Cognitive Workload Measurement using Brain Computer Interface

Lieutenant Colonel Baljeet Singh Cheema

Computer Science and Engineering Department
Indian Institute of Technology
Kharagpur - 721302
April 2017

Cognitive Workload Measurement using Brain Computer Interface

Thesis submitted in partial fulfillment of the requirements for the degree of

Master of Technology

in

Computer Science and Engineering

by

Lieutenant Colonel Baljeet Singh Cheema
[15IT60D01]

under the guidance of

Dr. Debasis Samanta

Computer Science and Engineering Department
Indian Institute of Technology, Kharagpur
Kharagpur - 721302
April 2017

Declaration

I, Lieutenant Colonel Baljeet Singh Cheema, Roll no. 15IT60D01 registered as a student of M.Tech. in the Computer Science and Engineering Department, Indian Institute of Technology, Kharagpur, India, do here by declare that:-

a. The work contained in the thesis is original and has been done by myself under the general supervision of my supervisor.

b. The work has not been submitted to any other Institute for any degree or diploma.

c. I have followed the guidelines provided by the Institute in writing the thesis.

d. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

e. Whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.

f. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Dated:

[Lieutenant Colonel Baljeet Singh Cheema]
Roll No. 15IT60D01
Computer Science and Engineering Department
Indian Institute of Technology, Kharagpur.

Certificate

This is to certify that this thesis entitled **Cognitive Workload Measurement using Brain Computer Interface**, submitted by **Lieutenant Colonel Baljeet Singh Cheema** to Indian Institute of Technology, Kharagpur, is a record of bona fide research work carried under my supervision and I consider it worthy of consideration for award of the degree of Master of Technology of the Institute.

Dated:

[Dr. Debasis Samanta]
Associate Professor
Computer Science and Engineering Department
Indian Institute of Technology, Kharagpur

Acknowledgement

Foremost, I would like to place on record my sincere gratitude to my advisor Dr. Debasis Samanta for his invaluable guidance, constant encouragement and mindful attention during the project work giving me extraordinary experiences through out the work. It is indeed an honor and a great privilege for me to have learned from his expertise and experiences which exceptionally inspire and enrich my growth as a student and a researcher want to be.

I owe my deepest gratitude to all the faculty members of Computer Science and Engineering Department for providing an excellent academic environment and support.

I must also convey my heartfelt thanks to the ever diligent staff of Computer Science and Engineering Department for providing support throughout.

This thesis would not have been possible without the blessings of my parents and love of my beloved wife and son, who have been backing me up throughout my life.

Last but not the least, I am indebted to all my classmates for the motivation and support extended throughout my work.

[Lieutenant Colonel Baljeet Singh Cheema]
Computer Science and Engineering Department
Indian Institute of Technology, Kharagpur
Date: ...../...../2017.

Abstract

Cognitive workload is a demand placed upon human for mental resources while performing a task. Cognitive Workload Measurement (CWM) refers to the task of estimating cognitive workload levels. CWM has application in various fields like aviation, driving, critical military or space mission, safety at work and wellbeing of human resources.

Traditionally, various subjective (NASA TLX, etc.) and performance measurement techniques have been used to assess the cognitive workload and are offline. Also, the subjectivity factor leads to incorrect measurements. On contrary, a non-invasive and online physiological measurement technique called Electroencephalography (EEG) can provide continuous record of cognitive workload. This thesis explores the feasibility a very famous, cost effective and commercially available wireless EEG device, EMOTIV Epoc+ to measure various cognitive workload levels with the help of *working memory* tasks like n-Back and Dual n-Back. Ten healthy male and female subjects volunteered to undergo various levels of *working memory* task depicting increasing workload.

A very novel "Fully Online and automated artifact Removal for brain Computer interfacing" (FORCe) algorithm was used to remove the artifacts captured during EEG data recording. Channel selection based on "Mutual Information" (MI) was done followed by *feature extraction* of EEG signal to generate *feature vector*. For further simplification of model, *feature optimization* was done with the help of "minimum Redundancy Maximum Relevance" (mRMR) algorithm.

A comparative study was carried out by application of various supervised "Machine Learning" (ML) algorithms like k-NN, Random Forest, SVM, Decision Tree, LDA and MLP to the set of labeled *feature vectors* . The proposed CWM technique has given results with accuracy ranging from 85% to 100% for various combinations of *two, three, four* and *five* classes of cognitive workload generated.

Keywords: Cognitive workload, Working memory, Channel selection, Feature extraction, Feature vector, Feature optimization, Machine learning.

Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

List of Symbols

$A_{m \times n}$	Matrix A of m rows, n columns
σ_x	Standard deviation of the signal
$\sigma_{x'}$	Standard deviation of 1^{st} derivative of the signal
$\sigma_{x''}$	Standard deviation of 2^{nd} derivative of the signal

List of Abbreviations

CWM	Cognitive Workload Measurement
BCI	Brain Computer Interface
EEG	Electroencephalography
TLX	Task Load Index
MRT	Multiple Resource Theory
FFT	Fast Fourier Transform
ICA	Independent Component Analysis
CAR	Common Average Referencing
SL	Surface Laplacian
PCA	Principal Component Analysis
CSP	Common Spatial Patterns
MEG	Magnetoencephalography
ECoG	Electrocorticography

fMRI	Functional Magnetic Resonance Imaging
fNIRS	Functional Near Infrared Spectroscopy
EMG	Electromyogram
WT	Wavelet Transform
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
FD	Fractal Dimension
AR	Autoregression
k-NN	k-Nearest Neighbors
SVM	Support Vector Machine
MLP	Multi Layer Perceptron
LDA	Linear Discriminant Analysis
ANN	Artificial Neural Network
DL	Deep Learning
CNN	Convolutional Neural Networks
RNN	Recurrent Neural Networks

Chapter 1

Introduction

Brain is the most complex part in entire human body. Researchers and doctors have dedicated considerable amount of effort in decoding human brain. However, only a small percentage of brain functioning theory has been established so far and still, there is a long way to go. In today's world, where technology has taken over every aspect of our day-to-day life, study of mental/cognitive workload, its source of generation, measurement, etc. is becoming very important to build applications based on them and for betterment of human race. Cognitive workload measurement (CWM) has great role to play in various fields like aviation, driving, critical military or space mission, safety at work and wellbeing of human resources. This chapter establishes the context of mental/cognitive workload and its measurement using Brain Computer Interface (BCI).

1.1 Context

Our research involves measuring human cognitive workload, which is traditionally a neurologist's or psychologist's field of expertise, but, we present a BCI solution for doing the same. To bridge the gap between the two, we feel that it is our moral duty to understand the neurologist's or psychologist's way of looking at things as regard working of human cognition is concerned to bring best out of this thesis. Hence, we present a brief theory of human brain and its functioning in subsequent sub-sections before going into details.

1.1.1 Human Brain

Brain is the most amazing and complex part of human body. An average human brain consists of approximately 100 billion neurons weighing about 1.5 kilograms and interconnected via axons and neurons. A neuron has cell body called soma, a long axon and many dendrites as depicted in the Fig 1.1.

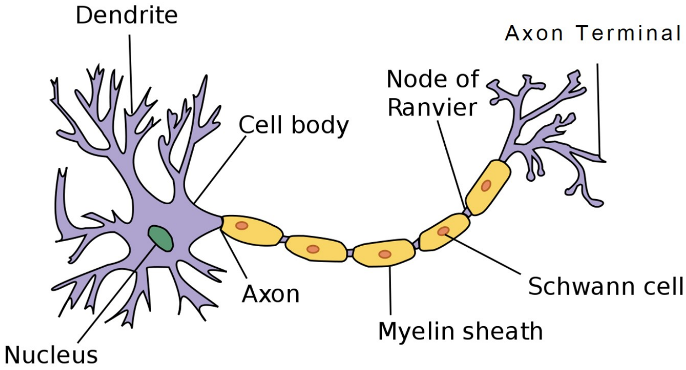


Figure 1.1: Neuron anatomy.

Neurons receive stimulus from other interconnected neurons about 103 to 105, through synapses. These stimulus travel through axon as electrical impulses and helps controlling body movements, emotions and other aspects of body coordination.

Parts of Human Brain

Our brain is responsible for cognition, sensory activity and controlling other voluntary and involuntary actions of the body. Different parts of the brain are responsible for specific tasks, hence, our brain can be classified into various parts as follows, based on the functions [1]:-

Cerebrum

The cerebrum or cortex is the largest part of the human brain, associated with higher brain function such as thought and action. It is further divided into four parts called "lobes" of brain:-

- ***Frontal Lobe*** is associated with reasoning, planning, parts of speech, movement, emotions, and problem solving. This part of the brain is responsible for cognition. The neural activity raises in this region of brain due to *cognitive workload.*
- ***Parietal Lobe*** is associated with orientation, movement, recognition, perception of stimuli.
- **Occipital Lobe** is associated with visual processing.
- **Temporal Lobe** is associated with perception and recognition of auditory stimuli, speech and memory.

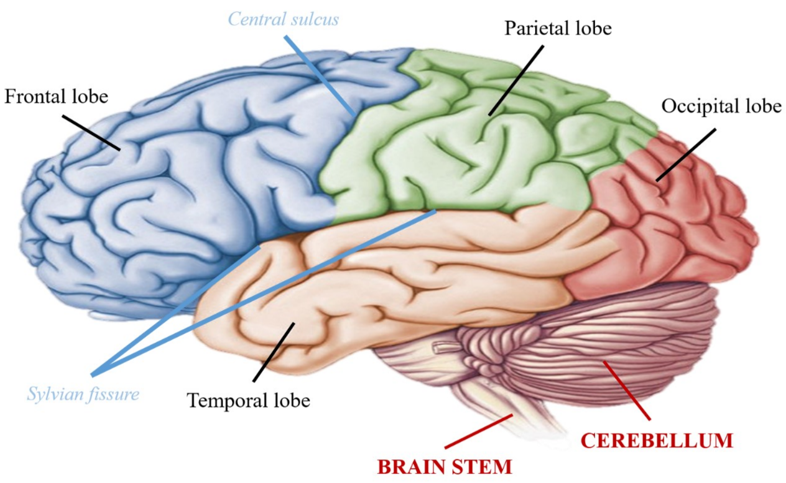


Figure 1.2: Human brain.

Cerebellum

The "emotional brain" also called limbic system, is found within the cerebrum. This system contains Hypothalamus, Thalamus, Hippocampus and Amygdala.

Brain Stem

Brain stem is found under the limbic system. It is responsible for basic important life functions such as heartbeat, breathing and blood pressure. This system contains Mid-brain Pons and Medulla.

1.1.2 Mental/Cognitive Workload

Mental workload is a function of both the cognitive demands associated with a task and the mental processing ability of the person performing the task (Parasuraman et al., 2008). We shall now discuss *working memory*, *mental workload* and *cognitive load* in detail.

Working Memory

Working memory is responsible for temporarily holding information available for processing. Working memory is important for decision making and reasoning. It is also known as short-term memory, but some theorists consider them to be distinct. They argued that working memory allows manipulation of information that is stored, whereas short-term memory only stores the information for short-term. Working memory is concept used in cognitive psychology, neuro-psychology and neuro-science [2].

Mental Workload

Mental workload is commonly described as the demand placed by a task for mental resources on a human subject. Mental workload measurement refers to the task of estimating the level of mental work load of a person in a particular situation from his/her physiological signals.

Mental workload is a neuro-ergonomics concept introduced in the 1940s. Researchers have attempted to define mental workload in different ways but no single definition is complete in itself. A commonly used definition of mental workload is: "The demand placed on humans for mental resources while performing a task". Mental Workload is an ISO 10075-1 standard and is defined therein as: "The total of all assessable influences impinging upon a human being from external sources and affecting it mentally". According to Jung and Jung (2001), the concept of mental workload is associated with the difference between the amount of resources available to a person and the amount of resources needed for the task. At the point when task demands exceed operator capacity, high levels of mental workload occur (Loft et al., 2007)[3]. According to Multiple Resource Theory (MRT), the human mind allocates several resources to different task demands like visual, auditory, cognitive, motor and speech either individually or collectively in order to optimally handle mental workload.

