Enhancing Performance of EEG based Machine Learning Algorithm via Feature Fusion and Dimensional Reduction Techniques

by

Ong Zhi Ying
(1741312604)

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

Faculty of Electronic Engineering & Technology
UNIVERSITI MALAYSIA PERLIS

2024

ACKNOWLEDGEMENT

First and foremost, I would like to express my greatest gratitude towards Dr. Saidatul and Dr. Vikneswaran for having them as my supervisors. I am fortunate and grateful to have them as my supervisors who inspired me to work on this project. Their wide knowledge has been of great value to me. Their willingness to motivate me contributed tremendously. Thank you for the continuous support of my research study, and their guidance, patience, motivation, enthusiasm, and immense knowledge throughout this project. I express my sincere appreciation to them for providing their invaluable assistance in writing coding and this report.

Besides my supervisors, I would like to thank Dr. Maz Jamilah from Engineering Mathematics Universiti Malaysia Perlis for her generous support, insightful suggestions and kindliness in assisting me. I am thankful to Mr. Ang Wei Chern from Hospital Tuanku Fauziah, Kangar, Perlis for his encouragement, insightful suggestions, and kindliness in assisting me. I would like to thank my fellow UniMAP subjects for contributing their time and EEG signals in completing the data acquisition section.

Millions of thanks to Assoc. Prof. Dr Ruslizam, Assoc. Prof. Ir. Dr. Ahmad Faizal, Assoc. Prof. Dr Shuhaida, and Assoc. Prof. Dr. Haniza who guided me in my postgraduate study such as application for conversion and submission of the thesis. I would like to thank the IRB/IEC Medical Research Ethics Committee (MREC) from the National Medical Research Registry (NMRR) for approving data collection from humans as a subject.

Next, I would like to express my sincere appreciation to the School of Mechatronic Engineering, UniMAP for the willingness to provide accessibility to the ANT Neuro EEGOTM Sport device and MRI Laboratory to ease the data acquisition process.

Last but not least, thanks to all my family members for their unconditional love, full support, understanding, and encouragement throughout my tough period in completing my studies. They were always there cheering me up and standing by me through the good and bad times. Words cannot express how grateful I am to my parents for all the sacrifices that they have made on my behalf.

Finally, I would like to thank all of my friends who have always been by my side, encouraged and supported me to strive towards my goal. Thanks to Dr. Sucinda who accompanied me throughout the period in completing my study. Thanks to Dr. Izzawati and Dr. Muhammad who always provided me with the information when I needed it.

Without any of them, this research project would not have been a success and I would never be able to finish my dissertation. Without them, I would not have been able to face the difficulties which arose during my research period. I would like to express my deepest gratitude and thanks to all of them for their encouragement, support, superior knowledge and experience throughout my degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABC	Artificial Bee Colony
ABI	Acquired Brain Injury
AC	Alternating Current
ACO	Ant Colony Optimization
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ApEn	Approximate Entropy
AR	Auto Regression
ATT	Adaptive Temporal Transformer
BCD	Box-Counting Dimension
BCI	Brain-Computer Interface
BLDA	Bayesian Linear Discriminant Analysis
bpm	Beats per Minute
C	Central
CAR	Common Average Reference
CCA	Canonical Correlation Analysis
CCA-C	Canonical Correlation Analysis by Concatenation
CCA-S	Canonical Correlation Analysis by Summation
CCDF	Canonical Correlation Discriminant Features
CFS	Correlation-based Feature Selection
CNN	Convolutional Neural Network
CNS	Central Nervous System
CR	Cardinality
CSP	Common Spatial Pattern
db	Daubechies
DE	Differential Evolution
DGAFF	Deep Genetic Algorithm Fitness Formation
DistEn	Distribution Entropy
DLDA	Diagonal Linear Discriminant
DOF	Degrees of Freedom
DWT	Discrete Wavelet Transform
ECG	Electrocardiography

EEG	Electroencephalogram
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
EMG	Electromyography
EOG	Electrooculography
F	Frontal
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FKNN	Fuzzy k-Nearest Neighbors
FLD	Fisher’s Linear Discriminant
fMRI	Functional Magnetic Resonance Imaging
FN	Number of False Negatives
FP	Number of False Positives
Fp	Pre-frontal
FT	Fourier Transform
FuzEn	Fuzzy Entropy
GA	Genetic Algorithm
GAN	Generative Adversarial Network
HE	Hurst Exponent
HELM	Hybrid Extreme Learning Machine
HICS	Hierarchical Incremental Channel Selection
HT	Hilbert Transform
ICA	Independent Component Analysis
IEC	Independent Ethics Committee
IQR	Interquartile Range
IRB	Institutional Review Board
ITD	intrinsic Time-scale Decomposition
KELM	Kernel Extreme Learning Machine
KNN	k-Nearest Neighbor
LDA	Linear Discriminant Analysis
LSTM	Long-term Short-term Memory
MATLAB	Matrix Laboratory
Max	Maximum
MDA	Minimum Distance Analysis

ME	Motor Execution
MET	Motor Execution Task
MFBF	Multifrequency Band Fusion
MIMFs	Multivariate Intrinsic Mode Functions
Min	Minimum
MI	Motor Imagery
MIT	Motor Imagery Task
ML	Machine Learning
MLP-NN	Multi-Layer Perceptron Neural Network
MREC	Medical Research Ethics Committee
mRmR	Minimum Redundancy and Maximum Relevance
MSATNet	Multi-Scale Adaptive Transformer Network
MSFE	Multi-Scale Feature Extraction
MST	Modified Stockwell Transform
NB	Naive Bayes
NMRR	National Medical Research Registry
O	Occipital
P	Brain Lobe: Parietal & Classification: Number of Predictors
PCA	Principal Component Analysis
PCC	Pearson Correlation Coefficient
PermEn	Permutation Entropy
PET	Positron Emission Tomography
PNN	Probabilistic Neural Network
PPV	Positive Predictive Values
PSD	Power Spectral Density
PSO	Particle Swarm Optimization
Q1	First Quartile
Q3	Third Quartile
RBF	Radial Basis Function Network
RF	Random Forest
RFE	Recursive Feature Elimination
ROI	Regions of Interest
SA	Simulated Annealing
SampEn	Sample Entropy

