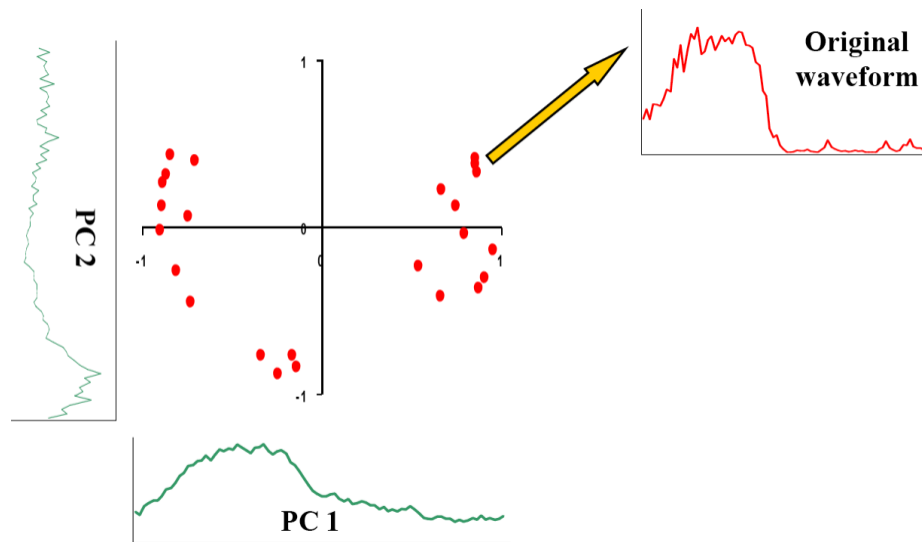


Figure 10 Depiction of the principal components space. The original time series is shown in vector shape in the two-dimensional PC spaces (red dots), and the yellow arrow show the first waveform connected to the data point.

2.3.2 Factor Analysis

This method is an elaboration of PCA. The main objective of this technique is not just to reduce the dataset. Finding hidden variables—which are the results of other dataset variables—is where it puts more

of its attention. They are not specifically measured by a single variable. Another name for latent variables is factors. Therefore, building a model that assesses these latent variables is referred to as factor analysis. It assists in both lowering the number of variables and identifying response clusters.

These scores can be utilized as indices of all variables for more analysis. Factor analysis, which is part of a general linear model (GLM), also assumes certain things, such as having a linear relationship, the lack of plurality, the inclusion of relevant variables in the evaluation, and the existence of actual correlations between variables and factors. To isolate the variable within the data set, many methods are used:

- **PCA** is the methodology that the scholars employ most often. Highest variance creates the first factor. The variation explained by the initial factor is then eliminated, and then the maximum variance for the following factor is subsequently obtained. This strategy is used to get at the last factor.
- **Common factor analysis** is the 2nd most preferred method by the scholars, and it extracts the average variance across all variables. Scanning Electron Microscope applies this technique.
- **Image factoring** is based on a correlation matrix. The factor that is used in image factoring is predicted using the OLS regression technique.
- **Maximum likelihood method** likewise operates on correlation matrices, but it factors using the maximum likelihood technique.

Least squares are surpassed by Alfa factoring and certain other FA techniques, another regression-based factoring technique is weight square.

2.3.3 Linear Discriminant Analysis

It is a technique for reducing dimensionality that is mainly used for supervised classification problems. Logistic regression fails in multi-classification. LDA enters the picture as a remedy for that flaw. It successfully separates training variables within all their classes. To improve the differentiation between multiple classes, it also varies from PCA as it determines the linear combination of the given input attributes. In the case of two variables, suppose they are randomly distributed on a 2D plane, making it impossible to divide them into two separate classes with a straight line. LDA is employed in these circumstances to maximize the difference between these two variables by converting a 2-D graph into a 1-D representation. The variables are projected onto a new axis as in Figure 9, which best divides them into their classes. Two procedures are used to create the new axis:

- By making the gaps between the means of the two groups as significant as possible.
- By reducing variation within each specific class.

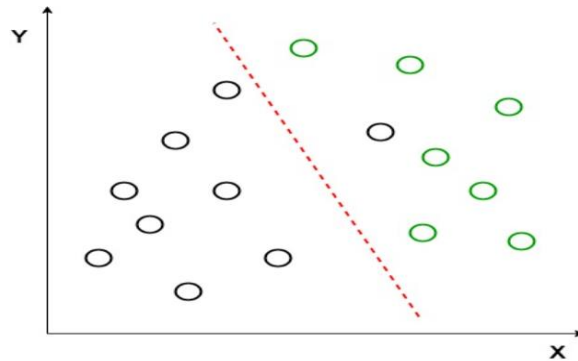


Figure 11 Separation of two variables based on LDA

LDA has two assumptions: that the matrices of covariance of the various classes are equal and that the information has a Gaussian distribution. Additionally, it presupposes that the data can be classified accurately by a linear boundary for decisions and that the information is linearly separable. It is a straightforward and computationally efficient approach that can perform well even when there are many more features than training examples.

It presupposes that the information follows a Gaussian distribution, which may not always be the case and can handle convergence (correlation among features) in the data. This assumption may not hold true in some datasets because it also presumes that the covariance matrices of the various classes are equal. It makes an inference that the data can be separated linearly, which might not be true for all datasets. In feature spaces with several dimensions, it might not function properly.

2.3.4 Truncated singular value decomposition method

The Truncated Singular Value Decomposition method (TSVD) is used to project a set of data having 'n' columns (features) onto a subspace with 'n' columns or fewer, without sacrificing the characteristics of the dataset. The SVD employs sparse data, whereas the PCA uses abundant data. Furthermore, PCA factorization uses the covariance matrix, whereas TSVD uses a data matrix.

Using mathematics to find the primary component, singular value decomposition, and truncated singular value decomposition. The truncated SVD differs from the conventional SVD in that it can be used more effectively with sparse data because it lacks a center for the data before computing the SVD.

2.4 Non-Linear Methods

With the aim of either visualizing the low dimensional data in the space or acknowledging the mapping (whether it's from the higher dimensional domain to the lower dimensional embedding and vice versa), nonlinear dimensionality reduction, also referred to as manifold learning, implies a variety of associated methods designed at projecting data with high dimensions onto lower-dimensional latent manifolds. Consider a dataset that appears like a matrix, where each row corresponds to a collection of characteristics (or features, or dimensions) that define a specific instance of something. When there are many qualities, the space of distinct possible rows grows exponentially. Thus, sampling the space gets more challenging the higher the dimensionality. This leads to numerous issues.

High-dimensional data processing algorithms frequently have very complex time schedules. For instance, numerous algorithms for machine learning struggle with large amounts of data. Data reduction can improve the efficiency of analytical algorithms and improve the prediction accuracy of machine learning algorithms. High-dimensional data is frequently challenging for humans to understand. As a result, data reduction to a condensed set of dimensions is advantageous for visualization. The term "intrinsic variables" is frequently used to describe the reduced-dimensional depictions of data. This description suggests that those are the values used to generate the data.

2.4.1 Kernel PCA

PCA is remarkably efficient for data that may be linearly segregated. the reduced dimension of the data set might not be accurate If applied to non-linear datasets. Consequently, this is the time when Kernel PCA is beneficial. The set of data is temporarily shifted to a more complexed feature dimensional space after undergoing a kernel function. The classes are changed in this case and can be identified and separated linearly with an even line. Additionally, a standard PCA is used, and data is reprojected into a space with fewer dimensions. It will be just as effective to use this LDR method there as it would be to use non-linear dimensionality reduction there. Kernel PCA implementation on two different variables is shown in Figure 10[36].

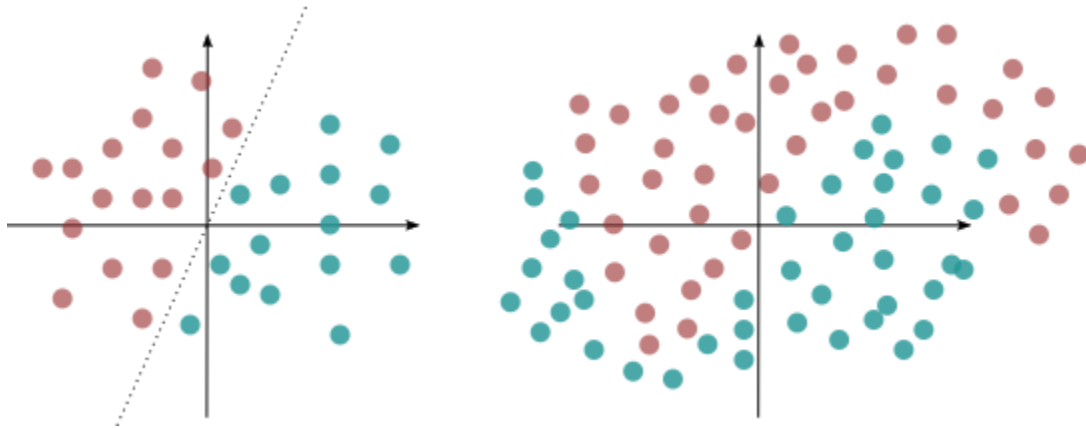


Figure 12 PCA (left) and kernel PCA (right)

The three key hyperparameters that govern the kernel PCA are the number of features to be kept, the type of kernel to be employed, and the coefficients of kernel. There are various varieties of kernels, including "linear," "poly," "rbf," "sigmoid," and "cosine." The Radial Basis Function kernel (RBF) is frequently utilized amidst these.

2.4.2 Multidimensional scaling

The process of simplifying data and reducing it to a lower dimension is described as scaling. A non-linear DR technique that graphically displays the discrepancies or gaps between the groups of characteristics. Longer distances are viewed as separate, while shorter distances are viewed as comparable.

MDS analyses the differences in the data by reducing the data dimension. Additionally, data does not lose its meaning when it is scaled down; regardless of its dimension, the two points of data remain relevant at the same distance. Only matrices with relational data, that is, correlations and distances, can be used with this technique. When MDS detects alterations in the data set, it generates a graph which determines the initial distance and displays their locations. Metric and non-metric MDS has been shown in Figure 11[37].