Cognitive Load

The work for cognitive load theory is marked with the beginning of Cognitive Science in the 1950s and the work of G.A. Miller. Cognitive load theory was proposed by John Sweller in the late 1980s. Miller was first to propose that our working memory capacity has inherent limits. His experimental results calculated that humans are normally able to hold only *"seven plus or minus two"* units of information in short-term memory. Cognitive load is the total amount of mental effort being used in the *working memory.* Sweller's cognitive load theory suggested that high mental workload would require the individual to allocate extra resources and may reduce processing efficiency and performance. Cognitive load is of three types (Paas et al., 2003a) [4] and (Sweller et al., 1998) [5].

- ***Intrinsic cognitive load***, is a load induced by the inherent nature of the items being processed, like task difficulty and would thus be fixed and inherent to the task.

- ***Extraneous cognitive load***, is induced by external factors, like time pressure, noise, situation, work organization, etc.

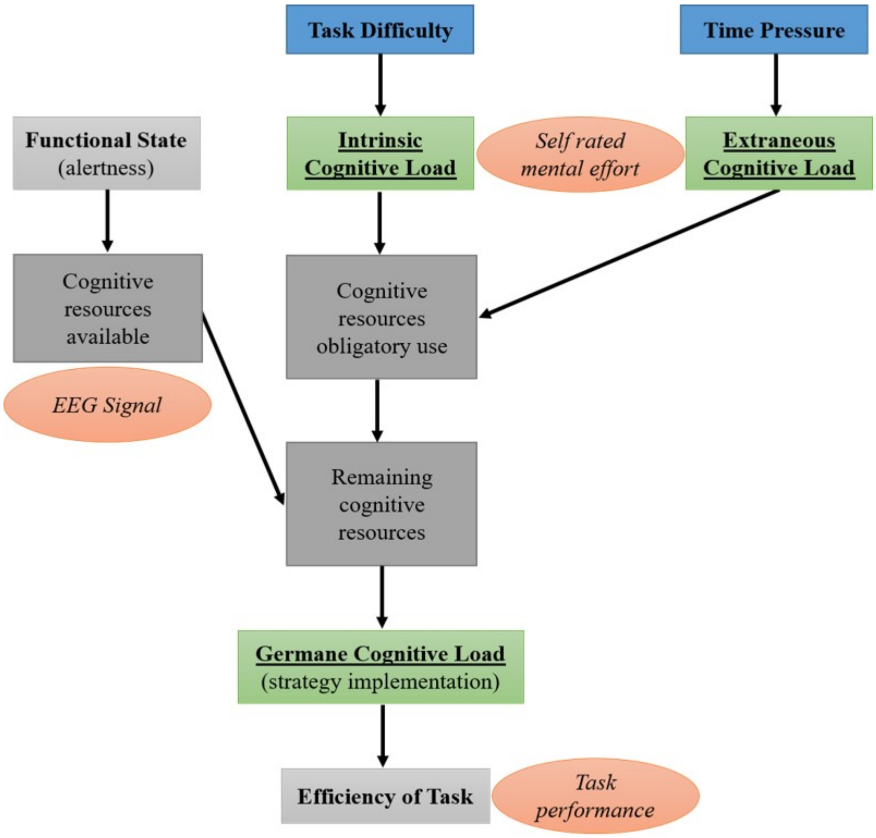


Figure 1.3: Cognitive load factors and cognitive load categories.

- ***Germane cognitive load***, is the load placed on working memory during schema formation and automation. More recently, Schnotz and Kurschner (2007) [6] proposed that germane load would correspond to the "conscious application of learning strategies (i.e. strategies, which are not automated yet), conscious search for patterns in the learning material in order to deliberately abstract cognitive schemata (i.e. mindful abstraction) and create semantic macrostructures, restructuring of problem representations in order to solve a task more easily (i.e. by insight), meta-cognitive processes that monitor cognition and learning".

According to cognitive load theory, intrinsic, extraneous, and germane cognitive loads are additive, in that the total load must not exceed available working memory resources if the task is to be completed. Relation between cognitive loadfactors and cognitive load categories is depicted in Fig 1.3 [7]. Details of how all three types of cognitive load has been manipulated or tweaked in this thesis would be discussed in Chapter 3, while discussing the design of experiment.

1.2 State of the Art

The tools/techniques used to measure cognitive load can be divided into three main categories as follows [7].

1.2.1 Subjective Measures

NASA-Task Load Index (NASA-TLX; Hart and Staveland, 1988) [8] is a subjective measurement technique which includes six subscales exploring the Mental Demand, Physical Demand, Temporal Demand, Own Performance, Effort, and Frustration Level. Subjective Workload Assessment technique (SWAT; Reid and Nygren, 1988) [9] describes three dimensions of operator workload: Time Load, Mental Effort Load and Psychological Stress Load.

1.2.2 Performance Measures

Within this group of methods, the participants' mental workload is inferred from their overt behaviour or performance, in particular response accuracy and response latency. Chi and Lin (1997) [10] demonstrated a trade-off between these performances criteria, as the time needed to complete a task increased when accuracy requirements increased, whereas a decrease in accuracy occurred when task rapidity requirements increased.

1.2.3 Physiological Measures

The principal attractions of psycho-physiological measures are continual and objective measurement of operator state. Psycho-physiology attempts to interpret the psychological processes through their effect on the body state, rather than through task performance or perceptual ratings. Few invasive, semi-invasive and non-invasive physiological techniques are illustrated below:-

Magnetoencephalography (MEG)

MEG detects the magnetic fields resulting from the electrical currents in neurons. These magnetic fields are orthogonal to the electric signals measured by EEG. The magnetic field are less distorted, and hence MEG provides better spatial and temporal resolution. MEG, however requires typically expensive and much sensitive devices. Also, the measurements are required to be taken at magnetically shielded rooms. This approach measures only shallow parts of brain and is too bulky to be suitable for everyday use.

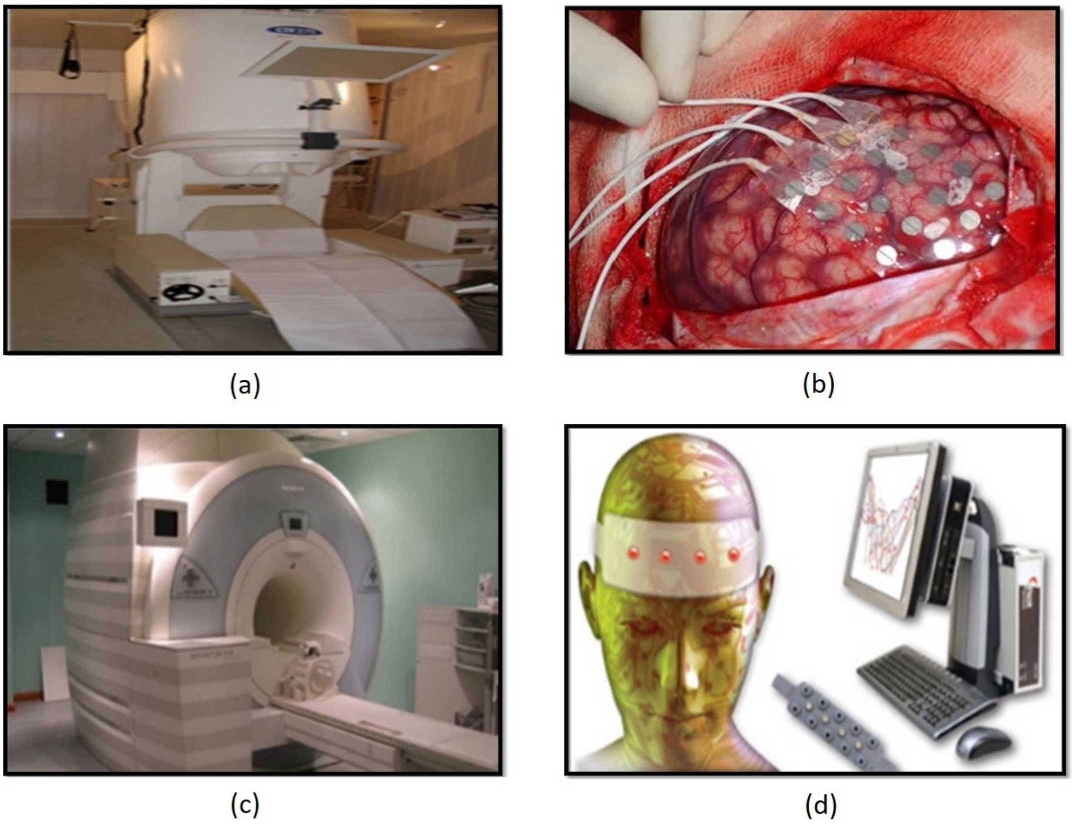


Figure 1.4: (a) MEG, (b) ECoG, (c) fMRI, (d) fNIRS.

Electrocorticography (ECoG)

In ECoG, the electrodes are placed under the dura matter, directly on the surface of the cortex, without penetrating the cortex. ECoG provides better spatial, temporal resolution and better signal quality. These signals have higher amplitudes and are less vulnerable to the artifacts such as eye-blinks and eye movements. However, this is a semi-invasive technique, which requires risky surgery. ECoG are primarily used in experiments with animals. ECoG implants remains stable for several months and can be used to record signals, however long term stability remains unclear to date.

Functional Magnetic Resonance Imaging (fMRI)

fMRI is a non-invasive technique that relies on hemodynamics. fMRIs determine the blood oxygen level variations and thus provides high spatial resolution. They suffer from poor time resolution, and is susceptible to head motion artifacts. fMRIs like MEG are very expensive equipment and are not suited for individual and everyday applications.

Functional Near Infrared Spectroscopy (fNIRS)

fNIRS is also a non-invasive acquisition technique. It measures the changes in optical response of cerebral tissues to near-infrared light due to variations of haemoglobin concentrations. Infrared light can penetrate only small depth of skull, and hence provides shallow spatial resolution of the order of the centimetre while the time resolution is of around 200 ms. fNRIS are relatively inexpensive, portable and are gaining popularity for everyday use applications.

Temporal resolution of EEG is high and is in the order of milliseconds. This makes EEG an appropriate tool to capture fast and dynamically changing brainwave patterns in complex cognitive tasks. EEG signals have been used to detect changes in mental workload on computer based tasks [11].

Electroencephalography (EEG)

Electroencephalography (EEG) is an electro-physiological monitoring method to record electrical activity of the brain. German physiologist and psychiatrist Hans Berger (1873-1941) recorded the first human EEG signal in 1924. EEG is used extensively in neuro-science, cognitive science, cognitive psychology, neuro-linguistics and psycho-physiological research [12].

EEG wave groups

The analysis of continuous EEG signals or brain waves is complex, due to the large amount of information received from every electrode. Five different types of EEG waves are as discussed below:-

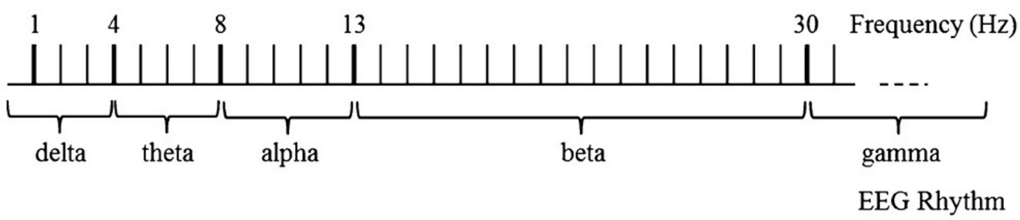


Figure 1.5: EEG frequency bands.

Delta Wave

Delta waves lie within the range of $0.1 - 4$ Hz, with variable amplitude. Delta waves are primarily associated with deep sleep, and in the waking state, were thought to indicate physical defects in the brain. It is very easy to confuse artifact signals caused by the large muscles of the neck and jaw with the genuine delta responses. This is because the muscles are near the surface of skin and produce large signals whereas the signal which is of interest originates deep in the brain and is severely attenuated in passing through the skull.

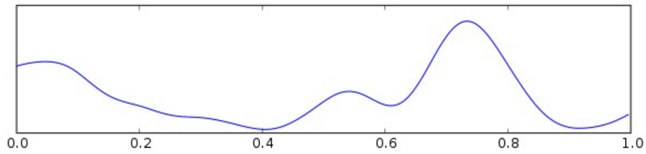


Figure 1.6: Delta wave.

Theta Wave

Theta waves lie within the range of $4-8$ Hz, with an amplitude usually greater than 20 μV. Theta arises from emotional stress, especially frustration or disappointment. Theta has been also associated with access to unconscious material, creative inspiration and deep meditation. The large dominant peak of the theta waves is around 7 Hz.