SMMSE	Standardized Mini-Mental State Examination
SPECT	Single-photon Emission Computed Tomography
Std	Standard Deviation
SVM	Support Vector Machine
SWLDA	Stepwise Linear Discriminant Analysis
T	Temporal
TN	Number of True Negatives
TP	Number of True Positives
TPR	True Positive Rates
UniMAP	Universiti Malaysia Perlis
Var	Variance
WHO	World Health Organization
WPD	Wavelet Packet Decomposition
WT	Wavelet Transform

LIST OF SYMBOLS

$	Dollar
%	Percentage
&	And
δ	Delta
θ	Theta
α	Alpha
β	Beta
γ	Gamma
μV	Microvolt
Φ^m	Function
Σ	Summation
ϵ	Belongs to
τ	Time delay
!	Factorial
λ	Eigenvalues
$\ln m!$	Normalization factor
$\Theta(\,)$	Heaviside function
$\|\| \|\|_1$	Chebyshev distance
Λ_2	Squares of canonical correlations / diagonal matrix of eigenvalues
a	Scaling factor
a_0	Average of the m values
A	Attenuation
$AgCl$	Silver chloride
b	Translation factor
c	Class
C_P^m	Proportion
Ca^{++}	Calcium ions
Cl^-	Chloride ion
cm	Centimetre
D	Distance matrix
D_{ij}^m	Similarity degree
d	Number of non-zero eigenvalues
d_{ij}^m	Maximum distance of reconstruction vector
f	Wavelet Transform: Function & Butterworth Filter: Frequency
f_c	Cut-off frequency

F_s	Filter selectivity
$h(t)$	Mother wavelet
$H(j\omega)$	Magnitude response
Hz	Hertz
I	Identity matrix
K	Welch's Method: Number of available frames & SampEn: Length & CSP: Number of trials
$K+$	Potassium ion
k	PermEn: Count of tuples & KNN: Number of neighbours
$k\Omega$	Kiloohm
M	Welch's Method: Number of points in each segment or batch size & DistEn: Number of bins & CSP: Covariance
m	ApEn, SampEn: Embedding dimension & PermEn: The order of permutation entropy
mV	Microvolt
N	Butterworth Filter: Filter order & SampEn, DistEn: Signal length, Sampling frequency & PermEn: Number of samples & CSP: Count of channels
n	Welch's, mean, Var, Std: Number of partition data (n = 0,1,…, M – 1) & FuzEn: Gradient of the exponential function & CCA: Number of training feature vectors
$Na+$	Sodium ion
P	Welch's Method: Modified periodogram value & PermEn, DistEn: Probability
p^{th}	Number of windowed blocks
R	Window hop size
R^m	Embedding space
r	Tolerance window, Width of the exponential function
s	Seconds
s_i	i^{th} motif
S	Covariance matrix
T	HE: Duration & CSP: Count of samples
t	PermEn: Time & DistEn: Bin
U_i^m	Reconstruction of EEG time series set
V_P	Vector
ω	Angular frequency

ω_c	Cut-off frequency which expressed as an angular value
w	Window function
w_j	j[th] eigenvector
W	CSP: Spatial filters & CCA: Transformation matrices
$\hat{W}$	Eigenvector
W^T	Transpose of W
x	Welch's Method, mean, Var, Std, PermEn: Signal & Pre-processing: Current sample
X	SampEn: Sequence & PermEn: Tuple & CSP, CCA: Data matrix
X^*	Canonical variate
x_{max}	Maximum values
x_{min}	Minimum values
y	Normalized sample
Y	Matrix
Y^*	Canonical variate
Z	Canonical Correlation Discriminant Features

Meningkatkan Prestasi Algoritma Pembelajaran Mesin berasaskan EEG melalui Gabungan Ciri dan Teknik Pengurangan Dimensi

ABSTRAK

Teknologi neuro-prostetik telah muncul dengan pantas sebagai penyelesaian utama untuk meningkatkan kehidupan manusia. Pada masa ini, terdapat beberapa cabaran dijumpai dalam peranti prostetik konvensional. Dalam usaha untuk meningkatkan tahap tertinggi kehidupan OKU, isyarat EEG merupakan salah satu kaedah alternatif untuk memulihkan fungsi motor yang hilang. Ini telah menghasilkan pembangunan algoritma pembelajaran mesin (ML) yang boleh digunakan untuk aplikasi neuro-prostetik. Neuro-prostetik adalah teknologi yang berkembang pesat yang bertujuan untuk membina saluran terus antara manusia dan peranti luaran tanpa menggunakan otot dan saraf periferi pada anggota badan. Walau bagaimanapun, neuro-prostetik berasaskan imejan motor (MI) terhad dalam bilangan arahan kawalan. Banyak ciri dan saluran membawa kepada peranti neuro-prostetik yang memerlukan spesifikasi yang lebih maju dan masa untuk memproses dimensi data yang besar. Keberkesanan algoritma ML berasaskan EEG selepas mengurangkan dimensi data diperlukan untuk menilai. Oleh itu, matlamat penyelidikan ini adalah untuk menyiasat isyarat EEG dengan mengekstrak maklumat tersembunyi dari isyarat pergerakan motor yang berbeza dan mengoptimumkan dimensi data tanpa menjejaskan prestasi algoritma ML. Isyarat EEG bagi pergerakan motor yang berbeza boleh dirakam menggunakan peranti EEG. Ciri penting boleh diekstrak dan dimensi data boleh dikurangkan menggunakan teknik pemprosesan isyarat lanjutan. Dalam kajian ini, protokol pemerolehan telah diwujudkan dan algoritma berangka telah dibangunkan untuk mengoptimumkan semua pembolehubah yang diekstrak untuk aplikasi neuro-prostetik. Pemprosesan isyarat lanjutan seperti analisis berbilang resolusi frekuensi masa dan analisis bukan linear menggunakan perisian MATLAB telah dijalankan untuk pencirian ciri-ciri gelombang otak. Ciri PSD berasaskan statistik dan entropi, parameter Hjorth, dan HE telah diekstrak daripada lima jalur frekuensi. Penguasaan ciri dari jalur frekuensi tertentu dalam jumlah jalur frekuensi telah disiasat dan saiz data yang besar berjaya dikurangkan dengan menghapuskan ciri dan saluran yang tidak berkaitan menggunakan plot kotak dan GA. Ciri gabungan dari berbilang jalur frekuensi untuk pencirian gelombang otak memberikan pemahaman yang tepat tentang membangunkan algoritma ML berasaskan EEG untuk aplikasi neuro-prostetik. CSP dan CCA dicadangkan untuk gabungan ciri jalur frekuensi berbilang. ANOVA dan klasifikasi seperti KNN, SVM, pengelas pokok, dan pengelas ensembel digunakan untuk peringkat pengujian dan pengesahan. Baki ciri dan saluran adalah ciri berasaskan statistik PSD dan Pz-CPz' (C26). CCA mengikut penjumlahan dan subruang KNN adalah kombinasi terbaik kerana ketepatannya, "recall", dan "precision" adalah 100.00%. Hasil penyelidikan menentukan kemungkinan gelombang otak sebagai alat utama untuk memulihkan fungsi motor yang hilang. Penemuan ini menyumbang kepada pemahaman awal tentang algoritma ML dan mencadangkan semakan lengan robot terutamanya pada bilangan arahan kawalan dan pengoptimuman ciri dan saluran EEG untuk aplikasi neuro-prostetik. Algorithm ML yang dicadangkan ini dijangka boleh diperluaskan kepada ujian bebas subjek dengan bantuan perkakasan dan penerimaan deria boleh disertakan pada masa hadapan.