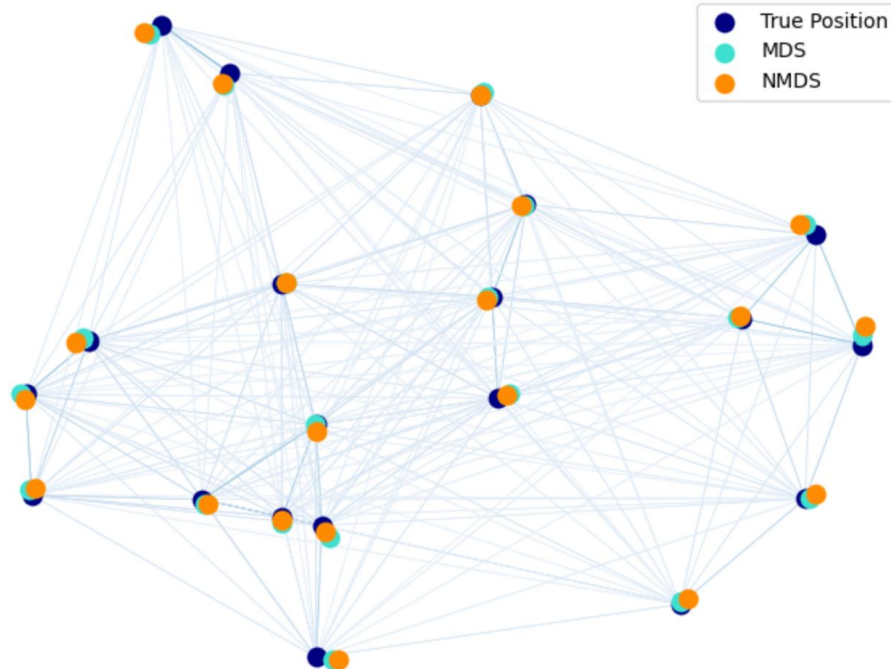


Figure 13 Visualization of MDS and NMDS with the original datasets

2.4.3 Isometric mapping (ISOMAP)

A non-linear DR technique is basically an extension of kernel PCA and MDS. It reduces dimensionality by connecting each feature depending on the curves or geometric distances between their nearest neighbors.

The basis of ISOMAP is the setting up of a neighborhood network. The geometrical distance between each pair of points is then calculated using graph distance. This set of data is finally entrenched in a smaller dimension by breaking down the geodesic matrix's eigen values. By using the `n_neighbours` hyperparameter of the `Isomap()` class, it is possible to specify how many neighbors should be taken into account for each data point. The ISOMAP algorithm is implemented by this class.

CHAPTER 3

METHODOLOGY

3.1 Setting and Pre-processing for sEMG features

Fig. 12 shows the whole process sequence of data gathering. Part A of Fig. 12 depicts the process of sEMG signals. Signals from the Otto Bock MyoBock system are processed, filtered, and quantized. We used signals from the fundamental hand movements, including thumb adduction, thumb abduction, thumb flexion, and thumb extension, index extension, index flexion, middle extension, middle flexion, ring extension, ring flexion, little finger extension, and little finger flexion.

10 dynamic double differential OttoBock MyoBock 13E200 surface EMG electrodes³ are used to record muscle activity in surface EMG. These electrodes produce an amplified, bandpass-filtered, and rectified variant of the raw sEMG signal, and are used to collect muscle activity. Frequency spectrum of the signal ranges between 0-25 Hz, as opposed to the raw sEMG signal's range of 15 to 500 Hz. The electrodes include an amplification gauge that may be set between 2,000 and 100,000 times. Qualitative tests showed that a secure enhancement gauge setting was in the center of the range, or roughly 14,000 times. Data contains the information of 7 intact subjects' 12 different movements having 10 repetitions. The subjects are required to replicate various motions portrayed in movies displayed on a laptop screen. Each subject's data for the exercise includes one MATLAB file having synchronized variables. The laptop's data acquisition software was Windows XP and custom software was acquired for this research. The variables incorporated in the MATLAB documentation are:

- Number of Subjects
- Exercises
- emg (electrodes' sEMG data)
- stimulus (Movement repetitive by the subject)
- re-stimulus (Subject's movement repetition)

For this variable, the length of the movement's label is modified a-posteriori to match the actual movement.

- repetition (repetition of the stimulus)
- re-repetition (repetition of the re-stimulus)

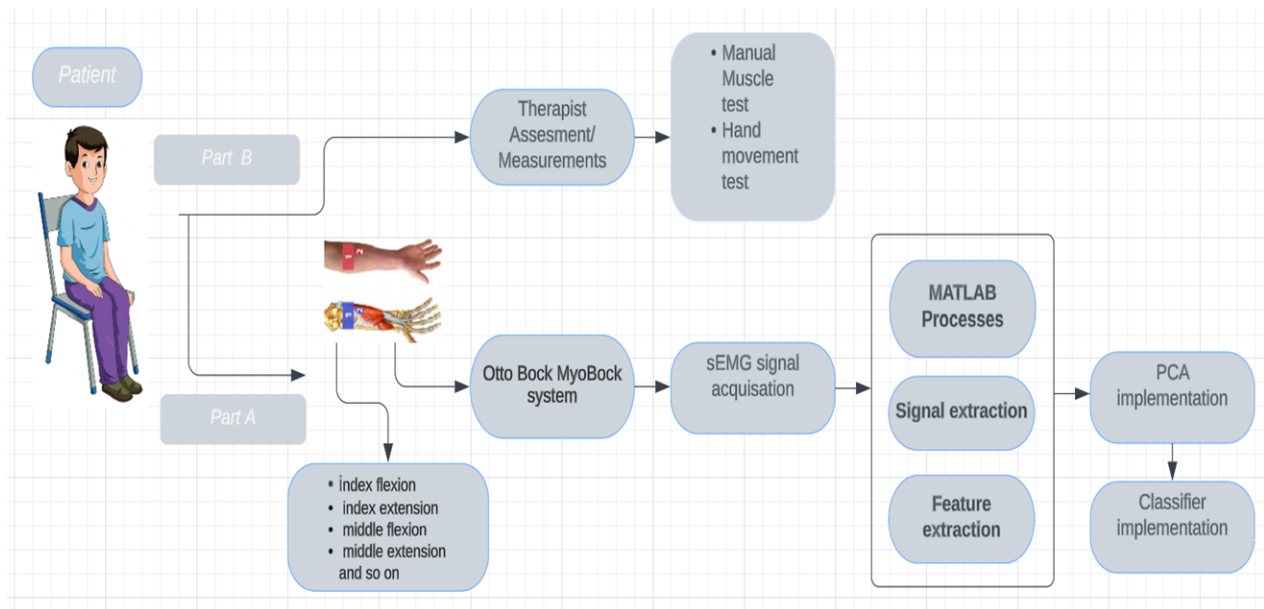


Figure 14 Flow of sEMG signal for PCA and classifier implementation

3.1.2 Pre-processing

Pre-processing of electromyography (EMG) data involves several steps to clean and prepare the signals for further analysis. The data is loaded in MATLAB and unique labels (0-13) representing the respective movements are assigned to it such as 0 is for the rest period and 1 to 13 are the movements. Data is further sorted with respect to the re-stimulus variable which is the repetition of the movements. It is further separated into 12 variables i.e., m1-m12 representing each movement which is processed at the 2483 samples as it provides better results when the total sample for respective movement is equal to the lowest sampling rate. Here are some common pre-processing techniques used in EMG data analysis:

- **Signal filtering:** Electrical interference, motion objects, and other types of noise frequently taint EMG readings. Techniques for filtering are used to reduce background noise and improve muscular activity. Commonly used filters are discussed below.
- **Low-pass filter:** It reduces high-frequency noise while keeping the EMG signal's lower-frequency components, which describe muscle activity, intact.
 - **Notch filter:** It eliminates power-line distortion as it impacts the data collectively by focusing on the precise frequency of 50 Hz. A notch filter is employed to enhance data quality.
 - **Smoothing:** To make the EMG signal easier to interpret and comprehend, smoothing techniques decrease high-frequency fluctuations. Moving average, median filtering or low pass filtration with a suitable cutoff frequency are examples of smoothing techniques.

- **Butterworth filter:** For the smoothing purpose, a 4th-order bandpass filter is employed to the filtered data to eliminate the effect of unwanted frequencies from the dataset i.e., spikes and dead beats. The sampling frequency of the bandpass filter is 2000 Hz with a 20 Hz higher cutoff frequency and 500 Hz lower cutoff frequency.
- **Rectification:** EMG signals are typically bipolar, with both positive and negative components. Rectification of the signal is achieved by taking the absolute entries of the data and converting it into a unipolar representation. This step ensures that the amplitude of muscle activity is accurately represented.
- **Normalization:** EMG signals can have varying amplitudes due to factors such as electrode placement or muscle size. Normalization aims to scale the EMG signal to a common range, typically by dividing the signal by its maximum value or by a reference value.
- **Segmentation:** EMG signals are frequently investigated in smaller time intervals called segments. Applying segmentation improves spectral leakage and grants a more accurate interpretation of the signal within each segment EMG data is mostly divided into segments after specific motor tasks or events of concern. Segmentation permits the assessment of EMG activity throughout identifiable periods, such as in muscle contraction. 200-500ms data is required for segmentation and 30-50% overlapping size is optimal according to the literature review. The window size of 450ms and the overlapping size of 225ms for the extracted EMG data. Overlap segmentation provided better classification results instead of disjoint segmentation. The segmentation was done for each subject with the motion of all channels.
- **Artifact Removal:** In some circumstances, non-physiological artifacts in EMG signals need to be eliminated. The movement of the electrodes or additional forms of interference may be the cause of these artifacts. These artifacts can be found and eliminated using various methods, including automatic thresholding, independent component analysis (ICA), and visual inspection.

The precise pre-processing procedures differ based on the EMG information or analytical goals, along with the peculiarities of the EMG signals being handled, therefore it's crucial to keep that in mind. To choose the best pre-processing methods for a particular application, it is advised to reference relevant literature or subject matter experts.