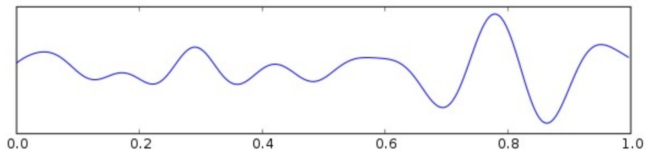


Figure 1.7: Theta wave.

Alpha Wave

The rate of change lies between $8-13$ Hz, with $30-50$ μV amplitude. Alpha waves have been thought to indicate both a relaxed awareness and also inattention. They are strongest over the occipital (back of the head) cortex and also over frontal cortex.

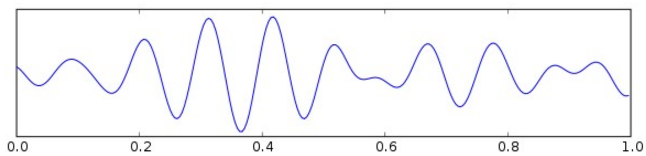


Figure 1.8: Alpha wave.

Alpha is the most prominent wave in the whole realm of brain activity and possibly covers a greater range than has been previously thought of. It is frequent to see a peak in the beta range as high as 20 Hz, which has the characteristics of an alpha state rather than a beta, and the setting in which such a response appears also leads to the same conclusion. Alpha alone seems to indicate an empty mind rather than a relaxed one, a mindless state rather than a passive one, and can be reduced or eliminated by opening the eyes, by hearing unfamiliar sounds, or by anxiety or mental concentration.

Beta Wave

The rate of change lies between 13−30 Hz, and usually has a low voltage between 5−30 μV. Beta is the brain wave usually associated with active thinking, active attention, focus on the outside world or solving concrete problems. It can reach frequencies near 50 Hz during intense mental activity.

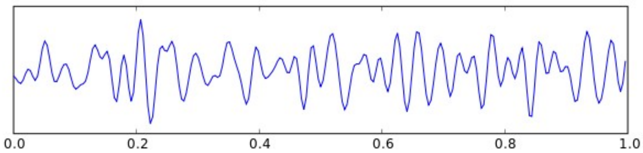


Figure 1.9: Beta wave.

Gamma Wave

Gamma waves lie within the range of 30 Hz and up. It is thought that this band reflects the mechanism of consciousness - the binding together of distinct modular brain functions into coherent precepts capable of behaving in a re-entrant fashion (feeding back on themselves over time to create a sense of stream-of-consciousness).

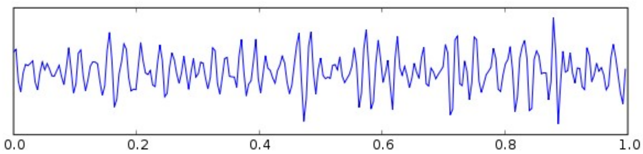


Figure 1.10: Gamma wave.

It is observed that mental/cognitive workload increase is consistently correlated with an increase in the frontal theta $4-8$ Hz and beta $13-30$ Hz band power and parietal alpha $8-13$ Hz band power.

Theta waves are related to the deepest state of mediation (body asleep/mind awake). Alpha waves are related to the case of dreaming and relaxation. Beta waves are the dominant with the waking state with large attention. Gamma waves are highly related to the decision-making mode of the brain. Gevins et al. reported that increased memory load was associated with increased theta band power in the frontal midline area. In addition to the theta-band activity changes, most studies also observed alpha band activity changes. Gevins et al. [13] reported that alpha signal power decreased with increased working memory load in the parieto-occipital midline areas. Several aforementioned studies that investigated classification of EEG signals distinguished discrete levels of working memory load.

1.3 Brain Computer Interface

A BCI, sometimes called a mind machine interface (MMI), direct neural interface (DNI), or brain machine interface (BMI), is a direct communication pathway between an enhanced or wired brain and an external device [14]. BCIs are often directed at researching, mapping, assisting, augmenting, or repairing human cognitive or sensory-motor functions. BCI system converts the electrophysiological signals generated by various mental/cognitive activities to device specific messages or commands. Based on the location of the sensors used to record the aforementioned signals, BCI systems can be categorized into:-

- ***Non-invasive***, when the sensors are placed on the scalp, e.g. Electroencephalography (EEG), Magnetoencephalography (MEG).
- ***Semi-invasive***, when the electrodes are placed on the exposed surface of the brain, e.g. Electrocorticography (ECoG).
- ***Invasive***, when micro-electrode arrays are placed directly into the cortex of human brain.

Non-invasive systems, records the electrical activity, the magnetic activity or hemodynamics response due to neural activity in brain. Non-invasive BCI system are now reliable enough, relatively inexpensive, and are primarily used for BCI research. All the methods based on the EEG devices have to address the inherent challenges, like its poor spatial resolution and low signal to noise ratio. MEG senses the magnetic potential formed due the activity of neurons. It is orthogonal to the electrical activity recorded using EEG devices, and gives higher spatio-temporal resolution. However, due to need for sensitive sensors and magnetically shielding, they have limited research potential.

Semi-invasive systems uses electrocorticography (ECoG) and due to it being located nearer to the site of neutral activity, provides better spatial resolution and higher signal to noise ratio. ECoG relies on the same neuro-physiologic mechanisms as EEG, and requires similar signal processing approaches.

Invasive techniques provides a very high spatial resolution, temporal resolution and signal to noise ratio. The recordings differ significantly from that of noninvasive devices and hence require different signal processing approach.

BCI System

BCI system for capturing EEG signal and then measuring CWM is divided into two phases in this thesis:-

- **Offline BCI CWM Model** - This model would be generated offline. The processes involved in generating this model is clearly described in the Fig 1.11.

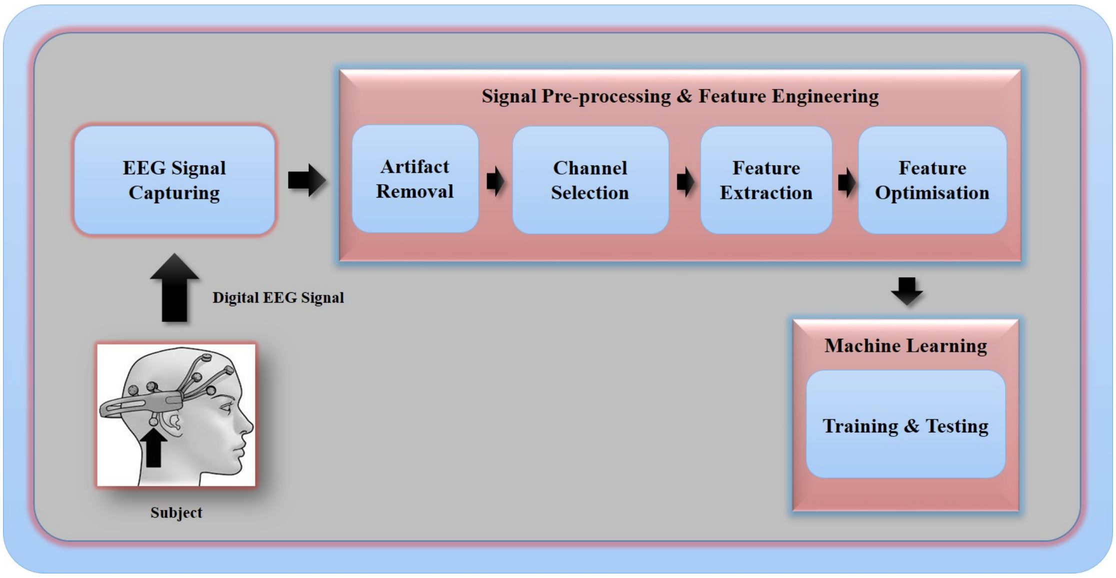


Figure 1.11: CWM model.

- **Online BCI CWM System** - This is an online system. The sub modules of this system are shown in the Fig 1.12. Each submodule in Fig 1.11 and Fig 1.12 would be explained in detail in Chapters to follow.

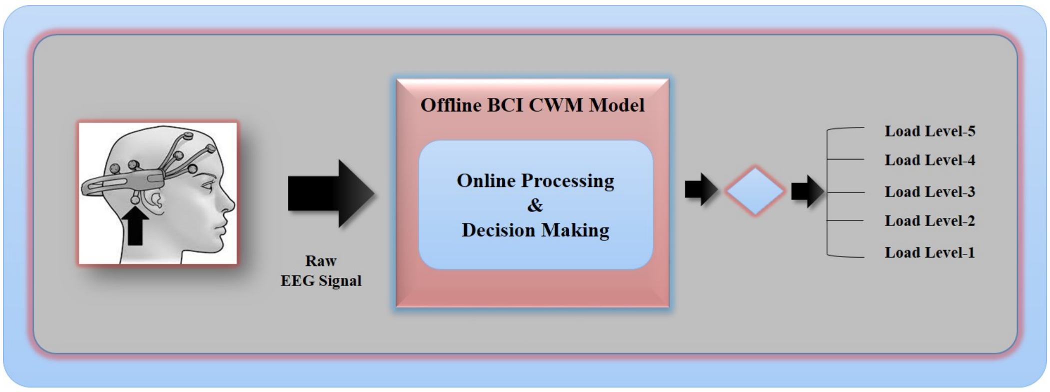


Figure 1.12: CWM system.

1.4 Issues and Challenges

The major issues and challenges with BCI data is BCI signal capturing, processing, hardware issue, algorithms (usually ML algorithms) to discover knowledge from the brain signals and then translating signals to decision (e.g. prediction, evaluation, measurement, etc.). Few immediate issues and challenges are mentioned below:-

- EEG capturing electrodes should be highly efficient in order to bring BCI systems outside laboratory conditions.
- For mental/cognitive state monitoring applications, the BCI should be asynchronous and continuous.
- BCI signal processing algorithms must be robust to external noise, as electrodes will most probably be unable to suppress all kinds of noises.
- Reducing or suppressing the BCI calibration time is essential to enable immediate use of the BCI.
- High dimensionality curse to be handled for faster and accurate results.
- Explore and design various feature extraction and classification algorithms to decode mental/cognitive states.
- Integrate BCIs with other modalities should also be made as simple and seamless as possible.

1.5 Motivation and Scope of Work

Research in the field of BCI using EEG signals has recently taken a boom. Availability of cheap non-invasive EEG device for collecting physiological brain signals has opened path for open research in this field. However very little has been achieved as regard cognitive workload measurement and its application in various areas like aviation, driving, military/space missions. Also, cognitive workload measurement applications has big scope in Indian Armed Forces and hence the motivation of our research.

As discussed earlier, subjective and performance measurement techniques suffer from subjectivity issue which introduces bias in these assessment techniques leading to generation of incorrect model. To overcome this issue, we can use a very carefully designed physiological measurement techniques such as EEG, that can provide continuous record of cognitive workload.

Various artifacts can cause noise in such physiological signal that can again lead to generation of incorrect model. To get rid of artifacts many novel algorithm have been proposed by researchers which can be used to suit our requirements. Curse of dimensionality is a major issue especially when one works with applications which are online. To reduce the dimensionality of data various channel selection techniques can be used.

Various feature extraction techniques can be applied to generate feature vector. To reduce the computation overload and avoid overfitting, identification of relevant effective features could be done with the help of various efficient algorithm available which are based on correlation/mutual information between variables, reducing the dimensionality of feature vector.

To classify input EEG signal, various ML algorithms can be applied to the set of feature vectors and the model would be trained using these feature vectors after dividing it into training-set and test-set data. A comparative study can be done for various famous classification algorithms and results can be compared to establish facts.

1.6 Research Objective

Objective of this thesis is to measure cognitive workload by BCI technique using non-invasive EEG signals. Following steps would be involved in achieving this:-

1. **Design of Experiment**
   Choosing correct experiment for generation of EEG data depicting various levels of cognitive workload is an important issue. An experiment that satisfies all the parameters of cognitive load should be satisfied to ensure the authentication of data collected.

2. **Feature Engineering**

   (a) ***Signal Preprocessing***
   EEG data is generally contaminated with noise. Removal of various artifact is not only difficult but is lossy, as it may accidentally remove some features which could be otherwise important for calculation of Mental Workload. A very carefully designed algorithm should be considered for removing artifacts.

   (b) ***Channel Selection***
   Our aim is to find a robust channel selection algorithm for enabling faster calculation and generating accurate model.

(c) ***Feature Extraction***
Feature extraction is more of an 'Art' than just a 'Technique'. It is important to study and calculate various features and evaluate the results before deciding to make it a part of model.