Enhancing Performance of EEG based Machine Learning Algorithm via Feature Fusion and Dimensional Reduction Techniques

ABSTRACT

Neuro-prosthetics technology has swiftly emerged as a go-to solution for improving human lives. Currently, there are some challenges found in conventional prosthetic devices such as cosmetic prosthetic hands and myoelectric prosthetic hands. In the drive to improve the highest level of a disabled individual’s life, the EEG signal is one of the alternative methods to restore lost motor functions. This has resulted in the development of a machine learning (ML) algorithm that can be used for reliable neuro-prosthetics applications. Neuro-prosthetics is a fast-growing technology which aims to build a direct channel between humans and external devices without using the muscles and the peripheral nerves in the limbs. However, motor imagery (MI)-based neuro-prosthetics are limited in the number of control commands. Numerous features and channels lead to neuro-prosthetics devices requiring more advanced specifications and time to process the large data dimension. The effectiveness of the EEG-based ML algorithm after reducing the data dimensionality is required for evaluation. Consequently, the goal of this research is to investigate the EEG signals by extracting hidden information from the signals of different motor movements and optimizing the data dimensionality without affecting the performance of the ML algorithm. The EEG signals of different motor movements can be recorded using EEG devices. Important features can be extracted and the data dimensionality can be reduced using advanced signal processing techniques. In this study, the acquisition protocol was established and the numerical algorithm was developed to optimize all variables extracted for the neuro-prosthetics applications. Advanced signal processing such as time-frequency multi-resolution analysis and non-linear analysis using MATLAB software was conducted for the characterization of brain wave features. Statistical-based and entropy-based features, Hjorth parameters, and HE of PSD were extracted from five frequency bands. The dominance of features from particular frequency bands in total frequency bands was investigated and the enormous data size was successfully reduced by eliminating the irrelevant features and channels using box plots and GA. The fusion features from multiple frequency bands for brain wave characterization provided an accurate understanding of developing an EEG-based ML algorithm for the neuro-prosthetics application. CSP and CCA were proposed for multiple-frequency band feature fusion. ANOVA and classification such as KNN, SVM, tree classifiers, and ensemble classifiers were used for the testing and validation stage. The remaining features and channels were statistical-based features of PSD and Pz-CPz’ (C26). CCA by summation and subspace KNN were the best combinations due to the accuracy, recall, and precision being 100.00%. The research outcome dictates the possibility of brain waves as an ultimate tool to restore lost motor functions. This finding contributes to an early understanding of the ML algorithm and proposes the revision of the robot arm especially on the number of control commands and the optimization of EEG features and channels for neuro-prosthetics applications. It is expected that this proposed ML algorithm can be extended to subject-independent tests with the help of hardware and the sensory reception can be included in the future.

CHAPTER 1 : INTRODUCTION

1.1 Research Background

In this era, the advancement of technology had a great impact on human life. However, elderly or disabled individuals still face difficulties in their daily lives (Chaudhry et al., 2022). Amputation of limbs is a transformative and debilitating life event. This event can drastically impair life's quality. Consequential psychosocial stressors have an impact on personal relationships, careers, and even threaten an individual's sense of self (Karczewski et al., 2021). Throughout the world, the total number of amputations is estimated to be about 65 million people and approximately 34% are upper-limb amputations (ATscale, 2020). Some people were born without limbs (Editorial of UPMC Health Beat & Rehabilitation, 2015). 5-6 million people faced partial hand amputation because of various health issues, traumatic accidents, and even wars (Chaudhry et al., 2022). This indicates that approximately 1 in every 50 individuals suffers from loss of motion (Sahu & Shukla, 2019).

In the United States, roughly 1.6 million people with limb loss in the year 2005, approximately 2 million people with limb loss in the year 2015 (Karczewski et al., 2021) (Editorial of UPMC Health Beat & Rehabilitation, 2015), and the total number of people suffering from limb loss is anticipated to reach 3.6 million by the year 2050. 35% of people with limb loss which is 541 thousand people have their upper extremity amputated. Among the people with upper limb loss, amputations above or below the elbow accounted for 41, 000 cases while amputations of fingers or hands accounted for 500, 000 cases. Upper limb amputations are disastrous incidents for patients despite the amputations being a small part of the upper limb. This causes an inability to do activities of daily life (Thiamchoo & Phukpattaranont, 2022). Figure 1.1 shows one of the upper-limb amputation patients.

Figure 1.1 Upper-limb amputation patient (Henson, 2021).

As the number continues to climb, the use of advanced prosthetics has been pushed to enhance (Karczewski et al., 2021). Prosthetics are devices that are designed to replace lost body parts or help the existing body parts to function more effectively (Editorial of UPMC Health Beat & Rehabilitation, 2015). According to the World Health Organization (WHO), only 10% of amputation patients can access prosthetic devices. This is mainly due to the high cost of the commercially-available and certified devices, as well as the shortage of personnel and infrastructures (Geneva, 2017). Biddiss et, al. 2007 also reported that the weight of upper-limb prostheses, lack of comfort, fatigue, or dexterity is just as or more functional without using the upper-limb prostheses causing prosthesis abandonment. 10% to 50% of users have stopped using upper-limb prosthesis devices due to the devices being inconvenient for daily use. The percentage of users who stopped using upper-limb prostheses varies with age (Biddiss & Chau, 2007).