3.2 Feature Extraction

The analysis and interpretation of sEMG signals depend heavily on extracting sEMG features. This research can acquire significant knowledge about the operation of muscles, performance, and potential problems by extracting pertinent features from the raw sEMG data. The following are some major implications of sEMG features extraction:

- **measurement of muscle activation levels** can be measured by feature extraction. Metrics such as MAV or RMS provide details on the overall level of muscle activation for a particular task or movement. The effectiveness of therapies or rehabilitation programs can be assessed using these metrics, which can also be used to compare the degrees of muscle activation between different circumstances or people.
- **Muscle fatigue assessment** is the measurement of muscular exhaustion made possible by sEMG feature extraction. Utilizing characteristics like mean frequency or signal entropy, it is possible to quantify fatigue-related modifications to sEMG signals, such as amplitude reductions or spectral content shifts. To evaluate muscle function and choose the best training or breaks during workouts or rehabilitation, these features offer objective indicators of muscle exhaustion.
- **Muscle synchronization and coordination** is the examination of sEMG signal timing and synchronization aspects can reveal information about muscle coordination patterns. Co-contraction indexes cross-correlation analysis, or characteristics like burst duration and onset time can show how several muscle groups or individual muscles cooperate during a given movement or task. Optimizing movement efficiency, spotting muscle imbalances, and creating tailored therapies for rehabilitation or enhancement of performance can all be aided by an understanding of muscle coordination.
- **Detection of motor unit recruitment strategies** is the evaluation of motor unit enrollment patterns is made possible by sEMG feature extraction. The selection order and tactics of motor units of a muscle might be inferred by researchers by analyzing characteristics like intensity or frequency shifts in the sEMG signals. This knowledge is useful for understanding how muscles are activated during various tasks, spotting atypical recruiting patterns in neurological conditions, and assessing how therapies affect the recruitment of motor units.
- **Recognition of gestures or motion** are the system for gesture or motion detection that can use extraction of features from sEMG data. Algorithms based on machine learning can be trained to identify and recognize various movements or gestures by collecting patterns or properties in the sEMG signals, such as the time-domain or frequency-domain properties. This has uses in virtual reality, human-computer interaction, and prosthetics, among other areas.

To characterize muscle activity, important information must be recovered from the raw EMG signals during the essential feature extraction stage of electromyography (EMG) data processing. The following are some commonly used features extraction techniques for EMG data.

Time Domain Features:

- Mean absolute value (MAV): A time window's mean value of the absolute amplitudes.

- Root mean square (RMS): A time window's average squared amplitude is averaged as a square root.
- Zero crossing rate (ZCR): The total time an EMG signal intersects the zero axis in a specified time window is known as the ZCR.
- Variance: Signal variability within a time span is measured.

Frequency-domain features:

- The frequency where half of the EMG energy is above, and the other half is below known as the median frequency (MDF).
- Mean frequency (MNF): The average frequency of power spectrum.
- Peak frequency: the rate at which the spectrum's maximum power is present.
- Spectral moments: Frequency distribution statistics including the centre, skewness, and kurtosis.

Time-frequency features:

- Wavelet transform: This technique divides the EMG signal into time-dependent and frequency-based components. The consequent coefficients for wavelets or time-frequency depictions, like scalograms or spectrograms, can be used to extract features.
- Short-time Fourier transform (STFT): Similar to wavelet transforms, it analyses a signal in the time-frequency domain using fixed-size Fourier windows.

Higher-order statistics:

- Hjorth parameters: Statistical descriptors of the complexity, motility, and activity of the EMG signal.
- Bispectrum: Examines the EMG signal's phase relationship among various frequency components.

Time-domain shape features:

- Slope sign changes (SSC): The frequency of gradient variations in the EMG signal throughout the course of a time window.
- Willison amplitude: The occurrence of an amplitude surpasses a given limit throughout the course of a time window is known as the Willison amplitude (WAMP).
- Simple square integral: The squared sum of amplitudes inside a time window is known as a simple square integral (SSI).

Statistical features:

- Skewness: A measure of the inconsistency in the EMG signal distribution.
- Kurtosis: A measurement of the signal from an EMG distribution's peak or flatness.
- Auto-correlation: Evaluates how closely the EMG signal resembles a later version of itself.

TD features are used to give better classification results and require low computational complexity. The surface EMG receives a signal from each channel. MATLAB examines, investigates, and extracts features from the gathered data on the laptop. The 19 features that are extricated from dataset in MATLAB includes mean absolute value (MAV), waveform length (WL), zero crossing (ZC), root mean square (RMS), slope sign change (SSC), variance (VAR), mean absolute deviation (MAD), simple square integral (SSI), integrated EMG (IEMG), modified mean absolute value 2 (MMAV2), Wilson amplitude (WA), skewness (SKEW), mean frequency (MF), median frequency (MDF), total power (TP), and mean power (MP). These features were extracted from each segment before being given to feature reduction technique. The following table shows the mathematical formulas of extracted features.

Names of the features	Mathematical formulas
Mean absolute value	$MAV = \frac{1}{M} \sum_{p=1}^M y_p $
Waveform length	$\sum_{p=1}^M \Delta y_p $
Zero crossing	$ZC = \sum_{j=1}^{M-2} u[(x_{j+1} - x_j)(x_{j+1} - x_{j+2})]$
Root mean square	$R.M.S = \left\{ \frac{1}{m} \sum_{p=1}^m (y_p - \bar{y})^2 \right\}^{\frac{1}{2}}$
Cardinality	$Step\ 1: z_n = sort(x_n), n = 1: N\ Step\ 2: CARD = \sum_{n=1}^{N-1} z_n - z_{n+1} > s$
Slope sign change	$\sum_{m=2}^{M-1} (y_p - y_{p-1}) \cdot (y_p - y_{p+1}) > s$
Variance	$\frac{1}{M-1} \sum_{m=1}^M (y_m - \mu)^2$

	$m=1$
Mean absolute deviation	$\frac{1}{n} \sum_{i=0}^n x_i - m(X) $
Simple square integral	$SSI = \sum_{i=1}^N x_i ^2$
IEMG	$IEMG = \sum_{i=0}^N x_i [1]$
Modified mean absolute value	$MMAV = 1N \sum w_i x_i $
Willison amplitude	$\sum_{m=1}^{N-1} [f(y_m - y_{m+1})];$ $f(y) = \begin{cases} 1, & \text{if } y \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases}$
Skewness	$\frac{1}{m(SD)^3} \sum_{p=1}^m (y_p - \bar{y})^3$
Mean frequency	$MF = \frac{\sum_{i=1}^M f_1 PSD_1}{\sum_{i=1}^M 1 PSD_1}$
Median frequency	$\sum_{k=1}^{MDF} Q_l = \sum_{l=MDF}^M Q_l = \frac{1}{2} \sum_{k=1}^M Q_l$
Total power	$\lim_{s \rightarrow \infty} \frac{1}{2s} \int_{-s}^s y(s) ^2$
Mean power	$MP = \sqrt{\frac{1}{T_2 - T_1} \int_{T_1}^{T_2} [f(t)]^2 dt}$

Table 1 Mathematical formulas of extracted features.

3.3 Feature Selection

The type of features retrieved and the decision regarding the machine learning method are the two most important and defining elements in establishing how to assess the condition of the hand.

Because different traits are relevant to different diseases and deformities, discovering pertinent features is a random process. The properties of sEMG are categorized as time, frequency, or combined time-frequency. Selecting traits within the time, frequency, and time-frequency domains is still challenging [38–40]. Frequency domain characteristics of sEMG signals are less significant than time domain characteristics. However, multiple studies [41–43] have shown that better results have been provided by time domain rather than the frequency domain for sEMG signal identification regarding speed and accuracy.

In-depth studies of time domain features have occasionally been carried out. The waveform length, slope sign change, mean absolute value (MAV), and zero crossing (ZC) parameters that Hudgins et al. [40] proposed in many papers are now frequently recognized as essential for any EMG-based identification analysis. Cardinality is also a potential characteristic for pattern detection [44]. Elamvazuthi et al.'s [45] classification of people into healthy, unhealthy, neuropathy, and myopathy groups using root mean square (RMS) data had an accuracy range of 77.5-83.5 percent. Any EMG-based identification analysis, minimum, maximum, median, standard deviation, and signal-to-noise ratio is now generally viewed as requiring zero crossing (ZC), mean absolute value (MAV), waveform length, and slope sign change respectively. Feature selection, a pre-processing method, identifies the main characteristics of a particular situation. It has historically been used to solve a variety of issues, such as analyzing biological data and other systems for rehabilitation.

In medical applications, choosing features has been utilized successfully to both reduce dimensionality and better understand the root causes of disease. One method for reducing dimensionality is feature selection; in this method, pertinent characteristics are chosen while redundant and unimportant features are eliminated. Approaches for feature selection span a wide range of active research areas, including medical imaging, image classification, facial recognition, text classification, biomedical signal processing, and analysis of DNA microarray data. This research on several topics that use real patient data demonstrates the applicability of feature selection approaches in medical problems and shows how these methods perform in practical situations.

Medicine is one of the most fruitful and developing fields for feature selection and machine learning applications, where the aim isn't merely to mitigate the dimensionality of the problems but also to minimize the costs involved; for example, information can be extracted from images or the reasons behind discrepancies in disease diagnosis among image-analysis experts can be understood. Average amplitude power [46] is another tool frequently used to assess muscle state. The remaining characteristics, including the minimum, maximum, median, standard deviation, and SNR ratio, are quite common. The place of the hand and the location of the electrode channel are two more attributes.

An illustrative example of difference between extraction and selection of features is represented in Figure 13.

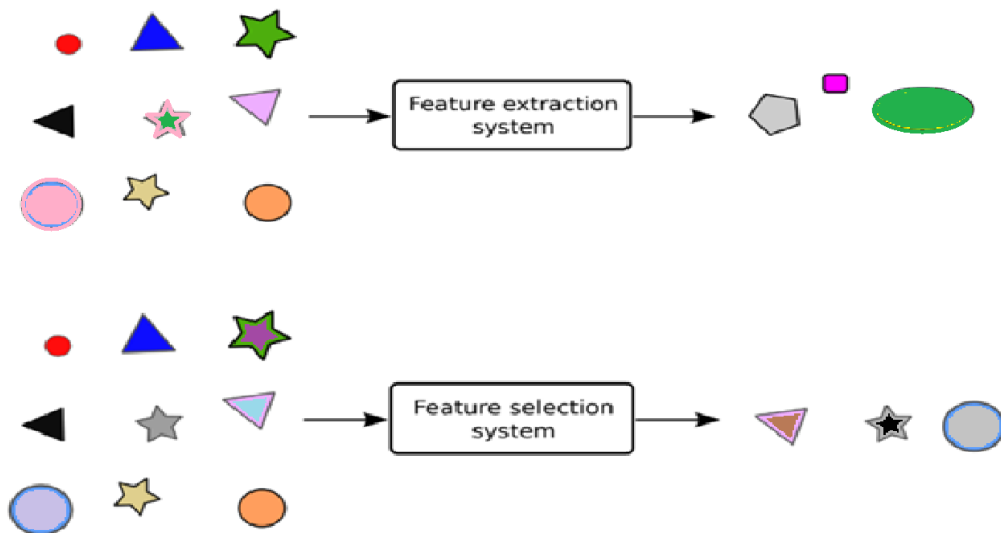


Figure 15 Difference between feature extraction and feature selection

3.4 PCA implementation

After following the entire procedure of principal component analysis described earlier in section 2.3.1, PCs are produced. PCs are further calculated on the given value of variance, for our dataset, the variance is 95%. Variance expresses the contribution of each PC in the dimensionality reduction of data, after the final calculation, it shows the PCs that give more than 95% variance. The reduced dimensional data is stored in the variable MappedX in MATLAB. Machine learning algorithms are implemented on the reduced data for the classification accuracy and to analyze the selected features.