(d) ***Feature Optimization***
To reduce the dimensionality of data to make the system real-time, it is important to select important features only. A very sound algorithm has to be considered for feature optimization without loss of important features accidentally.

3. **Modelling for Cognitive Workload Measurement**
In ML, various classifiers are available for final model generation. Our aim is to find the best classifier suitable for classification of cognitive workload data with high accuracy.

1.7 Organization of the Thesis

Chapter 1, the introductory chapter, establishes the context of CWM using BCI in today's scenario. It also highlights the current practices and state of the art in this field. Further, issues and challenges, motivation and scope of work and challenging objectives of this research are discussed. We conclude this chapter by presenting the organization of this thesis.

Relevant survey of existing work in the field of this thesis is discusses in **Chapter 2**, also highlighting the major contributions by the researchers along with the limitations.

In **Chapter 3**, we discuss the design of experiment along with justification for the same. Various experiments considered are discussed in detail along with the relevance to the cognitive workload theory.

In **Chapter 4**, details of signal preprocessing and feature engineering aspects like, channel selection, feature extraction, feature optimization are discussed along with the details of the algorithm used to achieve the same.

In **Chapter 5**, we implement various classification algorithm for CWM model generation.

In **Chapter 6**, we conclude the thesis by discussing the exhaustive results thus obtained after testing the model generated in Chapter 5, along with discussing the research contribution and future work.

Chapter 2

Literature Survey

Literature survey is an important step before proceeding with the details of thesis. Aim of literature survey is to understand the present techniques being used in the field of CWM. In this chapter we do a brief survey related to the objectives of Chapter 3, 4 and 5.

2.1 Survey for Design of Experiment

Design of experiment is a very challenging job in any research. In subsections to follow, we highlight few important tasks/tests being used by researchers to generate cognitive workload.

2.1.1 Flanker Test

In cognitive psychology, the Eriksen Flanker Task is a set of response inhibition tests used to assess the ability to suppress responses that are inappropriate in a particular context. The target is flanked by non-target stimuli which correspond either to the same directional response as the target (congruent flankers), to the opposite response (incongruent flankers), or to neither (neutral flankers). The task is named after Barbara. A. Eriksen and Charles W. Eriksen. In this tests, a directional response (usually left or right) is assigned to a central target stimulus. There are three types of stimuli used [15] in an Eriksen Flanker task, viz. congruent stimulus, incongruent stimulus and neutral stimulus.

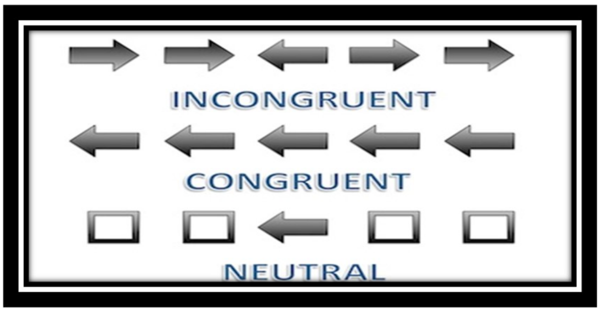


Figure 2.1: Flanker test.

The speed and varying size of stimuli presented, generates mental load on the subject. The problem for considering "flanker test" in our thesis is that the subject always knows the target position and hence doesn't get disturbed much by flank stimuli, which fails to produce the actual cognitive load we are looking for. Flanker test has been used in [16].

2.1.2 Arithmetic Test

In arithmetic test the subject is asked to perform arithmetic computations without help of any external means of recording the calculations. Here we increase the cognitive load by increasing the difficulty in the task by increasing speed, the number of operations or digits. However, someone who is good in arithmetic calculations shall perform better than others, making the entire experiment biased due to subjectivity factor. Farshid Anvari et al. [17] used this task to calculate cognitive load.

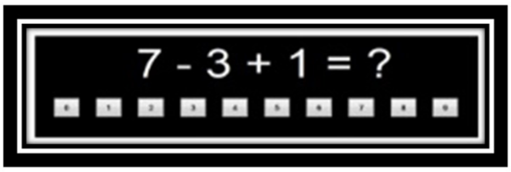


Figure 2.2: Arithmetic test.

2.1.3 Code Debugging Task

Code debugging involves a very demanding cognitive process. Researchers have attempted to present code debugging for generating cognitive load. In this method, the subject is presented with a piece of code for debugging. The task difficulty is increased by putting more bugs and introducing cryptic code. However, this method also suffers from the subjectivity issue due to skill levels of the subject participating in the task.

Figure 2.3: Code debugging test.

2.1.4 Stroop Test

In stroop effect the name of a color ("pink", "red", "blue") is printed on a computer screen in a color that is not denoted by the name for e.g., the word "orange" is printed in pink ink instead of orange ink. Here, it is difficult to name the color of the word as it takes longer and is more prone to errors as comparted to the case when the color of the ink is same as the name of the color. The effect is named after John Ridley Stroop, 1935.

Problem here with stroop test is that it generates so many eye artifacts that it would lead to noisy data. Though we shall use artifact removing algorithm, but our aim is just to mark a humble step in recording the cognitive workload measurement. Also, after learning the schema properly, subjects can play this task easily, thus defeating the main aim of generating varying levels of cognitive workload. Gwizdka J., 2010 used this task for mental workload assessment [18].

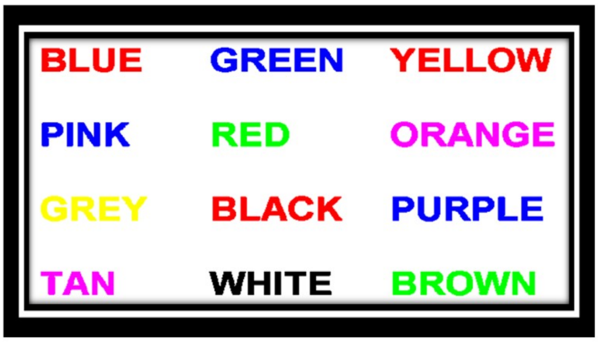


Figure 2.4: Stroop test.

2.1.5 n-Back Task

The n-back is a task that is commonly used as an assessment in cognitive neuro-science to measure a part of working memory and working memory capacity [19].The n-Back task involves remembering the sequence of alphabets and has been used widely used for generating cognitive workload[20] [21].

2.2 Survey for Signal Preprocessing and Feature Engineering

In this section we shall discuss various options available for signal preprocessing and feature engineering.

2.2.1 Signal Preprocessing

Signal preprocessing is very important as it removes any external disturbances and artifacts that may interfere with the EEG signals required for processing and feature extraction. Eye movement or eye blinking, any muscular activity, ECG and EMG are some of the common artifacts present in the recorded EEG signals. Basically, EEG signals are low amplitude signals with a very poor signal to noise ratio. There are various methods designed by biomedical researchers. Various filtering techniques have been employed for the removal of artifacts. The frequency range of the artifacts are usually undefined. Therefore the use of basic filters for artifact removal is insufficient. Researchers have used adaptive filtering techniques that adapts to the change in frequency by a change in the coefficients of the linear filter to generate a signal close to noise signal for cancellation. The most frequently used methods are ICA, CAR, SL, PCA, CSP and Adaptive Filtering [22].

Independent Component Analysis (ICA)

ICA was first applied to EEG by Makeig et al. in 1996 [23]. ICA separates the artefacts from the EEG signals into independent components based on the characteristics of the data without relying on the reference channels. The data is preserved during the ICA artifact removal. The ICA algorithm decomposes the multi channel EEG data into temporal-independent and spatial-fixed components. ICA algorithm is efficient in computation. When the size of the data is large, ICA shows high performance.

Common Average Referencing (CAR)

CAR works by removing the noise by subtracting the common activity from the position of interest. The common activity can be the noise present in the EEG signal [24]. Artifacts gives low SNR in EEG signals. CAR outperforms all referencing methods and leads to good classification accuracy [25].

Surface Laplacian (SL)

An estimate of current density entering or leaving the scalp through the skull is referred to as the Surface Laplacian of the skull. SL is robust against artifacts generated at uncovered regions by the electrode cap and it solves the electrode reference problem [26]. SL is a way of viewing the EEG data with high spatial resolution.

Principal Component Analysis (PCA)

PCA was invented in 1901 by Karl Pearson and later developed independently by Harold Hotelling in 1930 [27]. The PCA transforms the correlated vectors into linearly uncorrelated vectors. These uncorrelated vectors are called as "Principal Components". This is a classical method of Second Order Statistics. It depends on decomposition of covariance matrix. PCA helps in reduction of feature dimensions. The application of PCA in a BCI system yields best classification results [28]. The PCA is good but it is not as good as ICA.

Common Spatial Patterns (CSP)

CSP was first presented by Koles and it can detect abnormal EEG activity [29]. CSP performs transformation of EEG signal into a variance matrix that maximally discriminates between different classes. CSP uses spatial filtering and with spatial information it detects the patterns in EEG. CSP does not require a-priori selection of subject specific frequency bands and knowledge of these bands and requires use of many electrodes. It is sensitive to artifacts and electrode positions.

Adaptive Filtering

Adaptive filters have the ability to modify signal properties according to the specific characteristics of the signals being analyzed. Noise removal using filters removes noise along with important information. If the signal and noise are overlapping then filters will remove the signal of interest. This problem can be overcome by the adaptive filters. Adaptive interference cancellation is a very efficient method to solve the problem of signals and interferences with overlapping spectra. By using the least mean square algorithm (LMS) the artefacts from EEG signal can be efficiently removed. With the use of LMS algorithm optimization of mean square error is achieved. In [30] for the removal of artifacts a new algorithm for the adaptive filters was proposed and it is named as Recursive Least Squares (RLS) algorithm and it has proved that the artifacts in the ECG signals are removed and a considerable improvement has observed in the SNR of ECG signal.

2.2.2 Channel Selection

Channel selection aims to chooses the optimal subset of channels from complete set of available channels. The channel selection is done to improve the model performance, provide faster processing and remove dimensionality curse, and locate brain area that is responsible for generating neural activity. Steps involved in channel selection are explained in the Fig below. The subset generation step is a heuristic search and it is terminated when a stopping criterion is satisfied (search is successful or a threshold is reached). In the last step, the selected channel subset is validated via a-prior knowledge about the data [31] [32] .

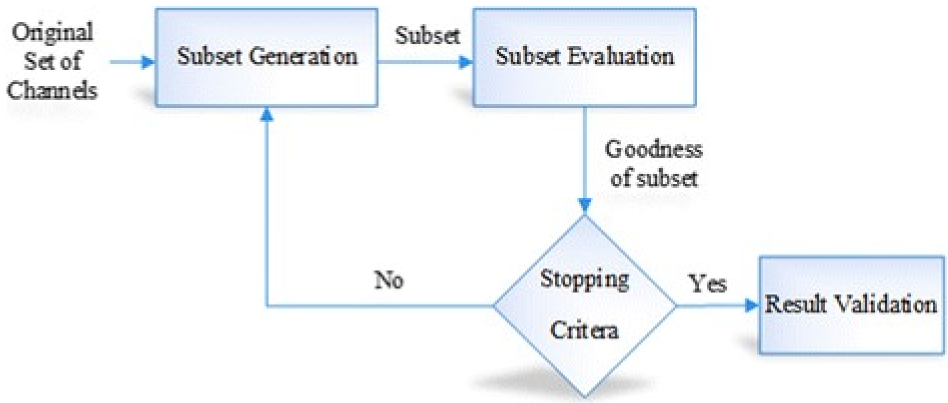


Figure 2.5: Channel selection steps.

Filtering Technique

Filtering techniques use an evaluation criterion which is independent, such as consistency measure, dependency measure, information measure, distance measure and to evaluate the candidate channel subsets, which are generated using a search algorithm. Advantage of filtering techniques is its high speed, scalability and independence from the classifier. But they give low accuracy, as they do not consider channel combinations. Fig below shows a general flowchart for the filtering techniques. In this flowchart,

S_0 represents the initial subset and Sbest represents the selected best subset of channels. Also, $D(C_0, ..., C_{n-1})$ represents a pool of n channels for selection, and M refers to an independent evaluation criterion. The γ represents the value of the evaluation criterion for each subset of channels. The "evaluate" function refers to an evaluation process.