The prosthetic arms are purely mechanical and tedious (Chaudhry et al., 2022). The drawbacks of the existing conventional prosthetics and traditional myoelectric prosthetics provide the focusing areas for the development of a new system. The focus areas are to improve the robustness and dexterity of the prostheses for daily activities. The new system would go beyond the traditional myoelectric control of the prostheses (Piozin et al., 2022). As a result, there is a growing demand and necessity for the development of an alternative interface that can be utilised to communicate with autonomous systems by the population with severe disabilities (Chaudhry et al., 2022).

A non-invasive method of measuring brain activity with an electroencephalogram (EEG) is widely common in research. The word “electro” refers to electrical, “encephalo” is related to the brain, “graphy” refers to the technique or method, and “gram” refers to the written record of the bioelectrical signals of the brain. An electroencephalogram is defined as the electrical activity of an alternating type recorded from the scalp surface after being picked up by metal electrodes and conductive media (Teplan, 2002). An electroencephalograph is an old-fashioned term for electroencephalography which is for the machine or instrument used to record an electroencephalogram. It provides insight into the status of the electrical activity of the brain (Roy, 2022). Electroencephalography is a medical imaging technique (Alam & Samanta, 2017). It is a completely non-invasive procedure of recording and interpreting scalp electrical brain activity (The Editors of Encyclopædia Britannica, 2023). There is virtually no risk or limitation when the electroencephalographic reading repeatedly to humans (Teplan, 2002). It is usually achieved through the placement of electrodes on the scalp.

EEG is a very powerful tool in the field of neurology and clinical neurophysiology. The evidence shows that EEG signals are feasible to find repeatable and stable brainwaves (Dũng et al., 2016). EEG enables the observation of gross electrical fields of the brain. The changes in neural mass activity associated with various mental processes are also reflected by EEG (Alam & Samanta, 2017). Furthermore, EEG is more efficient compared to other methods such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) which have time resolution in terms of minutes and seconds. This is because approximately 0.5 – 130 milliseconds are required to propagate an action potential across a single neuron. EEG explores the activity of brain and able to detect the changes within milliseconds. In addition, EEG measures the electrical activity of brain directly while fMRI and single-photon emission computed tomography (SPECT) record blood flow changes or PET record metabolic activity changes. These changes are indirect electrical activity markers that belongs to the brain (Roy, 2022).

The early researches are more towards the investigation of EEG for behavioural research on cognition neuroscience processes (Darriba & Waszak, 2018), medical diagnosis of neurological disorders (Bosl et al., 2018; Chaudhry et al., 2022; D. Chen et al., 2017), neurofeedback treatments (Franchi et al., 2021; Guillard et al., 2021), and neurorehabilitation such as motor impairment following stroke (Cantillo-Negrete et al., 2021; Ramos-Murguialday et al., 2013; Tam et al., 2011) or spinal cord injury (Benabid et al., 2019; Wodlinger et al., 2015) These researches set the stage for restoring and potentially upgrading human physical and mental capabilities (E. Mohammadi et al., 2021). EEG also has been widely studied in other fields such as video games (Congedo et al., 2011), biometric authentication (Dũng et al., 2016; Jayarathne et al., 2016;

Svetlakov et al., 2022), neuro-marketing (Naim, 2022; Stasi et al., 2018), and brain-computer interface (BCI) (Ak et al., 2022; Lazarou et al., 2018; Venkatachalam et al., 2020). BCI systems are commonly applied in lie detection, security, telepresence, alertness monitoring, education, gaming, art, virtual reality, and human augmentation (E. Mohammadi et al., 2021). It is also able to improve the precision of vehicles and robot control in hostile environments such as space, enable humans to live in intelligent e-homes, integrate new electronic body enhancements, and play and interact in novel ways (Bakardjian, 2010).

Considering the nature of the application of the prosthetic, the EEG signal is appropriate for the application of the prosthetic due to its ease of use and good temporal resolution the signal (Piozin et al., 2022). EEG signals can be easy to obtain and quantify without discomfort (Zúquete et al., 2010). Therefore, the issues of conventional and myoelectric prosthetics are addressed by introducing neuro-prosthetics into the robot control (Chaudhry et al., 2022). Neuro-prosthetics technology is hardware and software that establishes a novel communication pathway between the human brain and prosthetic devices directly. It allows people to send commands to the external world using brain activities rather than relying on the brain's typical output pathway of peripheral nerves and muscular activities (Chumerin et al., 2013). Therefore, a motor imagery (MI)-based machine learning (ML) algorithm can bypass the conventional channels of communication (Millan et al., 2010). It allows disabled individuals to control the functionality of external devices voluntarily (E. Mohammadi et al., 2021) and helps to restore lost motor functions (Jiang et al., 2015) by translating the motor intention into a control signal (Sahu & Shukla, 2019). In short, the EEG signals can be used for building a robust ML algorithm for the neuro-prosthetics application.

However, the EEG-based ML algorithm for the neuro-prosthetics application still faced some limitations in control command and data dimensionality. The number of control commands remains limited. It limits the usability of ML algorithms in control applications (Jiang et al., 2015) due to none of the ML algorithms are currently able to fully restore limb function (Karczewski et al., 2021; Sakti et al., 2021). Next, there is redundancy in data due to the number of optimized features is still large and the high number of channels (Jindal et al., 2022; Venkatachalam et al., 2020). Identifying the most effective features is still a challenging task for neuro-prosthetics. Furthermore, optimizing the EEG dimension without affecting the accuracy is still a challenging task for EEG-based ML algorithms (Lazurenko et al., 2022; Shedeed & Issa, 2016). The evaluation of the EEG-based ML algorithm is required after reducing the EEG data dimensionality. Therefore, the goal of this research was to provide an ML algorithm for neuro-prosthetics devices. To ensure the effectiveness of the neuro-prosthetics devices, the dimension of the EEG signals was diminished by eliminating data redundancy without losing and distorting useful information.

In this research project, a database for the neuro-prosthetics application was established. 36 subjects were recruited for data collection according to the central limit theorem. Advanced signal processing was carried out to extract the most useful information features which can be used for neuro-prosthetics. In EEG signal processing, there were several stages involved namely pre-processing, feature extraction, feature selection, channel selection, feature fusion and classification. Pre-processing is the stage of removing unwanted signals and feature extraction is to extract the most useful information features. Feature selection and channel selection select the most suitable and relevant features and channels for classifiers. The features and channels were optimized for a neuro-prosthetics application. Feature fusion was proposed in this uni-model to further enhance the performance of ML algorithms for neuro-prosthetics applications. Classification of EEG signals was carried out to validate the proposed ML algorithm. At the end of this study, the dominance of features from particular frequency bands in total frequency bands was investigated to further improve the performance of the ML algorithm for neuro-prosthetics.