3.5 Implementation of machine learning algorithm

One of the most active areas in machine learning is supervised learning. A predictive model must be trained using a collection of data that includes the desired outputs for it to forecast the outcome for data that have not yet been observed. This issue can be classified as a regression or classification challenge, based on the result type (discrete or continuous). The information algorithms utilized for supervised learning are usually represented as a matrix of samples and features (rows and columns). A suitable number of samples is often required by machine learning algorithms because too few samples might result in overfitting and poor generalization. However, they lack as many features as probable due to the dimensionality curse, which is the issue of analyzing high-dimensional datasets do not present in low-dimensionality settings.

The dataset must be shrunk to escape the scourge of dimensionality, which is connected to the optimization difficulties achieved by extensive enumeration in product spaces. This is accomplished by locating a matrix comparable to the original matrix but with fewer columns. This matrix can be deployed more effectively than the initial one because it has fewer features. Discovering matrices with smaller columns is referred to as dimensionality reduction.

Increasing generalization capability and classification accuracy or decreasing learning time and model complexity are two ways a drop in input dimensionality might enhance performance. The choice of appropriate characteristics can lower measurement costs and enhance problem comprehension. The influence of feature selection can be significant in some circumstances; for instance, in the analysis of microarray data, only 2 of the 7,129 features can be used to enhance classification performance. After feature selection, categorization is performed using MATLAB's classification learner app software. Different classification approaches are used in this study since it is supervised learning based on the efficiency and accuracy of these algorithms on the data, classification techniques like Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Quadratic Support Vector Machine (SVM), Cubic Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Ensemble Subspace KNN, Decision Tree (DT) [49, 50], Kernel SVM, Logistic Regression, Kernel Naive Bayes and Artificial Neural Network (ANN) based are all used in the training of features to be used.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Results

For each subject, six classifiers were trained and tested using a 5-fold cross validation scheme. The details of each feature are provided in table 1. MATLAB 2022 was used to design and analyze the proposed system. For each classifier, the CA was recorded, averaging 5-folds. The results were further validated by conducting statistical analysis using one-way analysis of variance (ANOVA) test. A probability value of less than 5% was considered significant.

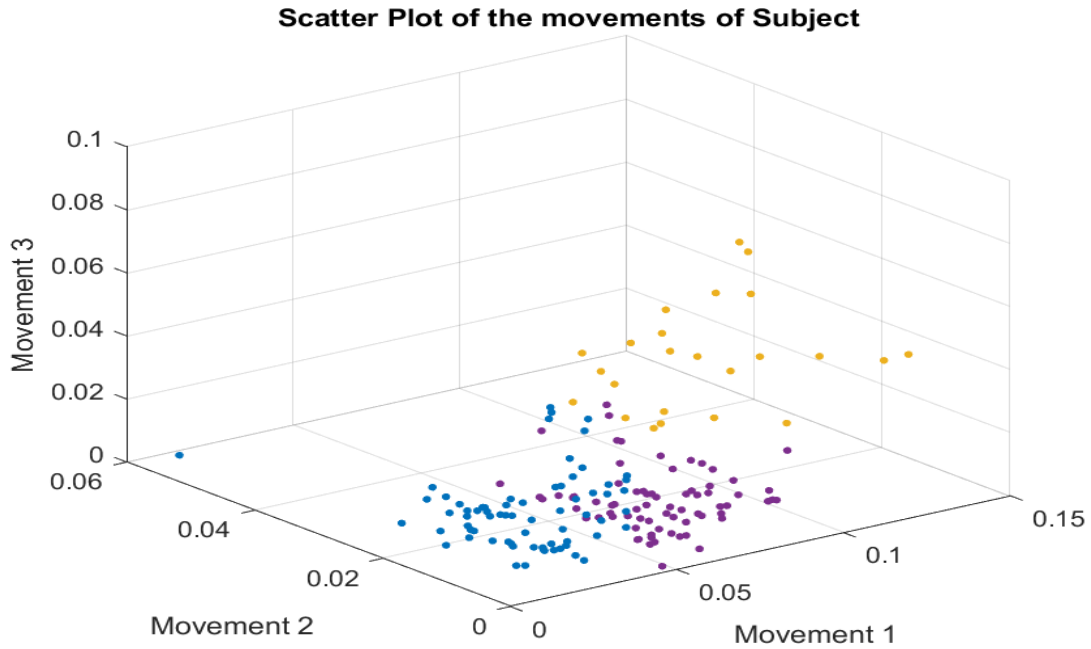
Two experimental components comprise our study. Firstly, the scatter plots of the subjects after the feature reduction and the second stage of the experiment, machine learning algorithms are evaluated to determine the cross-validation mean accuracy of each subject and for all subjects together. These are the components that are important for the results. Scatter plot exhibiting the distinctive feature's placement in the subspace are discussed in the study below.

4.1.1 Scatter Plots

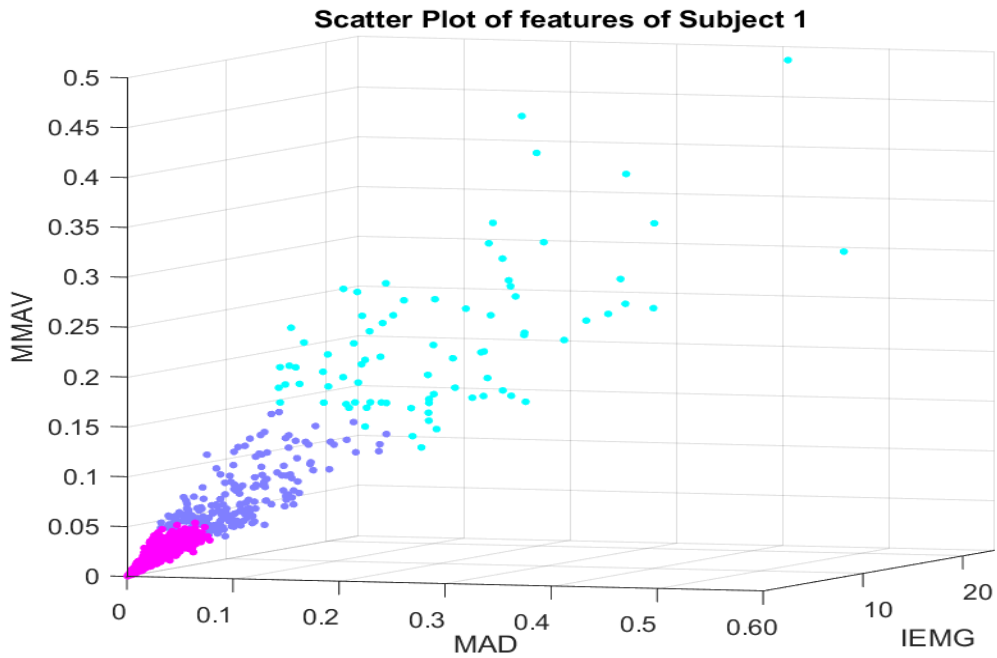
Scatter plots analyzes the relationship between two variables based on the overall pattern of the data points. It can identify the correlation among the variables whether positive, negative, or unrelated to each other. Scatter plots also identify outliers and can make predictions because of pattern of the dataset.

Scatter plot of Subject's Movements:

Displaying the extracted feature's first three movements i.e., index flexion, index extension and middle flexion, respectively.

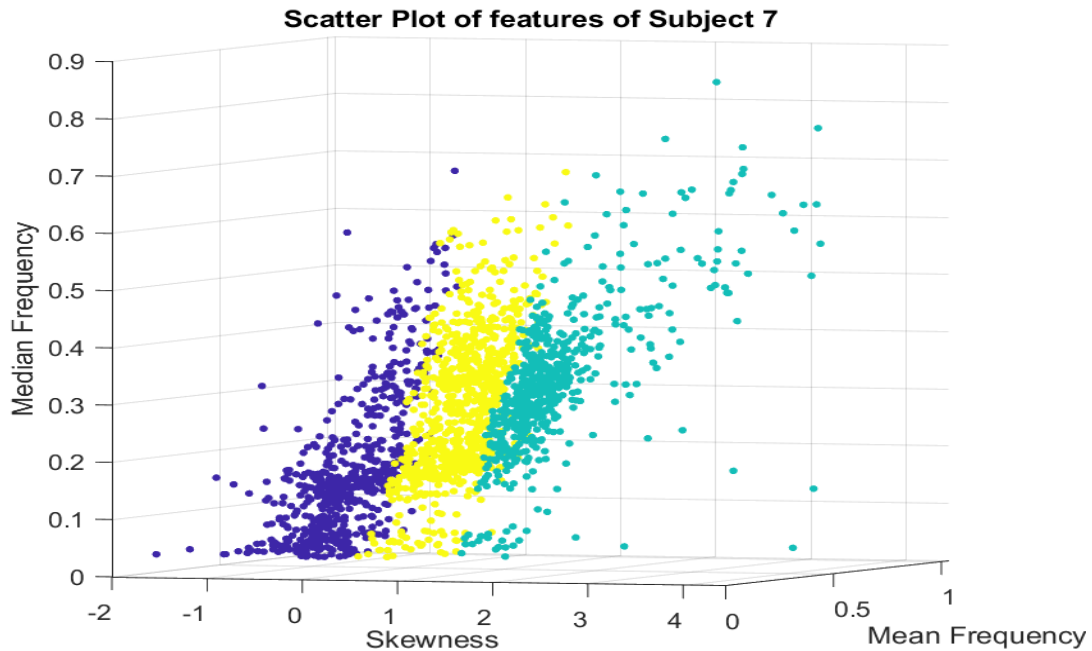


Scatter plot of Subject 1:



For this purpose, 3 different features i.e., Modified Mean Absolute Value (MMAV), Mean Absolute Deviation (MAD), and Integrated EMG (IEMG) of subject 1 have been extracted and displayed.