Wrapper Technique

In case of wrapper techniques, a classification algorithm is used to evaluate the candidate channel subsets, which are generated by a search algorithm as shown in Fig below, in which A denotes a classifier, and γ_{best} represents the best value of the evaluation criterion. The evaluation of every candidate is obtained by training and testing a specific classification algorithm [33]. Consequently, they are more computationally expensive than filtering techniques and they are prone to overfitting.

Hybrid Techniques

A hybrid technique is a combination of a wrapper technique and filtering technique. It tries to take advantage of both in avoiding the pre-specification of a stopping criterion. Hybrid techniques utilize both a mining algorithm and independent measure for evaluation of the available channel subsets [10]. Two threshold values are evaluated: Θ_{best} corresponding to the case without a classifier and γ_{best} corresponding to the case with a classifier. The independent measure is used to select the best subset for a given size Cr (cardinality), and then the mining algorithm is used to select the final best subset across cardinalities.

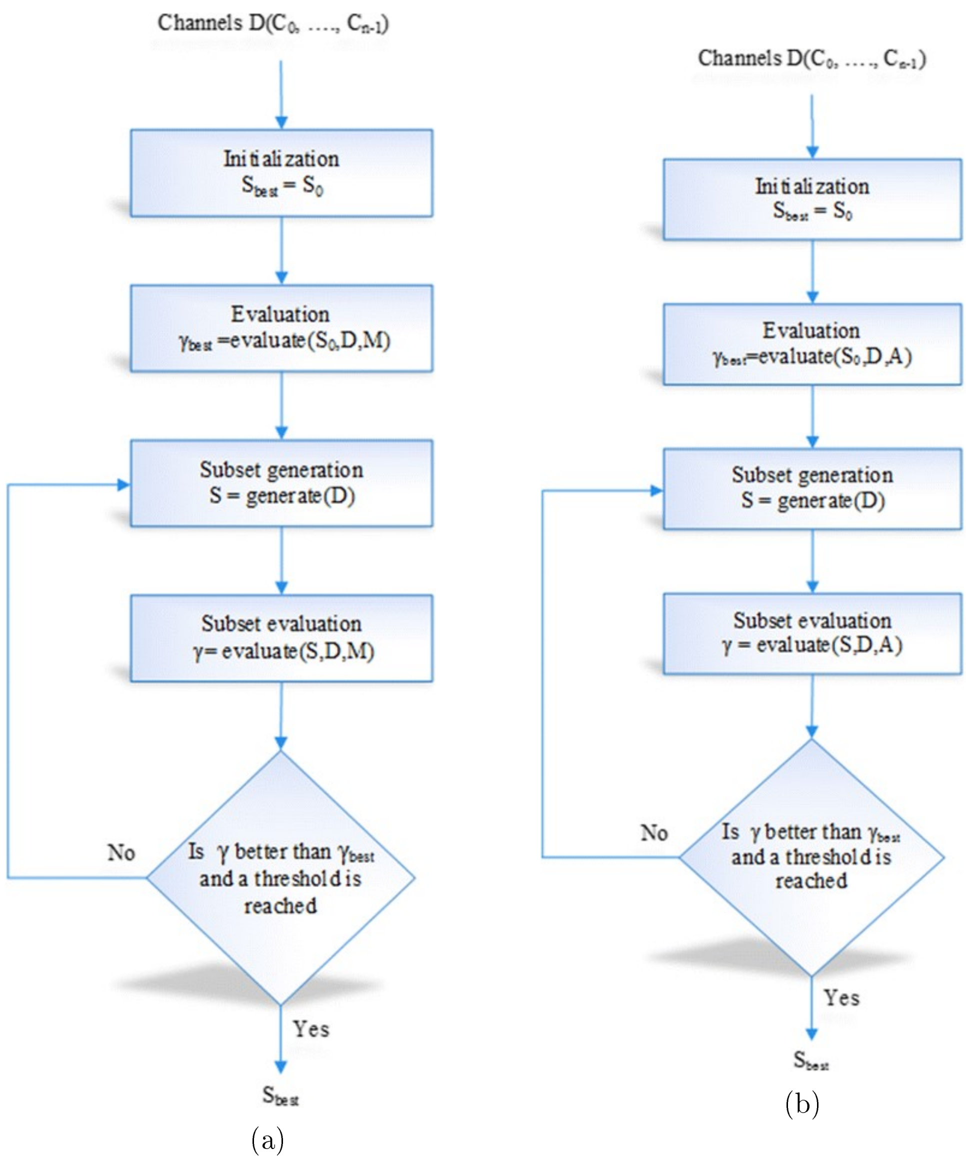


Figure 2.6: (a) Filtering technique, (b) Wrapping technique

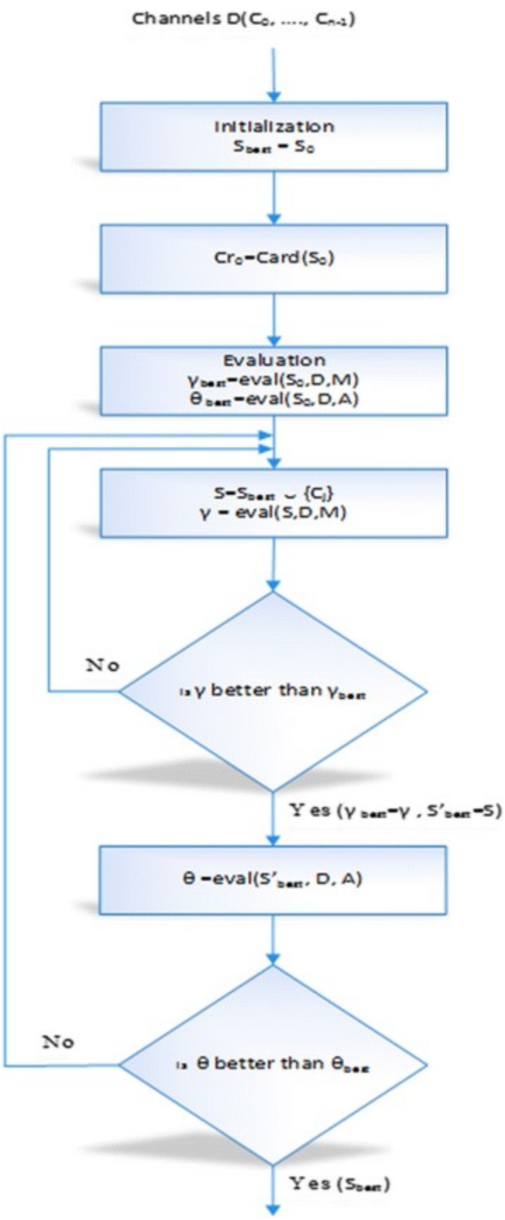


Figure 2.7: Hybrid technique.

2.2.3 Feature Extraction

A feature is a distinctive measurement or component extracted from the signal. Feature extraction process, extract the essential features needed for the classification of cognitive workload. Various feature extraction techniques like Wavelet Transformation (WT), Fast Fourier Transforms(FFT), Independent Component Analysis(ICA), Primary Component Analysis(PCA), Adaptive Auto Regressive parameters (AAR), multivariate AAR, Bilinear AAR, Kernel Density Estimation (KDE) etc. are available [34].

Independent Component Analysis (ICA)

ICA is one of the most commonly feature extraction techniques. ICA is a blind source separation technique (of knowing the source and channel characteristics), for the extraction of individual signals from mixture. It assumes that the different physical processes generate signals that are independent from each other. The observation signal is obtained by multiplying source signal with unknown mixing matrix. The aim of ICA will be to obtain un-mixing matrix so that original source signals can be reconstructed.

Principle Component Analysis (PCA)

PCA is a type of dimensional reduction or ordination analysis technique for feature extraction. Ordination analysis attempts to embed objects distributed in high dimensional space into lower dimensional space. In PCA, dimensional reduction is achieved by projection to lower dimensional space using linear transformation. Although PCA is a simple and classical method, it can often effectively reduce redundant information. PCA assumes that the data is linear and continuous.

Wavelet Transformation (WT)

Since the EEG signal is non-stationary in nature, it is befitting to use the time-frequency domain methods and WT is one of them. It does not impose the pseudo-stationary assumption on the data like the time and frequency domain methods. It is possible that WT represents the time function in terms of wavelets. The transforms are a family of functions derived from a generating function called mother wavelet using translation and dilation operations. Wavelet transform has an advantage of varying window size, being broad at low frequency and narrow at high frequency. This makes the WT suitable to get frequency data at low frequencies and precise time for the analysis of irregular knowledge patterns, such as impulses occurring at various time instances. WT uses multi scale structure.

- **Continuous wavelet transform (CWT)** - A CWT is employed to divide a continuous time function into wavelets. Unlike Fourier transform, the continuous wavelet transform acquires the power to construct a time frequency representation of a signal that gives excellent time frequency localization. However, its major weakness is that scaling parameter and translation parameter of CWT change continuously. Thus, the coefficients of the wavelet for all available scales after calculation will consume a lot of effort and yield a lot of unused information

- **Discrete wavelet transform (DWT)** - The disadvantages of CWT is overcome using DWT, which works on the principle of multi scale feature representation. Each step contains two digital filters $g(n)$ and $h(n)$ and two down samplers. The discrete mother wavelet $g(n)$ is high pass in nature, while its mirror image, $h(n)$ is low-pass.

Autoregression (AR)

AR methods estimate the power spectrum density (PSD) of the EEG using a parametric approach. Therefore, AR methods do not have problem of spectral leakage and thus yield better frequency resolution unlike non-parametric approach. Estimation of PSD is achieved by calculating the coefficients, that is, the parameters of the linear system under consideration.

Kernel Density Estimation (KDE)

KDE is a feature extraction technique to extract signals by computing density estimate using kernel smoothing function.

2.2.4 Feature Optimization

Feature optimization problem is similar as channel selection techniques. We shall not repeat those basic techniques like filter methods, wrapper methods and embedded methods again for feature optimization. Rather we shall cover other categorization for feature optimization. Feature optimization are classical algorithms widely used in machine learning and as such also very popular in BCI design (Garrett et al, 2003) [35]. There are too main families of feature optimization algorithms [36]:-

Univariate Algorithms

They evaluate the discriminative (or descriptive) power of each feature individually. Then, they select the N best individual features (N needs to be defined by the BCI designer). The usefulness of each feature is typically assessed using measures such as Student t-statistics, which measures the feature value difference between two classes, correlation based measures such as R2, mutual information, which measures the dependence between the feature value and the class label, etc. (Guyon and Elisseeff, 2003). Univariate methods are usually very fast and computationally efficient but they are also suboptimal. Indeed, since they only consider the individual feature usefulness, they ignore possible redundancies or complementarities between features.

Multivariate Algorithms

They evaluate subsets of features together, and keep the best subset with N features. These algorithms typically use measures of global performance for the subsets of features, such as measures of classification performances on the training set (typically using cross-validation Browne, 2000) or multivariate mutual information measures (Pudil et al, 1994) (Peng et al, 2005). This global measure of performance enables to actually consider the impact of redundancies or complementarities between features. Some measures also remove the need to manually select the value of N (the number of features to keep), the best value of N being the number of features in the best subset identified. However, evaluating the usefulness of subsets of features leads to very high computational requirements. Indeed, there are many more possible subsets of any size than individual features. As such there are many more evaluations to perform. In fact, the number of possible subsets to evaluate is very often far too high to actually perform all the evaluations in practice. Consequently, multivariate methods usually rely on heuristics or greedy solutions in order to reduce the number of subsets to evaluate. They are therefore also suboptimal but usually give much better performances than univariate methods in practice. On the other hand, if the initial number of features is very high, multivariate methods may be too slow to use in practice.

2.3 Survey for Modelling for Cognitive Workload Measurement

After feature extraction the signals are classified into various classes using various classifiers. Different types of classifiers include Linear Classifiers (LC), Artificial Neural Networks (ANN) based classifiers, Nonlinear Bayesian Classifiers (NBC) and Nearest Neighbour Classifiers (NNC). Of these classifiers, LC and NBC are mostly used in BCI design [25].

Linear Classifiers (LC)

Linear classifiers use the linear functions to classify signals into classes. The most frequently used linear classifiers are described below.

- **Linear Discriminant Analysis (LDA)**
  LDA is a generalization of Fisher's linear discriminant, a method used in statistics, pattern recognition and machine learning to find a linear combination of features that characterizes or separates two or more classes of objects or events.

LDA is closely related to analysis of variance (ANOVA) and regression analysis, which also attempt to express one dependent variable as a linear combination of other features or measurements. LDA creates models of the Probability density function respectively. LDA is simple to use and has very low computational requirements. It provides good results. For non-Gaussian distributions LDA may not preserve the complex structure in the data. LDA fails if the discriminatory function is not in mean but in the variance of the data.