1.2 The Research Problem/Problem Statement

There are many neuro-prosthetics applications all over the world. EEG for machine learning (ML) algorithms has received increasing attention in recent years. Therefore, the EEG signal is a platform to be explored for the ML algorithm. However, the domain of EEG-based ML algorithms still faced the problems of improving accuracy and robustness. Approximately 20% of ML users do not demonstrate motor imagery (MI) performance adequate to control the device effectively (Choi et al., 2017). Whereas the performance of the remaining users is mediocre (Santos et al., 2023). The accuracy of the MI-based ML algorithm for 10% to 50% of users is below 70% (Ek-Fliesberg et al., 2023; Shu et al., 2018). In addition, there are limitations in control command and data dimensionality.

1. The usability of ML algorithms in control applications involving multiple degrees of freedom (DOF), such as control of a robot arm is restricted by the limited number of control commands available in MI-based ML (Jiang et al., 2015). So far, limited studies found to investigate the fundamentals behind motor imagery tasks for the ML algorithm. Therefore, the functionality of neuro-prosthetics is limited due to none of the ML

algorithms are currently able to fully restore limb function (Karczewski et al., 2021; Sakti et al., 2021). The impact of using multiple limbs motor imagery tasks for generating EEG signals has not been investigated (Yi et al., 2016).

2. In signal processing, it is very difficult to get useful information from these signals directly in the time domain just by observing them due to brain signals containing a large amount of physiological and pathological information. The number of optimized features is still large. Accurately wise, the exact feature is still unknown performance to describe the capability of EEG signal for neuro-prosthetics. Identifying the most effective techniques for features is challenging due to the highly non-linear, nonstationary and artefact-prone nature of EEG data. In addition, there is redundancy in data due to the high number of channels (Jindal et al., 2022; Venkatachalam et al., 2020). The more the number of channels, the more the number of features. Additional time, equipment, and computational expenditures resulted from increasing the number of recorded electrodes (Cowley et al., 2017; Ghorbanzadeh et al., 2023; Qi et al., 2020; D. Wang et al., 2012). An extensive, quantitative comparison of the features and channels in the neuro-prosthetics application in recent literature is absent (Ramos et al., 2016; Stephe et al., 2022). To date, feature-level fusion is still new in neuro-prosthetics applications. There is limited literature on using feature fusion methods for neuro-prosthetics. The challenge of applying fusion features is how features can be combined to reduce data dimensionality and possibly improve the prediction rate

(Rahman et al., 2021). In short, optimizing the EEG dimension for neuro-prosthetics is still a challenging task (Lazurenko et al., 2022; Shedeed & Issa, 2016).

3. Little is known about the effectiveness of ML techniques for classifying the brainwave data. The results between different studies are difficult to compare due to different experiment environments and signal processing techniques. Studies have shown that the accuracy of results can be strongly affected by various factors such as data acquisition, signal processing techniques, and classification techniques (Sohaib et al., 2013). Although generally the easy-to-use nature of ML algorithms in neuroscience, ML algorithms can produce false and inaccurate results if implemented incorrectly (Shoorangiz et al., 2023). The real-time processing of EEG signal classification for implementation on embedded systems or edge computing devices is still an open research problem. The accuracy and reliability of EEG signals for programming human-like movements are required to improve. Therefore, research on the development of robust, reliable, and efficient ML algorithms for EEG signal classification should be pursued (Ramírez-Arias et al., 2022).

1.3 Hypotheses

Human brains control all activities of the human body including movement, imaging, and body responses. Brains receive and give feedback to the human body to react. Although the brain still gives the response, the lack of actuators (loss of limb) makes the task non-executable. Via EEG these instructions can be utilised to guide prosthetics. Therefore, it is believed that EEG signals collected from human brains contain a lot of useful information that can be used to restore lost motor functions. It is believed that different motor imaginary movements can be differentiated by extracting the EEG features. Consequently, the number of control commands available in motor imagery (MI)-based machine learning (ML) is increased. Furthermore, the redundant information extracted from irrelevant channels can be optimized. The most appropriate and representative features with the most informative EEG electrodes remain. Moreover, feature fusion is believed to further reduce the data dimensionality and improve the classification accuracy. It is believed that the best combination of data acquisition, signal processing techniques, and classification techniques can be figured out.

1.4 Objectives

This study aims to investigate utilizing electroencephalography (EEG) signals as the input for machine learning (ML) algorithm in the neuro-prosthetics application. The following objectives are formulated to investigate and propose an EEG-based ML algorithm for neuro-prosthetics applications.

1. To discriminate between seven different motor imagery movements for the neuro-prosthetics application.

2. To reduce the EEG data dimensionality and enhance the performance of the machine learning (ML) models by defining appropriate and representative features from extracted features as biomarkers, selecting the most informative EEG electrodes, and applying multiple frequency

bands feature fusion technique in uni-modality for neuro-prosthetics application.

3. To validate the performance of the machine learning (ML) algorithm by using classification techniques.

The motor imagery task (MIT) is the task that imagines the movements of the human body. Biomarker is defined as a biological marker which is a measurable indicator of some biological states or conditions.

1.5 Research Question(s)

The possibility of adapting EEG signals and different types of motor imagery (MI) movements may be the important factors that affect the performance that can be achieved in machine learning (ML) algorithms and are yet to be investigated. There are specific questions addressed in this work:

1. What types of different movements performed by subjects can be differentiated? How can the different movements be performed?

2. What features can be extracted from the EEG measurements? What feature should be selected for neuro-prosthetics application? How can EEG-based feature extraction be technically realized for neuro-prosthetics applications? What is the optimal placement of electrodes required for neuro-prosthetics application? How many EEG electrodes must be

involved in neuro-prosthetics applications? What type of feature fusion techniques can be applied in uni-modality? How can 'fusion features' effectively reduce data dimensionality and further improve the prediction rate for neuro-prosthetics applications?

3. What classification technique is suitable for neuro-prosthetics applications? How does the performance of the machine learning (ML) algorithm?