Scatter plot of Subject 7:

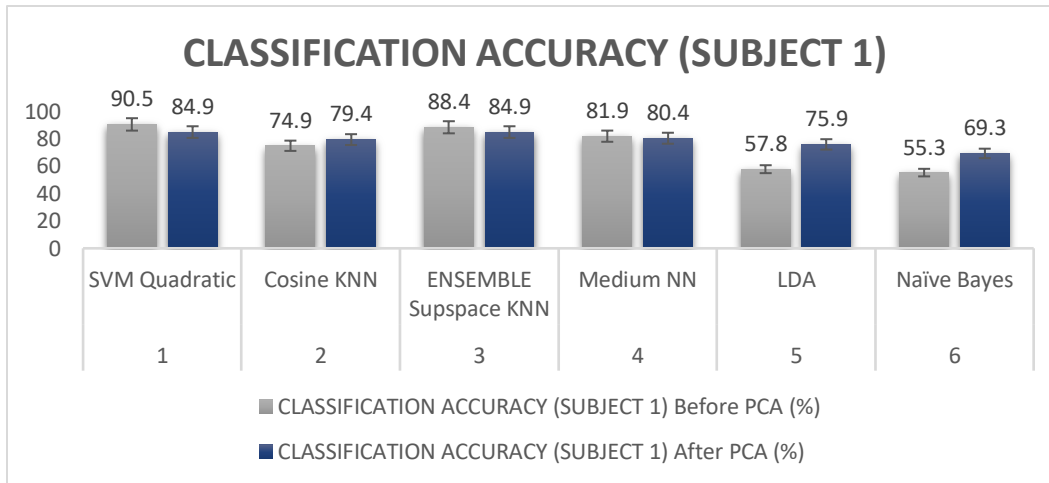


Three unique features i.e., Median Frequency (MDF), Mean Frequency (MF), and Skewness (SKEW) of subject seven have been extracted and displayed.

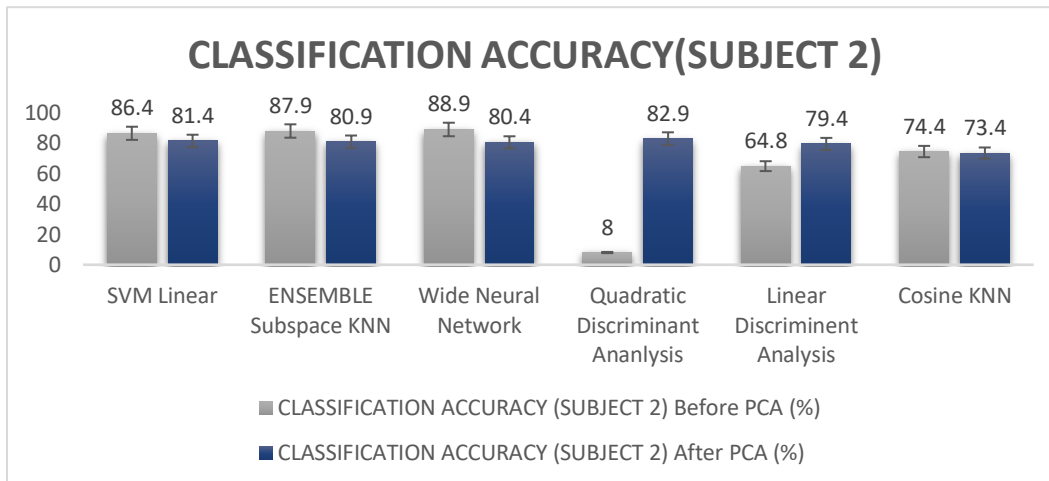
4.1.2 Cross validation means accuracy of machine learning algorithms

Cross validation means accuracy evaluates the performance of the machine learning algorithms and accesses how well a model simplifies unseen data by calculating its accuracy on various subsets of the given dataset. It also provides a robust estimate of the model's execution associated with evaluating it on individual train-test split. Applying various ML algorithms on each subject and all subjects together before and after the implementation of Principal Component Analysis (PCA) and calculating their accuracies. Graphical representation of each subject and all subjects have been studied well.

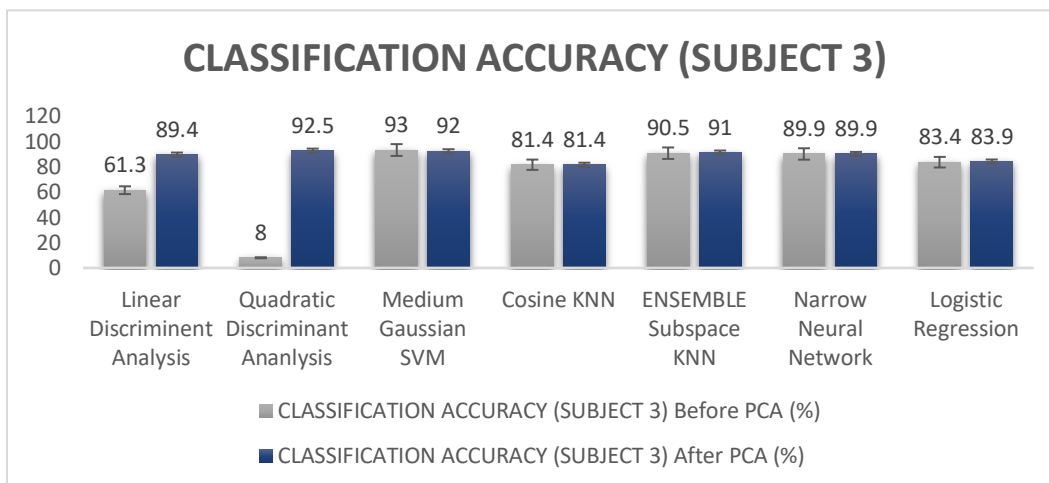
Subject 1:



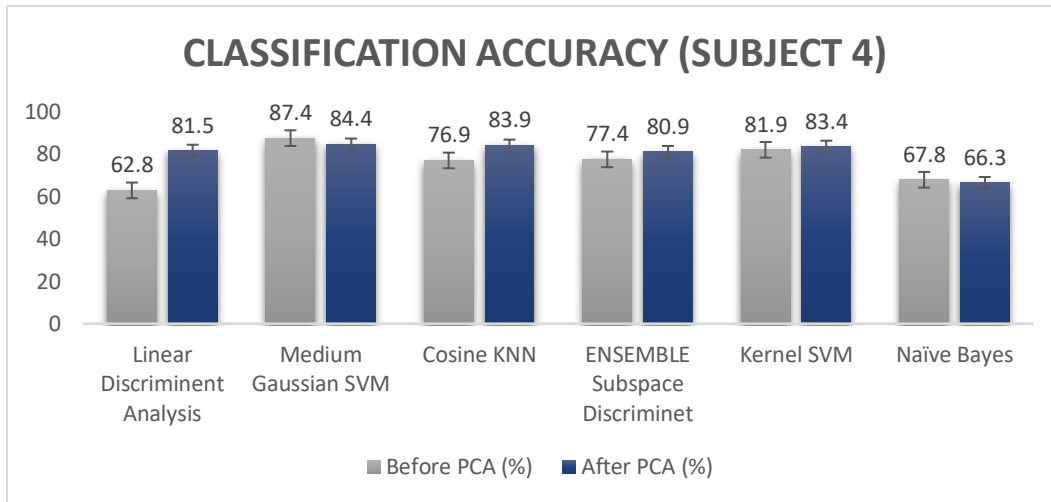
Subject 2:



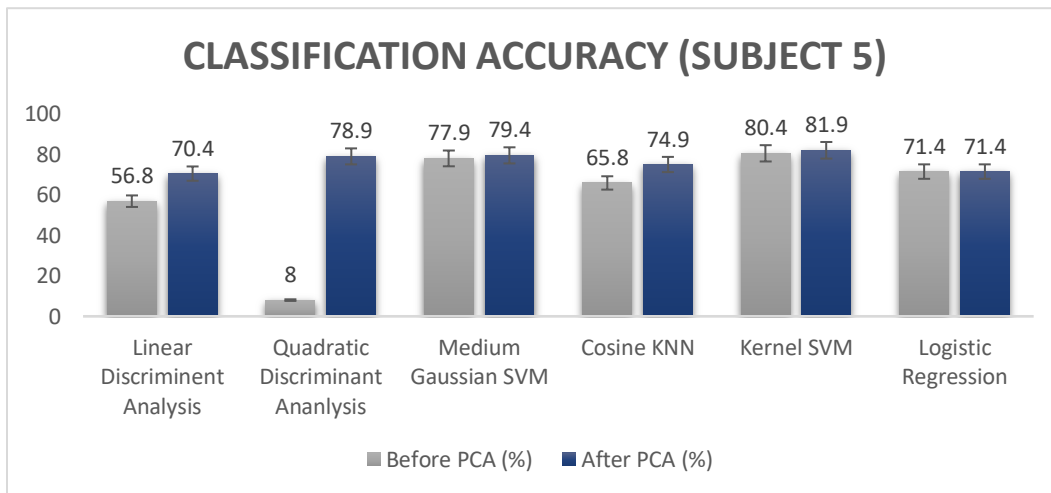
Subject 3:



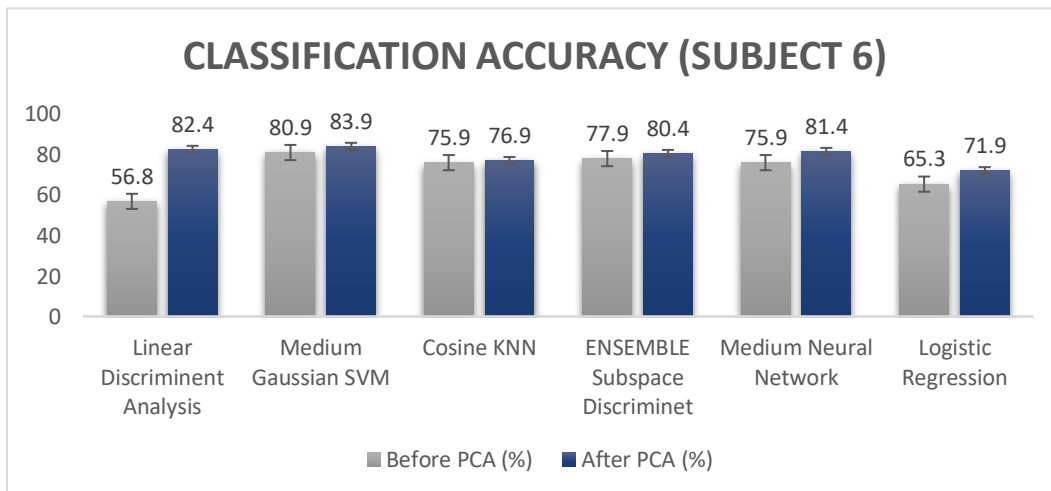
Subject 4:



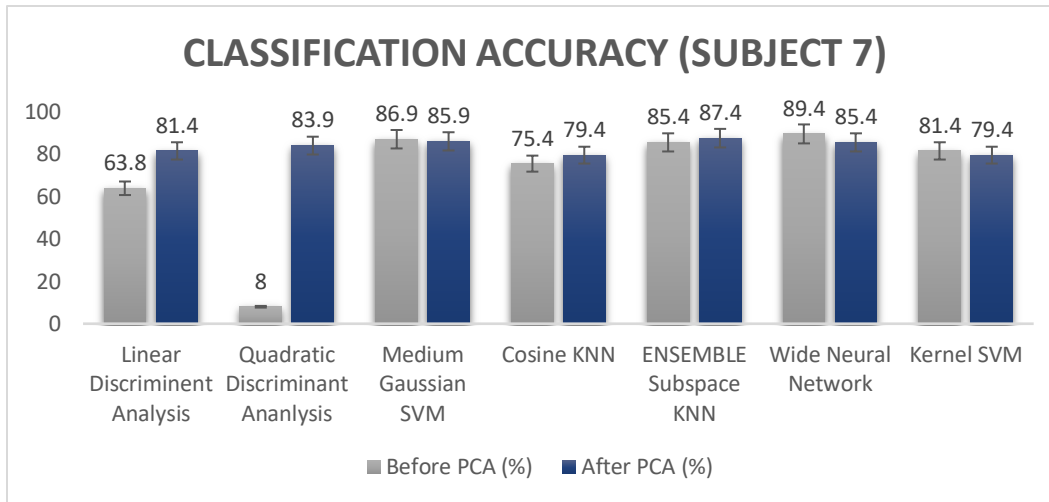
Subject 5:



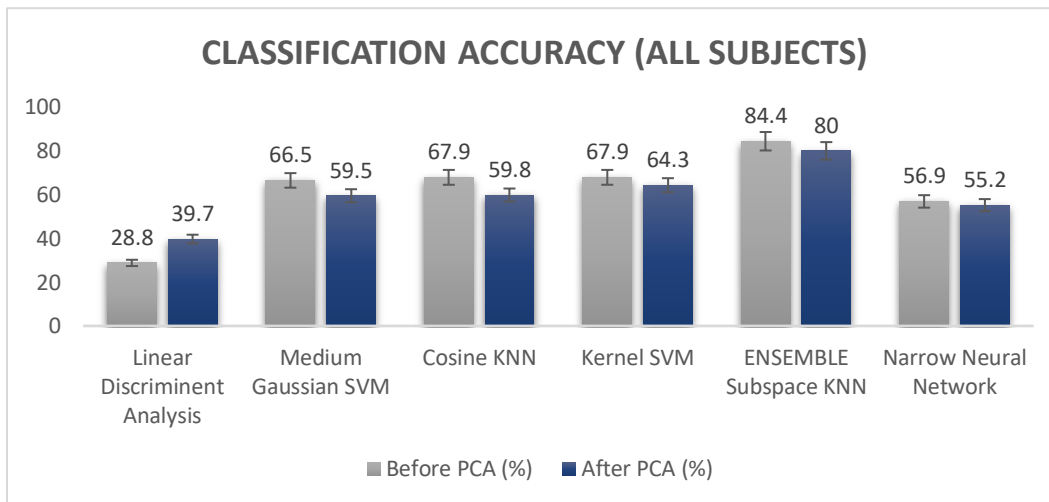
Subject 6:



Subject 7:



All Subjects:



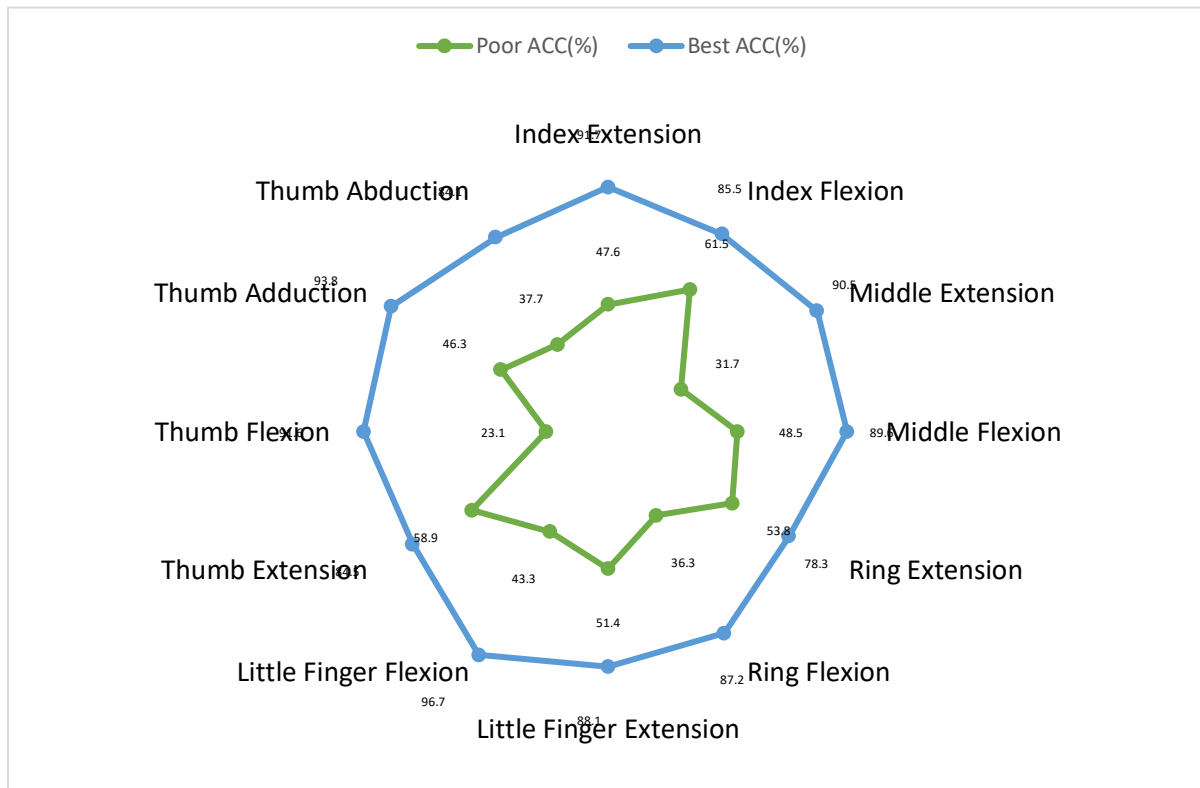
Cross validation means accuracy of ML algorithms based on variance:

The variance of the outcomes can reveal information about the stability and dependability of the predictions made by the model when evaluating the efficacy of ML algorithms applying cross-validation. The term "variance" describes the dispersion or variability of an assessment measure (such as accuracy) over various cross-validation folds. High variance means that the model is sensitive to the choice of training and evaluation subset. Prediction of the model can be varied depending on the individual subset of the data. It also shows that the model can be overfitting the training data and may not be able to generalize well to new and unseen data. Whereas low variance indicates the performance of the training dataset is more consistent across different validation folds. Generally low variance-based models are prioritized as they are more reliable and robust in nature. In our study, comparison of dataset has been done to analyze the

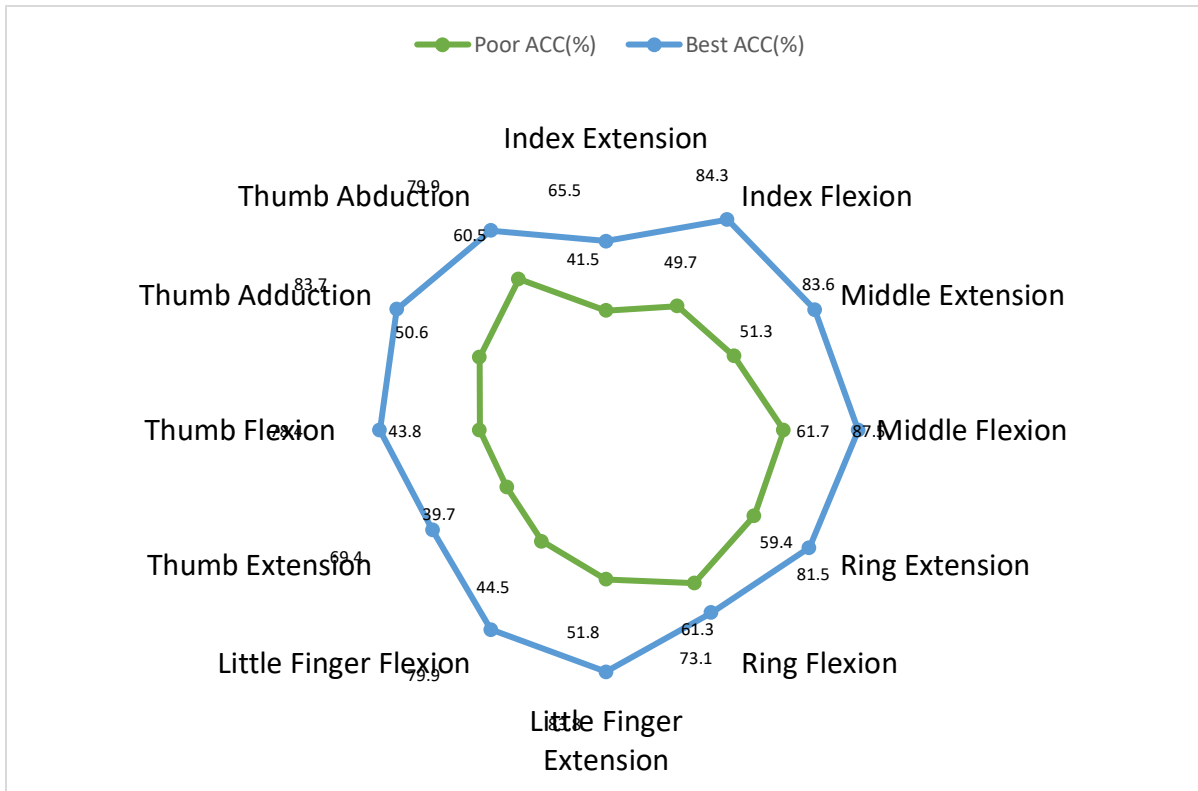
significant differences of variance. By increasing the variance levels, cross validation means accuracy has been increased respectively. Variance levels of 90%, 92%, 95%, 97%, and 99.7% have been implemented and their results are shown below.