- **Support Vector Machine (SVM)**
  SVM is a linear classifier that is used by most of the BCI applications. SVM was developed by Vapnik and was driven by statistical learning theory following the principle of structural risk minimization. SVM finds a hyper plane to separate the data sets. It separates data sets with clear gap that is as wide as possible to classify them into their relevant category. The hyper plane maximizes the margin that is the distance between the hyper plane and the nearest points from each class that are called as support vectors. The objective of this method is to provide good generalization by maximizing the performance of machine while minimizing the complexity of learned model. In by using a kernel-based SVM approach a mean classification accuracy of 87% was obtained. SVM has more performance and has high computational complexity. Shouyi Wang et al. in 2016 [11] used SVM for classification of EEG signal

Artificial Neural Network (ANN)

ANNs are non linear classifiers composed of large number of interconnected elements called neurons. Each neuron in ANN simulates the biological neuron and is capable of performing simple computational tasks. The most frequently used neural network is the Multi Layer Perceptron Neural Network (MLPNN) in which, the network is arranged into three layers viz., input layer, hidden layer and output layer. The advantage of MLPNN is that its fast operation, ease of implementation and requiring small training sets. The no. of inputs denotes the no. of features selected and, no. of outputs denotes the no. of classes formed. The complexity of an ANN is estimated by the no. of neurons in the hidden layer of it. The large the no. of neurons in hidden layer the more the complexity, less no. of neurons in hidden layer causes classification errors. No specific criterion was defined for making this decision in hidden layer. By using trial and error method the no. of neurons has to be decided.

Naive Bayes' Classifier (NBC)

NBCs produce non linear decision boundaries. They are generative in nature and enabled to perform more efficient rejection of uncertain samples than discriminative classifiers. Bayesian Classifiers assign a feature vector to its class with highest probability. Using Bayes' rule a posteriori probability of a feature vector is computed. Hidden Markov Model is a non linear Bayesian classifier and is a dynamic classifier. This classifier is suitable for the classification of time series. These are not as widespread as Linear Classifiers and NNs in the field of BCI applications.

Nearest Neighbour Classifier (NNC)

NNCs assign a feature vector to a class based on its nearest neighbours. If the feature vector is from the training set then it is named as k-Nearest Neighbour (k-NN) classifiers. k-NN is a non parametric method, it predicts objects values or class memberships based on the k-closest training examples in the feature space. It assigns the label of a test sample with the majority label of its k-nearest neighbours from the training set. k-NN is very simple to understand, transparent, easy to implement, and debug. In Table IV, comparison of various signal classification methods was given.

Chapter 3

Design of Experiment

Data plays a very important role and its consistency is of prime importance in any research. Open source data available online is not always as per your choice and design as regard various parameters are concerned. It is established that there is no match to data collected on your own as per your ground rules and choice of parameters and protocols. Due to lack of open source EEG data available online as per the requirements of this thesis, it was decided to collect one, at the BCI Research lab facility located in IIT Kharagpur. This Chapter describes our experimental framework for data collection. Before we proceed ahead, we hereby declare that an Ethical Clearance was taken for collecting EEG data from volunteer subjects as per procedures in vogue.

3.1 Working Memory Task

The n-Back and Dual n-Back tasks are continuous performance task, commonly used for assessment in cognitive neuro-science to measure a part of working memory and working memory capacity [19]. The n-back task was introduced by Wayne Kirchner in 1958 and the Dual n-Back task is a variation that was proposed by Susanne Jaeggi et al. in 2003. n-Back and Dual n-Back task for generating cognitive mental load has been chosen for experimental design to remove subjectivity factor in the captured EEG signals for better model generation.

3.1.1 n-Back Task

When n-Back task starts, a square(stimuli) appear every 3 seconds accompanied by the sound of an english alphabet in either of nine positions as shown in the 3x3 grid in Fig 3.3.

- **Sound n-Back**
  Subject is required to press key 'L' (sound match) on the keyboard, if the english ALPHABET heard at *stimuli* is the same as it was n trials back.

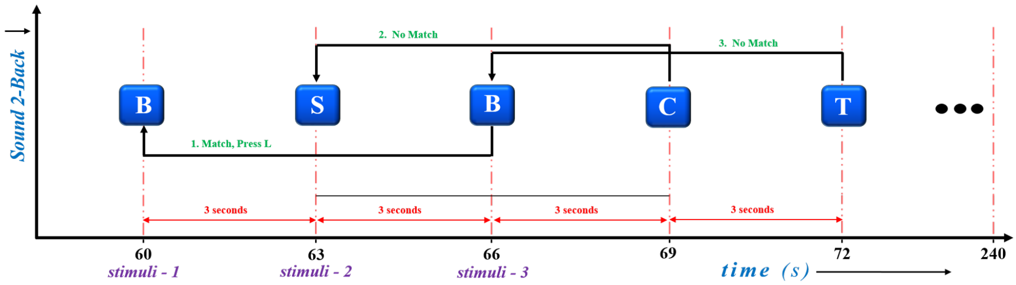


Figure 3.1: Sound *2*-Back task.

- **Position n-Back**
  Subject is required to press key 'A' (position match) on the keyboard, if the POSITION of the blue square at *stimuli* is the same as it was n trials back.

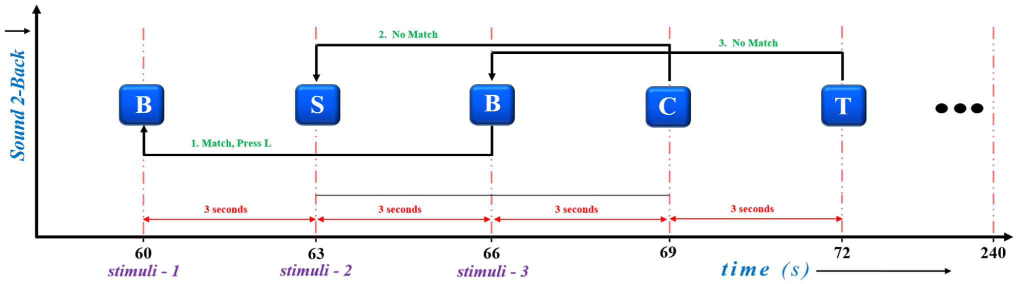


Figure 3.2: Position *2*-Back task.

3.1.2 Dual n-Back Task

This task involves remembering a sequence of spoken alphabet and a sequence of positions of a square at the same time, and identifying when a alphabet or position matches the one that appeared n trials back. Similarly other games are played based on above ground rules only. The Dual*ness* of the games induces more cognitive workload.

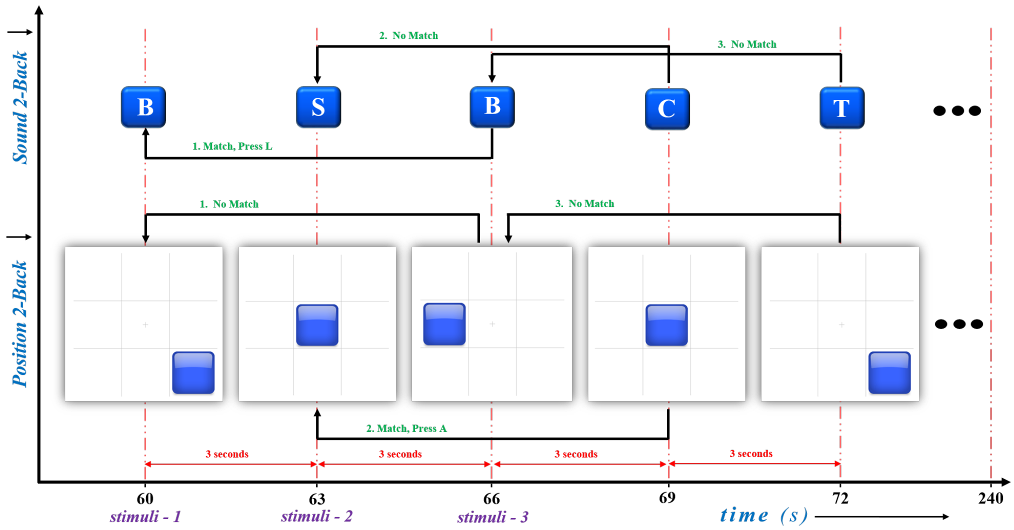


Figure 3.3: Dual *2*-Back task.

The n-Back/ Dual n-Back tasks have all three ingredients of cognitive load.

- ***Intrinsic load:*** The inherent difficulty of task can be increased firstly by increasing the value of "n" from 1 to 2 and secondly by migrating from n-Back to Dual n-Back task.
- ***Extraneous load:*** This can be increased by reducing the time between the stimuli. In our experimental setup we decided to keep it constant at 3 seconds.
- ***Germane load:*** This can be increased by increasing the value of "n" which leads to increased amount of information required to be stored and processed in working memory.

3.2 Hardware and Software Interface

Hardware Interface

Device used for this research work is EMOTIV Epoc+. It uses international 10-20 standards for locations of electrodes. This device uses 14 Channels for data collection from various part of brain. This device is comparatively cheap and has been used by many researchers in their work. EMOTIV Epoc+ is an EEG headset as shown in figure below, build by Australian company Emotiv Systems. EMOTIV Epoc+ captures data in high resolution using 14 EEG channels plus 2 references for accurate spatial resolution, at sample rates of 128 samples per seconds (SPS). The device operates at a resolution of 14 bits per channel and frequency response between $0.16 - 43$ Hz [37].

Figure 3.4: EMOTIVE Epoc+ headset and electrode positions.

EMOTIV Epoc+ has been extensively used for EEG based BCI research. Ramirez et al. Epoc+ is widely used for P300 based research, Duvinage et al. studies the performance of Emotiv EPOC for P300 based applications. Epoc+ is also evolving as a major player in BCI based gaming systems.

Table 3.1: EMOTIV Epoc+ specifications

Feature	**Description**
Number of Channels	14 (plus CMS/DRL reference, P3/P4 locations)
Channel Names	AF3,F7,F3,FC5,T7,P7,O1,O2,P8,T8,FC6,F4,F8,AF4
Sampling Method	Sequential Sampling. Single ADC
Sampling Rate	128 SPS
Resolution	14 Bits 1 LSB = 0.51 μ V (16 bit ADC)
Bandwidth	0.2 - 45 Hz, digital notch filters at 50Hz & 60 Hz
Filtering	Built in digital 5th order Sinc filter
Dynamic Range	8400 μ V(pp)

Software Interface

EMOTIV TestBench is a software that comes as part of the EMOTIV SDK. It allows to record and replay files in .EDF format. TestBench is a powerful tool that can be used to visualize the signals captured, convert the data into different file formats like csv etc. Some advance features like Fast Fourier transform (FFT), wireless packet acquisition/ loss display are also supported. As shown in figure, the software indicates the headset battery level, and also the sensors contact quality. The tool displays the data from all channels for 5 sec of rolling time window. We can view data from selective channels or for a particular frequency band. The tool is particularly helpful in offline recording of data for further analysis and classification. This is a very easy to use convenient software interface for capturing real-time EEG signal of a subject. It captures the EEG signal through Bluetooth dongle provided with it. The data was captured in European Data Format (.EDF files.). We converted the .EDF files to .MAT format for signal preprocessing.

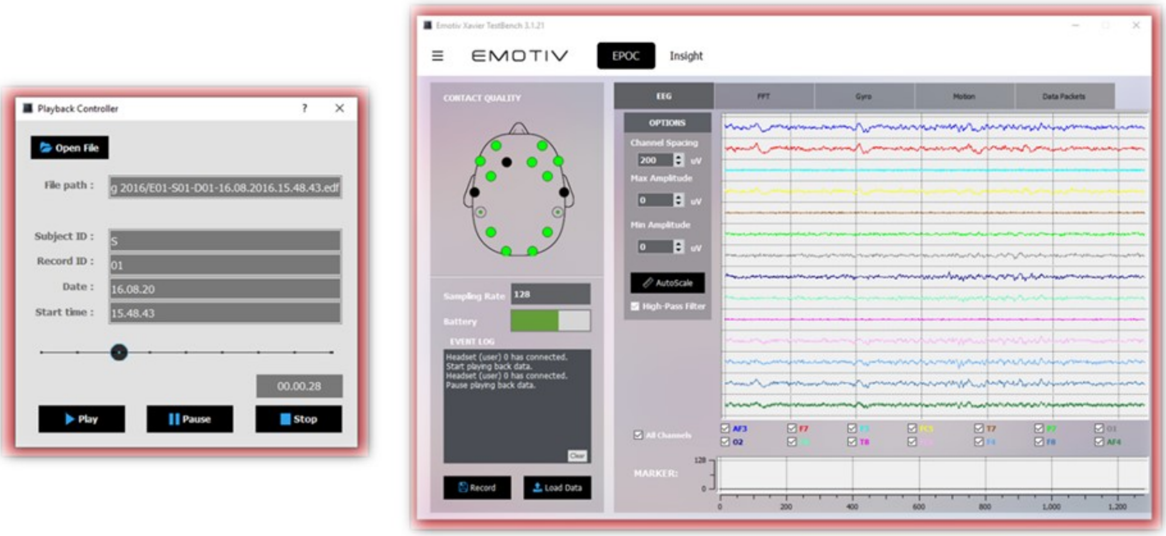


Figure 3.5: Test bench and recorder interface.