1.6 Research Scope/Limitations

The research scope was divided into two parts namely data sources and processing tools. The research project of part 1 which is data sources focuses on UniMAP students who are in good mental condition with no neurological disorders as subjects. The subjects are ages in the range of 20 to 30 years old and right-handed. The protocol of the experiment was established to induce relaxation, motor execution task (MET), and motor imagery task (MIT) only. In part 2 which is processing tools, the EEGO^{TM} sports device (ANT Neuro, Enschede, The Netherlands) with 32 channels is an EEG measurement tool that was used to collect EEG signals. The MATLAB software is a mathematical modelling software that was used to formulate the machine learning (ML) algorithm. The details are explained in Chapter 3.

1.7 Organization of the Dissertation

This report contains five chapters. Chapter 1 is the introduction and Chapter 2 is the literature review. The methodology is explained in Chapter 3. Chapter 4 is about the results and discussions. Chapter 5 is the conclusions.

Chapter 1: Introduction – This chapter briefly introduces motor imagery (MI)-based machine learning (ML) algorithms for neuro-prosthetics applications and the motivations behind this research project. It gives a brief description of the project background. It briefly discusses the concept of signal processing. Next, problem statements describe the drawbacks in current research related to MI-based ML algorithms for neuro-prosthetics applications. Hypotheses discuss and explain the use of EEG in MI-based ML algorithms for neuro-prosthetics applications possibly can be done. It states the possible findings of this research project based on the problem statement. Objectives describe the aims and purposes of this research project. Research questions are the questions that going to be answered in this research project. The scope of this study describes the importance and limitations of the research project.

Chapter 2: Literature review – This chapter discusses conventional prosthetics which consists of cosmetic prosthetics and functional prosthetics, the history of EEG signals, brain anatomy, brain lobes, neurological disorders, Standardized Mini-Mental State Examination (SMMSE), EEG frequency rhythm, artefacts that are possibly found in the EEG signals, signal acquisition, signal processing, testing and validation stage. Signal processing contains pre-processing, feature extraction: power spectral density (PSD) Welch's method, features, feature selection, channel selection, and feature fusion. Moreover, this chapter reviews various significant research related to this research project. The gap in the existing knowledge has been found by reviewing other related research. The literature that was investigated includes comparing the methods and results in their works to produce good methodologies for the completion of this research project. The proposed works or conceptual solutions have been briefly discussed in this chapter.

Chapter 3: Methodology – This chapter explains in detail the flow of the research, issues of reliability and validity, data sources and data collection techniques, experimental paradigm, processing tools, data analysis and interpretation approach. Flow charts also being presented in this chapter. The data analysis and interpretation approach are discussed based on pre-processing, feature extraction, feature selection, dominance of selected features, channel selection, feature fusion, and classification.

Chapter 4: Results and Discussions – This chapter provides the results and related works for this research project. The results are analysed and discussed in detail according to initial accuracy, feature selection, dominance of selected features, channel selection, feature fusion, overall results, recall, and precision. The results have been compared with the related works.

Chapter 5: Conclusions – This chapter is to conclude the findings of this research project. The novel theories and contribution of this research project also being discussed in this chapter. Some recommendations are suggested in this chapter as well.

CHAPTER 2 : LITERATURE REVIEW

2.1 Introduction

Chapter 2 consists of sixteen subchapters. Subchapter 2.1 introduces Chapter 2. Subchapter 2.2 discusses conventional prosthetics such as cosmetic prosthetics and functional prosthetics including their limitations while Subchapter 2.3 narrates the history of EEG signals. Brain anatomy and brain lobes are discussed in Subchapter 2.4 and Subchapter 2.5 respectively. Neurological disorders and Standardized Mini-Mental State Examination (SMMSE) are discussed in Subchapter 2.6 and Subchapter 2.7 respectively. Then, the explanation of EEG frequency rhythm is presented in Subchapter 2.8. Subchapter 2.9 provides an overview of the artefacts. Next, Subchapter 2.10 introduces signal acquisition while Subchapter 2.11 introduces signal processing. Signal processing consists of pre-processing, feature extraction method which is power spectral density (PSD) Welch's method, features, feature selection, channel selection, and feature fusion. Subchapter 2.12 discusses the testing and validation stage. There are four sections under the testing and validation stage namely statistical test, classification, cross-validation, and confusion matrix. ANOVA one-way test is the statistical test that is being discussed while k-nearest neighbor (KNN), tree classifiers, ensemble classifiers, and support vector machine (SVM) are the classification methods that are being discussed. Subchapter 2.13 is the summary of the literature review and Subchapter 2.14 is the gap in the existing knowledge. The proposed works or conceptual solutions are stated in Subchapter 2.15. Lastly, Subchapter 2.16 summarises the Chapter 2.

2.2 Conventional Prosthetics

Conventional prosthetics are the traditional prosthetics that are intended to improve the life quality of the disabled individual. A custom prosthetic is an option that helps to regain functionality and mobility when someone experiences the loss of limbs (Horton, 2016). The term “peg leg” appeared before the twentieth century. Many people created prosthetics based on the materials they had such as table legs or wooden chairs. This was because they could not afford professionally made prosthetics. Due to advances in technology, design, and materials, prosthetic devices have greatly enhanced over time (Editorial of UPMC Health Beat & Rehabilitation, 2015). Prosthetics have advanced significantly from solely meeting aesthetic purposes to having unique abilities that result in regaining functionality that was previously lost due to trauma or amputation (Horton, 2018).

The earliest-known prosthetic toe was fashioned of leather and wood in 950-710 B.C. It was unearthed during the 1800s and found affixed to a mummy from Egypt. In 600 B.C., Egyptians created the Greville Chester toe which is made of cartonnage and was found in 2000 close to the location of present-day Luxor. The cartonnage is composed of paper mache which is created with glue, plaster, and linen. Romans crafted the Capua leg from bronze and iron with a wooden core in 300 B.C. It is the earliest prosthetic leg. It was previously housed in the Royal College of Surgeons. However, Capua leg was devastated during World War II bombardment. Now, a replica is on display at the London Science Museum (Editorial of UPMC Health Beat & Rehabilitation, 2015).

Peg legs and hand hooks were popular for those who could afford them during the Middle Ages which is from 476 to 1000. Prosthetics intended to fit into stirrups or grip a shield were frequently worn by knights. However, the functionality of the prostheses was not the main focus. During this period, the number of tradesmen-crafted prosthetics was increasing. More detailed functionality of the limbs was given by the springs and gears which were often used by those who made watches. Wood, iron, steel, and copper were the materials that the most frequently utilized in prosthetics during the 1400s-1800s which is the Renaissance (Editorial of UPMC Health Beat & Rehabilitation, 2015).