The subsequent polar plots represent the best and worst accuracies of each subject and all subjects. The accuracy of every subject is based on the classifiers. Best performing subject shows the highest accuracy at the “Little Finger Flexion” movement i.e., 96.7%, and the lowest accuracy at the “Thumb Flexion” i.e., 23.1%. The worst performing subject shows the highest accuracy at the “Middle Flexion” movement i.e., 87.5%, and the lowest accuracy at the “Thumb Extension” i.e., 31.7%. For all subjects, the highest accuracy is at the “Index Extension” movement i.e., 88.7%, and the lowest accuracy is at the “Middle Extension” i.e., 21.1%.

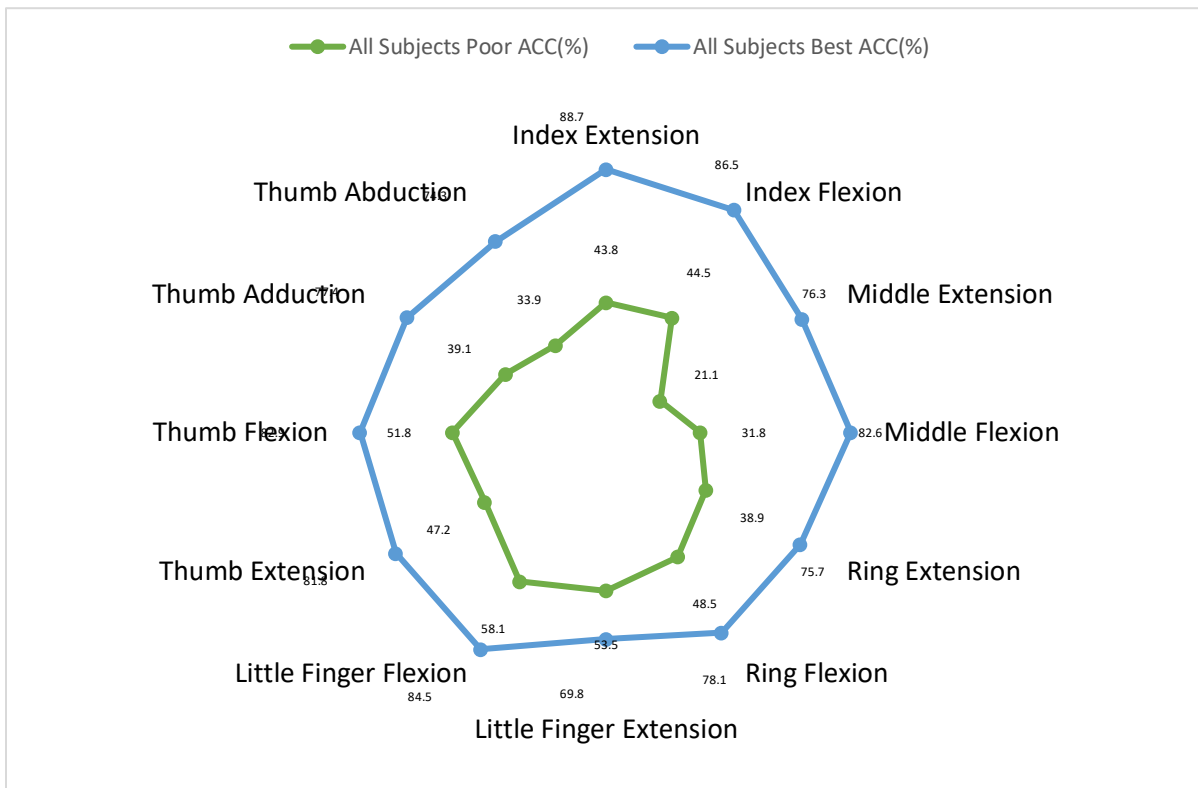
Best Subject:



Worst Subject:



All Subjects:



4.2 Discussion

Choosing 95% variance gives 65-70% accuracy before and after PCA, which means they slightly changed the classification accuracies. By changing the variance from 90% to 99.9%, an increase in accuracy can be seen in the results. By increasing the variance, the accuracies are also increasing which gives the idea about the working of PCA with respect to variance. Another method to determine the classification accuracy discussed in this research is to change the dimensions and observe the effect on classification accuracy. For this study, the data is reduced to 64 dimensions yielding classification accuracies between the range of 55%-65% which means the data is being compromised and information is lost. The data is reduced to 95 dimensions which are half of the total data to check the classification accuracies which are in the range of 70-80% meaning the data is being preserved well with these dimensions and gives better accuracy. PCA gives the best accuracy at half of its total dimensions i.e., 95 dimensions according to our dataset, the accuracy decreases before and after this dimension.

This approach has at least two possible uses, each of which has clear and illuminating examples from motor control literature. To characterize the temporal characteristics of activity related to various parts of the motor action, one potential use for the PCA might be to investigate the electromyographic recordings of specific muscles captured during several variations of one movement (or also different motor behaviors). This initial kind of PCA use is best illustrated by an investigation by Flanders and Herrmann [11], which was the first to employ PCA on EMG data. The study of electromyographic activity patterns captured by multi-electrode devices (up to 32 electrodes) from several muscles across the body throughout one or more motor activities is another possible use for PCA. Therefore, PCA can be used to find the spatial-temporal neuromuscular synergies that underlie motor behavior. Ivanenko and colleagues have conducted a number of insightful studies in this regard, using PCA in variable rotation to analyze EMG data collected from 16 to 32 muscles across the body over a number of movement conditions, including running and walking, different levels of body weight support (BWS), and the integration of movement with voluntary movements like kicking the ball or holding an item. [12][19][20][21].

4.3 Statistical Analysis

ANOVA (Analysis of variance) has been performed to compare EMG signal characteristics. It also helps determine the significant differences or relationships observed in the EMG data. The threshold for determining the significant difference between the EMG data before and after PCA depends on the α (level of significance). Optimal significant level is 0.05 or 0.01. Observing the p-value, data is analyzed whether they have a significant difference or not, if the p-value is less than or equal to the α , it shows the difference between the dataset and if the p-value is greater than or equal to the α , there is not significant difference

before or after implementing the PCA. ANOVA test has been performed to determine the p-values of the subjects and all subjects before and after PCA. If the p-value is in $0 \leq P \leq 0.5$ range, then the classification is said to be accurate. One way ANOVA test has been done is there are only two classes to be observed i.e. Before PCA and After PCA.

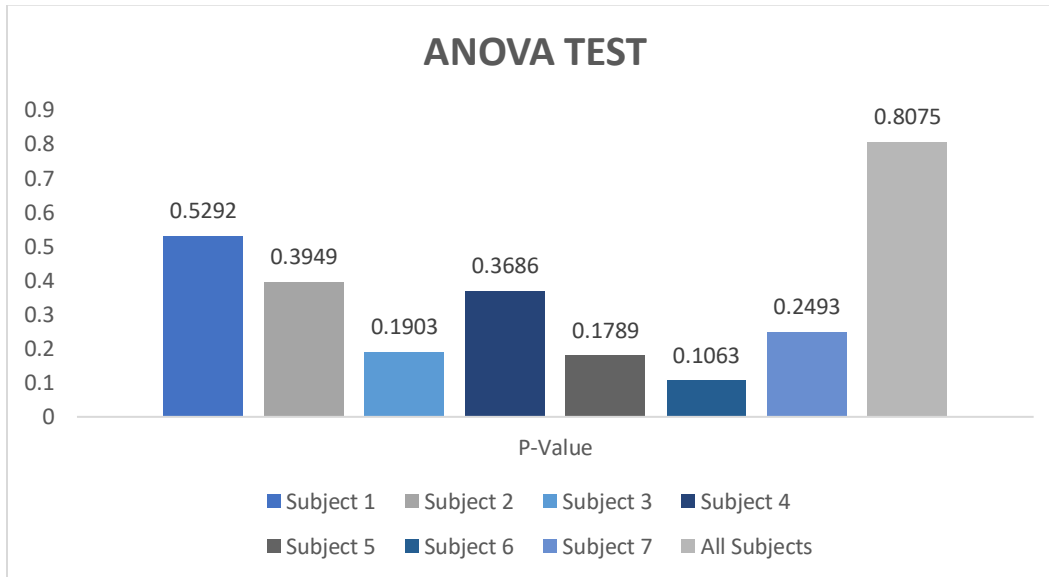


Table 2 One-way ANOVA (p-values)

Significant changes can be seen in these values which shows that the dimensionality reduction technique is working quite efficiently.

CHAPTER 5
CONCLUSION

5.1 Conclusion

In this study, a feature reduction technique has been implemented on the extracted data. The surface EMG signals were recorded for 7 healthy subjects. The subjects performed 12 basic hand movements along with a rest period. Nineteen time and frequency domain features from each segment of every movement's channel, presented in Table 1, were used to train, and test various classifiers. In conclusion, PCA is an effective technique for locating common temporal patterns in big collections of time series data, such as EMG signals, and it establishes a space with low dimensions in which the initial data points can be expressed as vectors and categorized. Classification accuracies of individual subjects before PCA and after PCA is more likely the same, in some subjects, the accuracy after PCA is higher which means reducing the dimensions yields more accuracy. Implementation of classifiers on individual subjects yields more classification accuracy rather than all subjects. The results showed that increasing the variance significantly increases the performance of the system as well as the complexity of the system's architecture. There is a trade-off between the performance of the system and its computational complexity. Choosing the classifier that best fits the dataset for individual subjects are SVM, KNN, ENSEMBLE BOOSTED TRESS, and ANN.