3.3 Experimental Setup

The choice of ideal, n-Back/Dual n-Back task as mentioned above was based on the need to induce five different load levels as depicted in the table below along with cognitive load class. We label our various cognitive workload level as Ci, where i = 1, 2, 3, 4, 5. For example, for Dual *1*-Back task the class label is C4.

Table 3.2: Cognitive workload levels

Task	Cognitive Workload Class
Ideal	C1
1-Back	C2
2-Back	C3
Dual *1*-Back	C4
Dual *2*-Back	C5

The experiments were conducted in controlled environment in an electrically isolated BCI laboratory at IIT Kharagpur. The tasks were conducted with a minimum distance of 5 m from all power sources and under adequate illumination, on an all-in-one PC with a viewing distance of 70 cm from the subject under experiment. To minimize any the artifacts generated, the subjects were asked to avoid any unnecessary physical movements and their hands were placed in a fixed position, so that they could still make a slight finger movements in response to the correct answer on the keyboard as required by task. Subjects were requested to refrain from excessive blinking of eye lid. The subjects were given a 60 second break between every pair of tasks, where they could move or blink. In the Idle task, the subjects were asked to sit relaxed and keep their eyes closed without any movement during the capture of EEG signal. The open source application "Brain Workshop" [38] was used for data collection experiments.

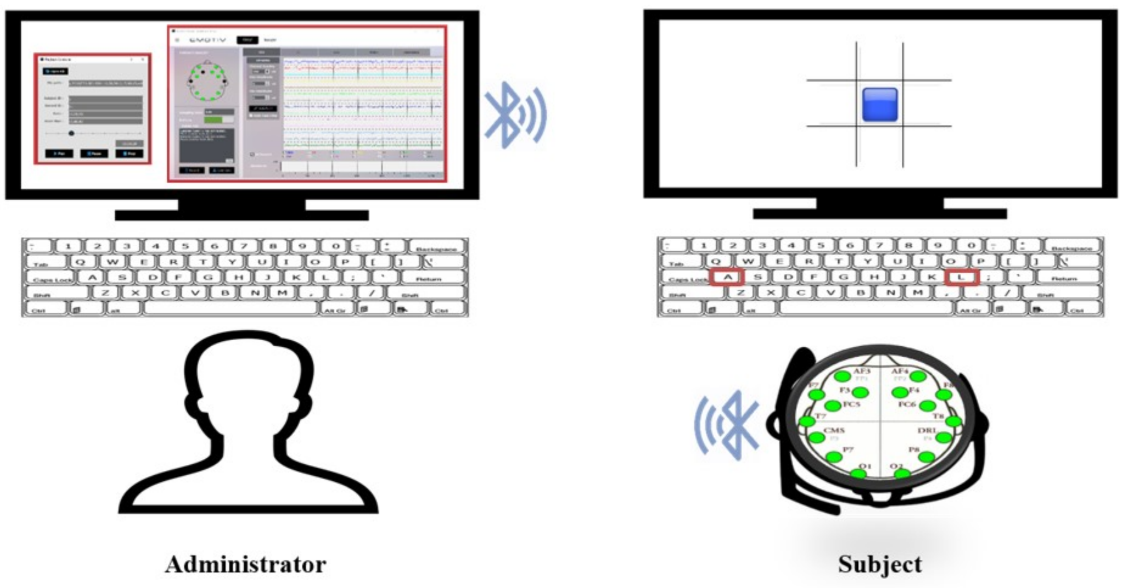


Figure 3.6: Experimental setup.

As marked clearly in Fig 3.3, each individual task had 60 audio/visual stimuli (depending on various tasks, as explained above in this chapter) at every 3 seconds (at which subject takes a decision of pressing or not-pressing 'A' or 'L' key). Hence, each task was presented to subjects for 180 seconds ($60 \times 3 = 180$).

Each subject was made to undergo total 5 experiments back to back, as shown in the Fig 3.7 below. Each experiment preceded with a 60 seconds break (where subject could just relax), making each setting of total $(60 + 180) \times 5 = 1200$ seconds or 20 minutes. In ideal task, the individual just sat ideal without any movement and eyes closed.

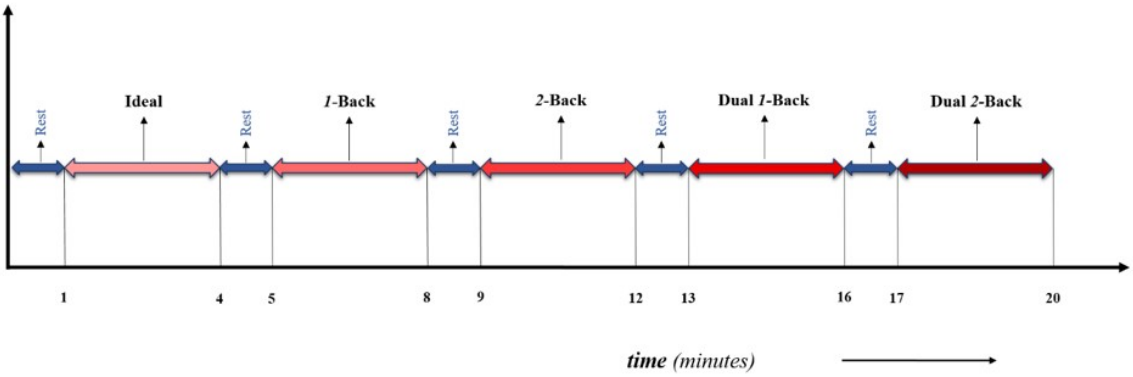


Figure 3.7: Experiment template.

3.4 Subjects and Metadata

Five healthy male and five healthy female volunteers engaged in graduate study participated in the experiment with age ranging between $20 - 24$ years. All subjects were right-handed and had normal/corrected to normal eye-sight. Subjects were allowed to wear spectacles during the experiments. All subjects were given adequate practice and idea of the experiments. It was ensured that the subjects did not carry mobile phones inside the Lab during the test. Room temperature was controlled to ensure adequate comfort during the experiments.

Due to ethical reasons, the identity of subjects participated in data collection is not being revealed. Also, the photographs displayed in Fig 3.8 and 3.9 are deliberately blurred to keep their photo identity confidential.

Figure 3.8: Male subjects.

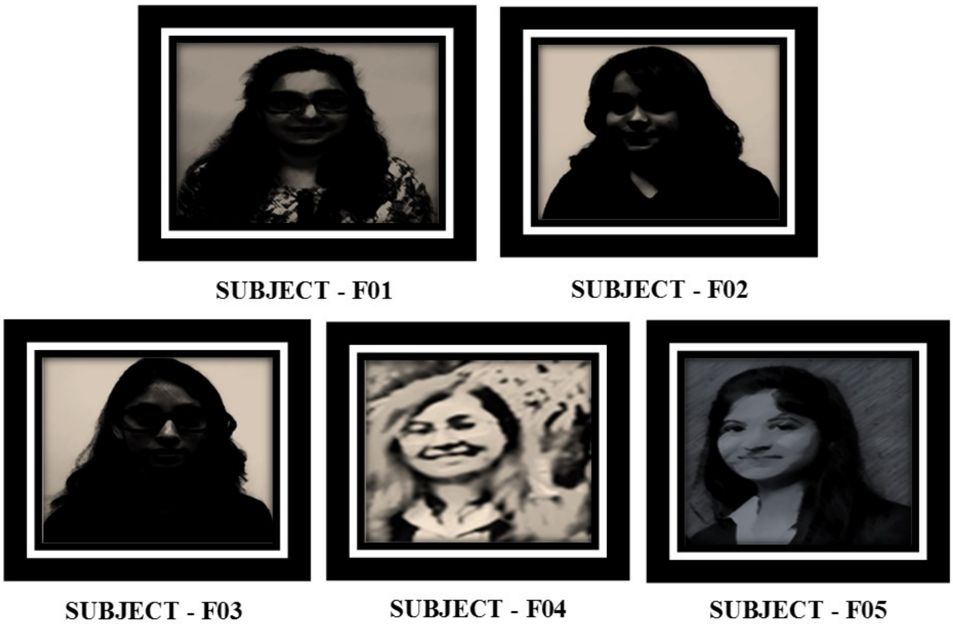


Figure 3.9: Female subjects.

Metadata of the EEG experiments conducted is summarized in the Figure below:-

Table 3.3: Metadata

Device **EMOTIV Epoc+**	**Sampling Rate** **128 Hz**	**Duration** **20 Minutes**	**Subjects** **Students, IIT Kharagpur**	
Date	**Start Time**	**Subject ID**	**Gender**	**Age**
25 Sep 2016	1554 H	M01	M	20
25 Sep 2016	1659 H	M02	M	21
26 Sep 2016	1751 H	M03	M	24
27 Sep 2016	2307 H	M04	M	20
27 Sep 2016	2341 H	M05	M	21
09 Jan 2017	1247 H	F01	F	22
09 Jan 2017	1730 H	F02	F	21
11 Jan 2017	1718 H	F03	F	21
12 Jan 2017	1745 H	F04	F	22
12 Jan 2017	1919 H	F05	F	22

The data collected from each subject with 14 EEG channels is 2,150,400 samples, as explained below:- Let $A_{m \times n}$ be a matrix of size $m \times n$ representing data from each channel. where,

$$
\begin{aligned}
m &= number\ of\ samples \\
n &= number\ of\ EEG\ channels \\
m &= time \times sampling\ rate \\
&= 20 \times 60s \times 128Hz \\
&= 153,600 \\
n &= 14 \\
\therefore m \times n &= 153,600 \times 14 \\
&= 2,150,400
\end{aligned}
$$

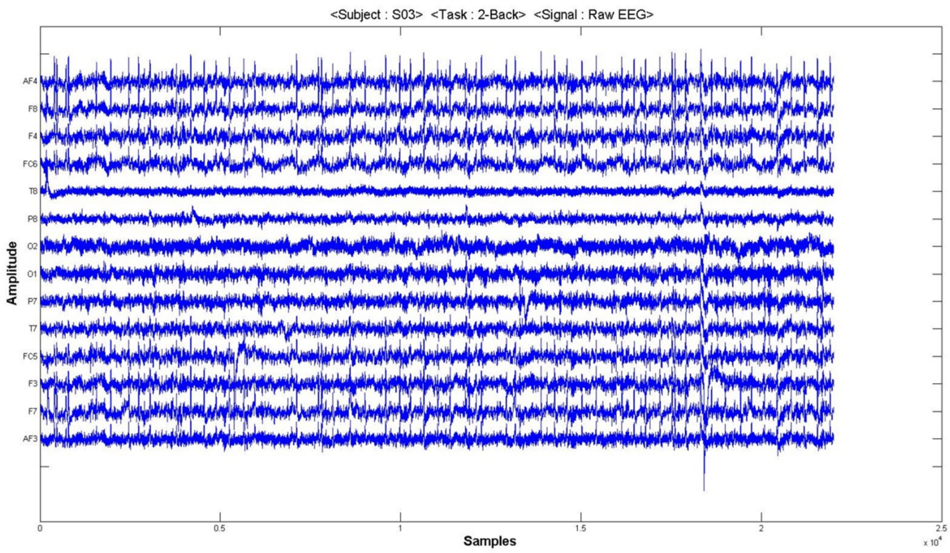


Figure 3.10: A sample 14 channel EEG data.