During the year 1863 which was the American Civil War, the United States started to have breakthroughs in the prosthetics industry. The aesthetically pleasing rubber hand, featuring movable fingers and several accessories such as hooks and brushes had been introduced. Following World War II in the year 1945, the majority of the limbs were composed of leather and wood. Although several benefits were provided by these materials to the wearer, the prosthetics were bulky and the leather was hard to maintain clean, especially since it absorbs perspiration. During the 1970s – 1990s, alternatives to leather and wood models were proposed. The materials were laminates, resins, polycarbonates, and plastics. These materials were light and easy to clean. Lightweight materials such as carbon fibre started being used to make prosthetics. Each patient received the custom-fit synthetic sockets to ensure a hygienic, personalized, and comfortable fit (Editorial of UPMC Health Beat & Rehabilitation, 2015).

From 2000 to 2014, advancements in prosthetic design led to the development of highly specialized prosthetics, such as motorized hand prosthetics that can be controlled using microprocessors and sensors, lightweight, high-performance running blades, and responsive feet and legs for navigating diverse terrain. Prosthetics have gone a long way and advanced significantly since the first documented wooden toe, largely due to the enhancement of materials and new technologies. Future developments in the field of prosthetics are facilitated by technological innovations including 3-D printing, robotics, and brain-computer interfaces (BCI). Researchers have studied BCI technology which sends signals to the prosthetic limb from the brain using surgically implanted electrodes (Editorial of UPMC Health Beat & Rehabilitation, 2015). However, this is a risky and painful procedure (Roy, 2022). Nowadays, non-invasive BCI technology has started to develop. Generally, there are two types of available commercial prosthetics devices namely cosmetic prosthetics and functional prosthetics (Thiamchoo & Phukpattaranont, 2022).

2.2.1 Cosmetic Prosthetics

Cosmetic prosthetic hands are primarily designed to approximate the look of biological hands to replace missing hands (Thiamchoo & Phukpattaranont, 2022). These cosmetic prosthetics are to compensate for body structure loss and achieve body balance (Widehammar et al., 2021). Cosmetic prosthetic hands are non-functional and passive (Horton, 2016). They are incapable of functioning as a natural hand (Thiamchoo & Phukpattaranont, 2022) due to the cosmetic prosthetic hands seldom focusing on functionality and mobility.

2.2.2 Functional Prosthetics

In contrast with cosmetic prosthetic hands, functional prosthetic hands enable limb amputation individuals to manipulate the prosthetic hands to perform daily activities independently (Thiamchoo & Phukpattaranont, 2022). This is to avoid the occurrence of compensatory overloading in the contralateral arm (Widehammar et al., 2021). Functional prosthetic devices are designed to partially restore upper limb functionality (Cowley et al., 2017). Functional prosthetic hands enhance the quality of life by controlling the device to complete varying motions. The movements that can be controlled are depending on the type of prosthetic hands. The functional prosthetic hands allow disabled individuals to engage in many activities like normal individuals (Horton, 2016). Figure 2.1 shows one of the functional prosthetic hands.

Figure 2.1 Functional prosthetic hand (Technology, 2020).

However, a longer time is required for a prosthesis user to complete movements. Uncoupling of grasping and reaching due to grasping is more prolonged than reaching. Furthermore, these reaching movements do not appear smooth. More discrete sub-movements are required to complete a task (Cowley et al., 2017). In addition, conventional upper-limb prosthetics are inconvenient for daily use due to their weight and lack of comfort. This causes fatigue or the dexterity is just as or more functional without using the upper-limb prosthetics (Biddiss & Chau, 2007; Piozin et al., 2022).

2.2.2.1 Myoelectric Prosthetics

Myoelectric prosthetics is one of the functional prosthetics that use the biological signals from the existing muscles to perform daily routine activities. It is indeed known that the users of traditional myoelectric prosthetics control the movements of the devices based on the use of surface electromyography (EMG) electrodes. The EMG can be recorded from the skin and it reflects the activities of muscle and peripheral nerves that are related to body movements directly (Thiamchoo & Phukpattaranont, 2022). The peripheral nerves contain both afferent and efferent fibres which allow bidirectional communication (Karczewski et al., 2021).

To train the prosthetic hands how to move, myoelectric prosthetic hands harness the power of muscles and nerves in the residual arm. With training, it learns the signals of the behaviour patterns that are meant to be accomplished from contracted muscles in the residual limb, then trained to react accordingly. Therefore, the signal detection components that are connected to the hand and the type of neuro-muscle needed largely depend on the amount of residual limb remaining. Below-elbow amputees rely on the wrist and forearm more heavily and the transmitters that are attached to the biceps and triceps are for above-elbow amputees (Horton, 2016).

However, these myoelectric prosthetics come with some limitations. Training the myoelectric prosthetic hands to react to various muscle signals required a significant training period (Horton, 2016). Constant stress is subjected to the stump on the arm of users while in use. Another limitation is the devices face difficulty in catching muscular activity properly due to the sweat from the prosthesis socket (Piozin et al., 2022). In addition, the cost of the commercially-available and certified myoelectric prosthetic devices is high ranging from US$4,000 to US$130,000 (Ku et al., 2019; K. H. Lee et al., 2017; Schacter, 2017; Setty et al., 2020), as well as the shortage of personnel and infrastructure (Geneva, 2017).

2.2.2.2 Brain-Computer Interface System for Neuro-Prosthetics

To help the disabled individual, an electroencephalogram (EEG)-based brain-computer interface (BCI) has been investigated (Bilal et al., 2022; E. Mohammadi et al., 2021). This is due to the BCI system showing immense potential applications in neuro-prosthetics, especially motor imagery (MI)-based BCI system (Venkatachalam et al., 2020). With the technology of BCI, conventional prosthetics have been upgraded to modern neuro-prosthetics. The development of this system over the past few decades has been incredibly rapid (Shedeed & Issa, 2016).

BCI system is capable of providing control and communication to paralyzed individuals or people who suffer from other severe mobility impairments (Shedeed & Issa, 2016). It converts a presentation of brain cell communication impulses into digital commands (Ghorbanzadeh et al., 2023). The MI-based BCI system serves as effective assistive technology to maintain or restore a lost motor function or communicate skills by imagining the movement of different body parts (Jiang et al., 2015). The BCI system can eschew conventional channels of communication (Millan et al., 2010) as Figure 2.2.