The subject-wise variability in the performance of the system has also been observed, suggesting that the different subjects provide varying results. Individual subject's classification accuracy before and after the implementation of PCA yields better results ranging between 70-85%. It classifies the data properly according to each movement and its channels, respectively. The classification accuracy of all subjects combined is observed, it ranges between 55-65% before and after the implementation of PCA. Reason for it could be the presence of outliers which make the data unable to be evaluated accurately, and the EMG data of each subject varies from one another, therefore, the PCA is not able to capture all the information due to the variation in the information and outliers makes the dataset less efficient, hence reducing the overall accuracy of the classification. Combining all subjects together makes the data's dimensions so high that PCA might not be able to capture the useful information, hence discarding the valuable features.

EMG data use for PCA discussed in this study demonstrates the effectiveness of this technique to extract features from surface EMG signals which may shed light on the nature of premotor control signals as well as the activation state of motoneurons, opening new avenues toward both neurophysiological and clinical/rehabilitation studies. The presented results can be used to design a better system for the diagnosis and pinpointed treatment of the upper limb injury based on the selection of features which are closely related to the injured muscles.

References

1. M Raez., et al., “*Techniques of EMG signal analysis: detection, processing, classification, and applications*”, Biological Procedures Online (BPO) in March 2006. V 8, P: 11–35.
2. Sachin., et al., “*Techniques for feature extraction from EMG signal*”, International Journal of Advanced Research in Computer Science and Software Engineering in January 2012. V 2, Issue 01.
3. A Phinyomark., et al., “*Feature extraction of the first difference of EMG time series for EMG pattern recognition*”, Computer Methods and Programs in Biomedicine in November 2014, Volume 117, Issue 2, P:247-256.
4. A Phinyomark., et al., “*Surface Electromyography (EMG) Signal Processing, Classification, and Practical Considerations*”, Biomedical Signal Processing in November 2019. P: 03-29.
5. J Jair., et al., “*Comparative analysis among feature selection of sEMG signal for hand gesture classification by armband*”, IEEE Latin America Transactions in June 2020. V: 18, I: 06, P: 1135 – 1143.
6. Tschanner V., et al., “*EMG. In: Nigg B, Herzog W, Eds. Biomechanics of the Musculo-Skeletal System*”, Hoboken, NJ (USA): John Wiley & Sons in 2007, 3rd Edition.
7. Rainoldi A., “*Spectral properties of the surface EMG can characterize/do not provide information about motor unit recruitment strategies and muscle fiber type*”, J Appl Physiol in 2008, P:105(5): 1678.
8. Kilby J., et al., “*Wavelet analysis of surface electromyography signals*”, Conf Proc IEEE Eng Med Biol Soc in 2004, V:1, P:384-7.
9. Maitrot A., et al., “*Signal-dependent wavelets for electromyogram classification*”, Med Biol Eng Comput, in 2005, V:43(4), P:487-92.
10. Gorsuch RL., “*Factor Analysis*”, Hillsdale, NJ (USA): Erlbaum in 1983.
11. Flanders M., et al., “*Two components of muscle activation: scaling with the speed of arm movement*”. J Neurophysiol 1992; 67(4): 931-43.
12. Ivanenko YP., et al., “*Five basic muscle activation patterns account for muscle activity during human locomotion*”. J Physiol 2004; 556(Pt 1): 267-82.
13. Cheung VC., et al., “*Central and sensory contributions to the activation and organization of muscle synergies during natural motor behaviors*”. J Neurosci in 2005; 25(27): 6419-34.
14. Tresch MC., et al., “*Matrix factorization algorithms for the identification of muscle synergies: evaluation on simulated and experimental data sets*”. J Neurophysiol in 2006; 95(4): 2199-212.
15. Naik GR., et al., “*Limitations, and applications of ICA for surface electromyogram*”. Conf Proc IEEE Eng Med Biol Soc in 2006; 1: 5739-42.
16. Tabernig CB., et al., “*M-wave elimination from surface electromyogram of electrically stimulated muscles using singular value decomposition: preliminary results*”. Med Eng Phys in 2008; 30(6): 800-3.
17. Gianfranco Bosco., “*Principal Component Analysis of Electromyographic Signals: An Overview*”, The Open Rehabilitation Journal in 2010. V: 3, P: 127-131.
18. Morrison DF. “*Multivariate Statistical Methods*”, New York (USA): McGraw-Hill 1990.

19. Ivanenko YP., et al., “*Temporal components of the motor patterns expressed by the human spinal cord reflect foot kinematics*”. J Neurophysiol in 2003; 90(5): 3555-65.
20. Ivanenko YP., et al., “*Coordination of locomotion with voluntary movements in humans*”. J Neurosci in 2005; 25(31): 7238-53.
21. Cappellini G., et al., “*Motor patterns in human walking and running*”. J Neurophysiol in 2006; 95(6): 3426-37.
22. Atzori M., et al., “*Building the NINAPRO Database: A Resource for the Biorobotics Community*”, Proceedings of the IEEE International Conference on Biomedical Robotics and Biomechatronic, June 2012.
23. De Luca, C.J., “*The use of surface electromyography in biomechanics*,” Journal of Applied Biomechanics, vol. 13, no. 2, pp. 135–163, 1997.
24. Castellini C., et al., “*Surface EMG in advanced hand prosthetics*,” Biological Cybernetics, vol. 100, no. 1, pp. 35–47, 2009.
25. Castellini C., et al., “*multi-subject / dailylife activity EMG-based control of mechanical hands*,” Journal of Neuroengineering and Rehabilitation, vol. 6, no. 41, 2009.
26. Fakuda O., et al., “*A human-assisting manipulator teleoperated by EMG signals and arm motions*,” IEEE Transactions on Robotics and Automation, vol. 19, no. 2, pp. 210–222, april 2003.
27. Castellini C., et al., “*Fine detection of grasp force and posture by amputees via surface electromyography*,” Journal of Physiology (Paris), vol. 103, no. 3-5, pp. 255–262, 2009.
28. Tenore., et al., “*Decoding of individuated finger movements using surface electromyography*,” IEEE Transactions on Biomedical Engineering, vol. 56, no. 5, pp. 1427–1434, 2009.
29. Tsuji H., et al., “*Discrimination of forearm motions from emg signals by error back propagation typed neural network using entropy*,” IEEE Transactions, Society of Instrument and Control Engineers, vol. 29, no. 10, pp. 1213–1220, 1993.
30. Hersh W., et al., “*The IMAGECLEFMED Medical Image Retrieval Task Test Collection*” Journal of Digital Imaging, V:22(6), P:648–655.
31. J. Nilsson, “*Implementing a continuously updating, high-resolution time provider for windows*,” The MSDN Magazine in 2004.
32. Zecca M., et al., “*Control of multifunctional prosthetic hands by processing the electromyographic signal*”, Critical Reviews in Biomedical Engineering, vol. 30, no. 4-6, pp. 459–485, 2002.
33. Bitzer S., et al., “*Learning EMG control of a robotic hand: Towards active prostheses*,” Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.
34. John P., et al., “*Linear Dimensionality Reduction: Survey, Insights, and Generalizations*”, Journal of Machine Learning Research in 2015.
35. Jackson J., et al., “*A user’s guide to principal components*”, The statistician, V:42, P:76,77 in March 1991.
36. Pelliccia, D. (2022). Retrieved from <https://nirpyresearch.com/pca-kernel-pca-explained/>.
37. Pedregosa., et al., “*Scikit-Learn: Machine learning in Python*”, Retrieved from <https://dl.acm.org>.
38. Adewuyi, A., et al., “*Evaluating EMG Feature and Classifier Selection for Application to Partial-Hand*

- Prosthesis Control*”, Front Neurorobot, 2016, V:10, P:15.
39. Ortiz, M., “*Cardinality as a highly descriptive feature in myoelectric pattern recognition for decoding motor volition*”, Front Neurosci, 2015, V:9, P: 416.
 40. Hudgins, B., et al., “*A new strategy for multifunction myoelectric control*”. IEEE Trans Biomed Eng, 1993. V:1, P: 82-94.
 41. Turgunov, A., et al. “*Using Different Features of Signal in EMG Signal Classification*”. International Conference on Information Science and Communications Technologies (ICISCT) in 2020.
 42. Phinyomark, A., et al., “*Feature reduction and selection for EMG signal classification*”. Expert Systems with Applications, 2012. V:39(8): P: 7420- 7431.
 43. Raurale, S., et al., “*EMG Wrist-hand Motion Recognition System for Real-time Embedded Platform*”. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). 2019.
 44. Tae, A.A.A., et al., “*Cardinality and Short-Term Memory Concepts based Novel Feature Extraction for Myoelectric Pattern Recognition*”. 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2021.
 45. Elamvazuthi, I., et al., “*Electromyography (EMG) based Classification of Neuromuscular Disorders using Multi-Layer Perceptron*”. Procedia Computer Science, 2015, P:76.
 46. Garikayi, T., et al., “*Analysis of surface electromyography signal features on osteomyoplastic transtibial amputees for pattern recognition control architectures*”. Biomedical Signal Processing and Control, 2018, V: 40, P: 10-22.
 47. Kim, K.S., et al., “*Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist- motion directions*”. Current Applied Physics, 2011, V:11(3), P: 740-745.
 48. Zhou, S., et al., “*Surface electromyography-based hand movement recognition using the Gaussian mixture model, multilayer perceptron, and AdaBoost method*”. International Journal of Distributed Sensor Networks, V:15(4), P:155014771984606.
 49. Wolczowski, A., et al., “*Control of dexterous hand via recognition of EMG signals using combination of decision-tree and sequential classifier, in Computer Recognition Systems 2*”. Advances in Soft Computing, P:687–694.
 50. Mukhopadhyay, A.K., et al., “*An experimental study on upper limb positioninvariant EMG signal classification based on deep neural network*”, Biomedical Signal Processing and Control, 2020, V:55, P:101

THESIS

ORIGINALITY REPORT

ASW

Dr. Asim Waris
Assistant Professor
School of Mechanical &
Manufacturing Engineering
(SMME) NUST, Islamabad

10% SIMILARITY INDEX	8% INTERNET SOURCES	5% PUBLICATIONS	2% STUDENT PAPERS
--------------------------------	-------------------------------	---------------------------	-----------------------------

PRIMARY SOURCES

1	www.kdnuggets.com Internet Source	2%
2	benthamopen.com Internet Source	1%
3	infoscience.epfl.ch Internet Source	1%
4	community.nasscom.in Internet Source	1%
5	Submitted to University of East Anglia Student Paper	<1%
6	arch.library.northwestern.edu Internet Source	<1%
7	Submitted to University of East London Student Paper	<1%
8	link.springer.com Internet Source	<1%
9	www.researchgate.net Internet Source	<1%