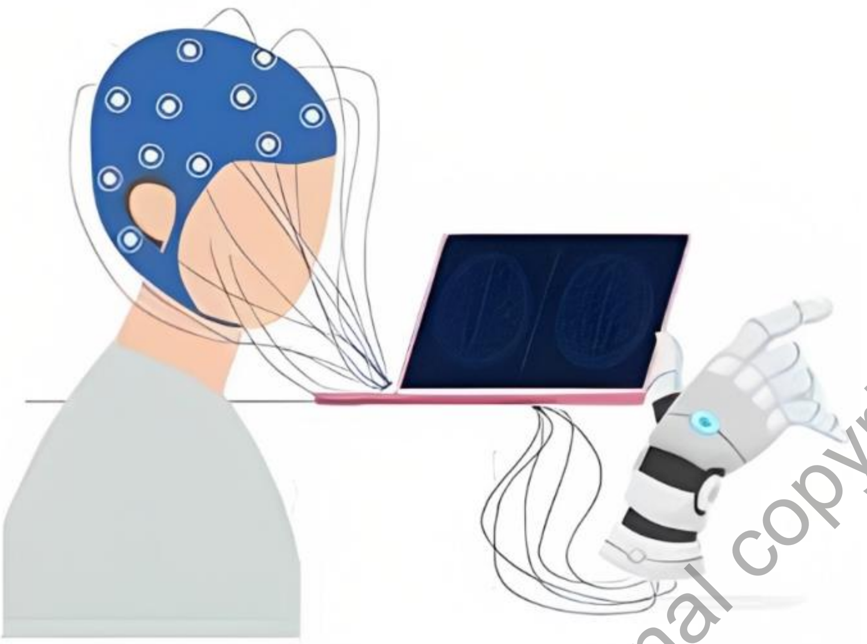

Figure 2.2 General idea of brain-computer interface (BCI) system (Mitra, 2018).

BCI system is a non-invasive system that records the data by using the sensors that are placed over the scalp rather than utilizing an invasive method to implant the sensors beneath the scalp. This is to prevent the high expense and the risk of surgery associated with planting (Shedeed & Issa, 2016). Neuro-prosthetics use EEG signals as input signals due to high time resolution and mobility potential in the user (Venkatachalam et al., 2020). A person with the ability to generate the mental processes that produce discernible EEGs at will may be able to utilize the EEGs to communicate. The viability of this communication relies on the degree to which the EEGs related to these mental processes can be recognized reliably and automatically (Alam & Samanta, 2017).

In an MI-based BCI system, the brain signals from corresponding brain activities are generated by the motor movement imagination without the actual moving of any limbs. These EEG signals can be collected, analysed, and characterized to convert into actionable information. These actionable signals are utilized to activate external devices without the usage of limb movement. The success of the BCI system depends on reliable and suitable signal processing methods for the noisy, non-stationary, low signal-to-noise ratio, and non-linear brain activity signals (Alam & Samanta, 2017).

Extraction of characteristics for efficient classification of brain activity and conversion into relevant actions are key components of the effectiveness of the BCI system (Alam & Samanta, 2017). The five main consecutive stages of a conventional BCI framework are signal acquisition, signal pre-processing, feature extraction, classification, and control interface (Chaudhry et al., 2022). However, these modern neuro-prosthetics are costly due to numerous features and channels. Numerous features and channels lead to the system requiring more advanced specifications and time to process. Consequently, these modern neuro-prosthetics are necessary to simplify by eliminating irrelevant features and channels.

2.3 History of EEG Signals

Encephalography has advanced dramatically for more than a century of history. In 1875, English physician Richard Caton observed the EEG from the exposed monkeys' and rabbits' brains. This led him to the conclusion that electrical currents existed in the brain. German neurologist Hans Berger used his ordinary radio equipment to amplify the brain's electrical activity measured on the human scalp in 1924. He declared that feeble electric currents produced within the brain can be recorded by being depicted graphically on a strip of paper without opening the cranium. He found out that the activity altered following the brain's functioning state like lack of oxygen, anaesthesia, sleep, and certain neural diseases like epilepsy (Khalifa et al., 2012). The first person who used an electroencephalogram to define the electric potentials of human brains was Hans Berger.

The foundations for many present applications of electroencephalography were laid by him. He suggested that when the subject's general status shifts from being relaxed to being alert, the activity of the brain changes consistently and recognisably. Adrian and Matthews validated the notion of "human brain waves" and discovered consistent oscillations between 10Hz and 12Hz in 1934. They termed it as "alpha rhythm" (Teplan, 2002).

2.4 Brain Anatomy

As far as organs go, the human brain is the most intricate (Roy, 2022). It consists of nerve cells and they are structured by axons, dendrites, and cell bodies (Dũng et al., 2016). Figure 2.3 shows the basic structure of neurons.

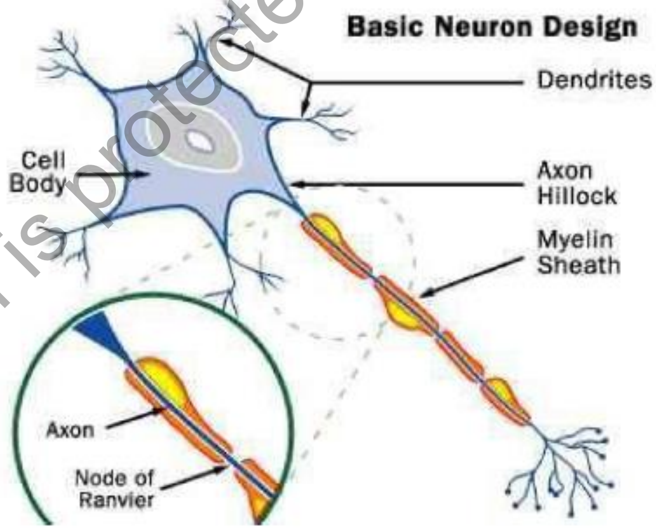

Figure 2.3 Basic structure of neurons (Khalifa et al., 2012).

EEG signals are generated from activities in neurons. The flows of local currents are produced in the brain when neurons are activated. Mostly, EEG measures the current flow when the dendrites of numerous pyramidal neurons in the cerebral cortex are synaptically excited. Cl-, Ca++, K+, and Na+ ions that are pumped through channels in