This data collected above is chopped into individual experiment data by removing 5 x 60 seconds break and 4 x 180 seconds of remaining experiments. Hence we are left with total $3, 22, 560$ data samples per experiment and $23, 040$ data samples per channel per subject. This data is stored in European Data Format (.EDF) file.

Chapter 4

Signal Preprocessing and Feature Engineering

In this chapter we shall discuss the signal processing and feature engineering aspects which include channel selection, feature extraction and feature optimization methods along with the proposed algorithm used to achieve the same.

4.1 Signal Preprocessing

EEG devices are capable of recording electrical activities other than the brain signals. Recorded EEG signal which is not originated from brain is termed as artifact.

4.1.1 Artifact Removal

Artifact removing ia an approach of recognizing, identifying source and removing of EEG artifacts and is an important process to ensure good quality of recorded signal. There are many varieties of artifacts present in EEG signal. Few of the artifacts as captured during our experiments are described below along with the 14-Channel EEG streams.

1. **Cardiac artifacts** - Heart gives two kinds of artifacts, mechanical and electrical. Both gets generated due to cardiac contractions and are easy to identify. Electrical artifact is due to ECG and mechanical is due to the circulatory pulse.
    (a) **Electrical** - Electrical artifact is due to ECG.
    (b) **Mechanical** - Mechanical is due to the circulatory pulse.

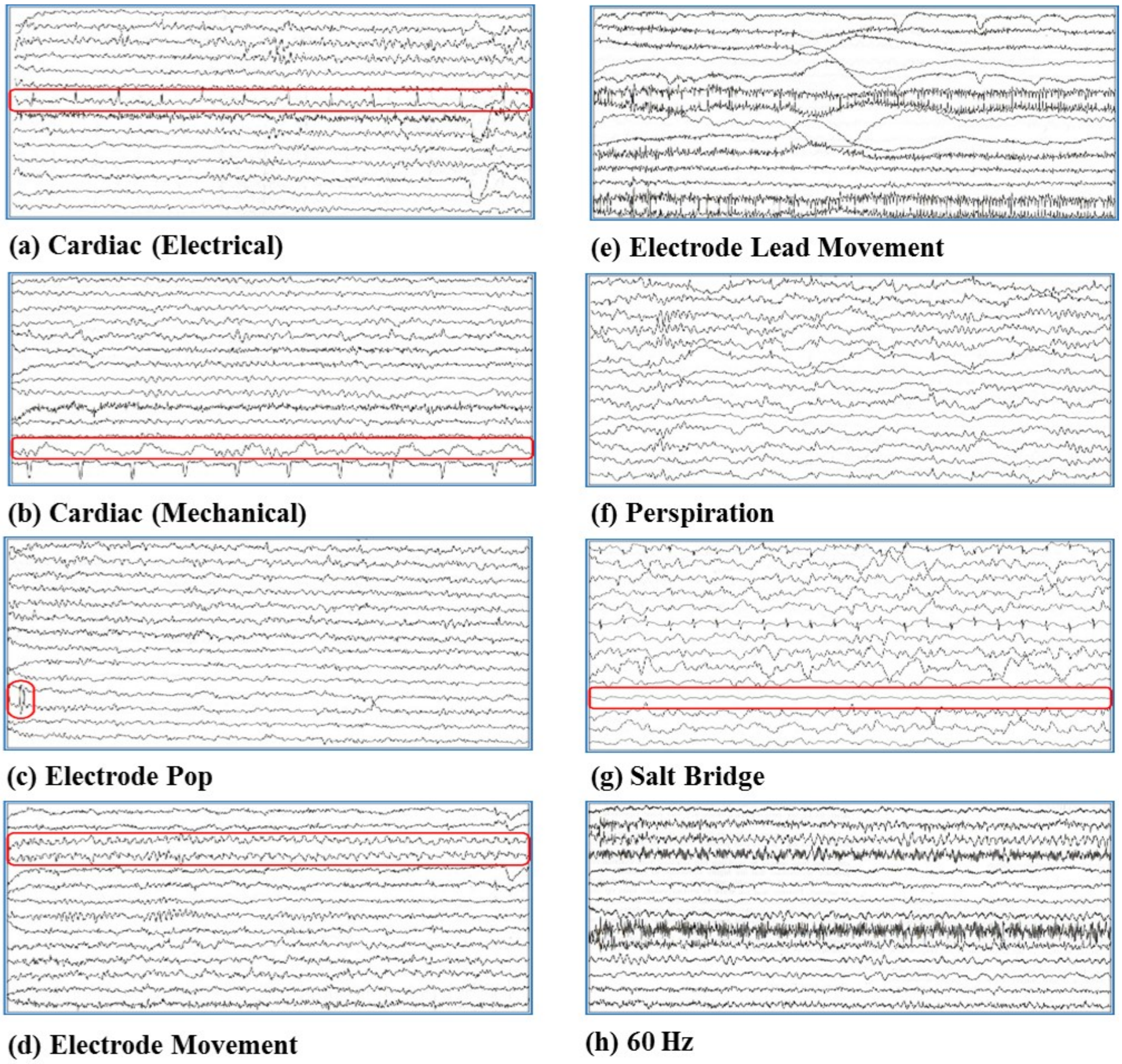
2. **Electrode Artifacts**

   (a) **Electrode pop** - They are due to electrode and skin interface which act as a capacitor and store electrical charge across the electrolyte around the electrode.

   (b) **Electrode Movement** - This happens due to movement of electrodes bacause of muscular movement in scalp.

   (c) **Electrode Lead Movement** - Electrode lead movement causes artifact due to varying voltages at electrodes due to movement.

   (d) **Perspiration (Sweat)** - Sweat causes artifacts due to unwanted electrical connection between electrodes and skin.



(a) Cardiac (Electrical) **(e) Electrode Lead Movement**
(b) Cardiac (Mechanical) **(f) Perspiration**
(c) Electrode Pop **(g) Salt Bridge**
(d) Electrode Movement **(h) 60 Hz**

Figure 4.1: EEG artifacts.

(e) **Salt bridge** - Due to electrolyte getting dried up, it causes unnecessary obstruction in capturing clean EEG.

3. **External device artifacts**

(a) **50/60 Hz Ambient Electrical Noise** - These artifacts are caused due to noise in 50/60 Hz electrical lines in the vicinity.

(b) **Electrical Motor** - Electric motor may produce high amplitude, irregular, poly-spike or spike artifact due to the switching magnetic fields within the motor.

(c) **Phone** - Mechanical telephone bells causes sinusoidal form of artifact.

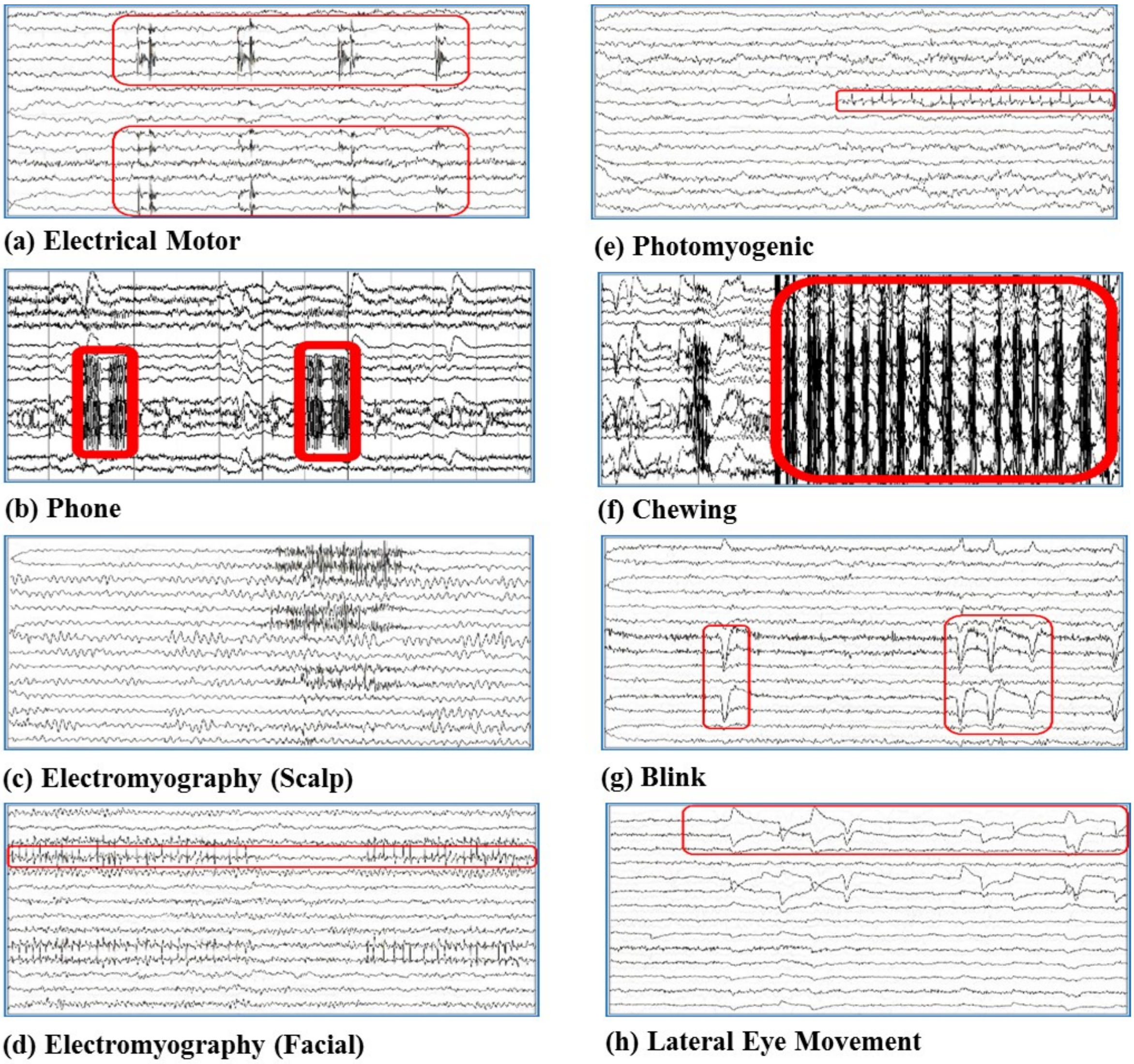


Figure 4.2: EEG artifacts.

4. **Muscle artifacts**

    (a) **Electromyography (scalp/facial muscle)** - They are caused due to muscle activity and it could last for less than a second to an entire EEG record.

    (b) **Photomyogenic** - These artifacts are due to the effect of light. The endeavor is to keep the lights to bare minimum in the EEG recording room.

    (c) **Glossokenetic(Chewing)** - Chewing or jaw clinching causes the worst possible noise called Glossokenetic across all EEG channels.

5. **Ocular artifacts**

    (a) **Blink** - Due to rapid movement of the eyes in up and down direction during blink causes this artifact.

    (b) **Lateral Eye Movement** - This artifact may persist for a very long duration corrupting the signal during lateral eye movement.

4.1.2 FORCe Algorithm

We have employed FORCe: Fully Online and Automated Artifact Removal for Brain-Computer Interfacing (FORCe) an automatic ICA-based algorithm FORCe. FORCe applies stereotyped artifact-specific spatial and temporal features to identify independent components of artifacts automatically. Without affecting the activity of neural sources, this algorithm removes the artifacts efficiently [39].

The FORCe algorithm works on a novel combination of wavelet decomposition, independent component analysis and thresholding of EEG signals. FORCe does not require additional signals (e.g., electro-oculogram or electro-myogram signals, which is generally recorded with EEG signals separately) and it works on small Epoch of EEG data. This algorithm outperforms the state-of the-art automated artifact removal methods, Fully Automated Statistical Thresholding for EEG artifact Rejection (FASTER) and Lagged Auto-Mutual Information Clustering (LAMIC). FORCe can remove a wide range of artifact as described in the previous subsection.

Algorithm Steps:-

1. Decompose EEG signal from each channel into a set of approximation and detail coefficients with wavelet decomposition. Denote $C_j^i \in C$ and j^{th} coefficient set from the set of all coefficients, from channel i. Wavelets decomposes the signal by convolving it with a mother wavelet function at a range of different frequency and time locations. Then it measures the strength of the signal as a coefficient

of the wavelet function. FORCe uses discrete wavelet transform (DWT)as this scales to the signal at a discrete set of frequencies and times. Mathematically, WT may be defined as:-

$$\omega(t,f) = \int_{-\infty}^{\infty} x(t) * \psi_{s,r}(t)dt$$

$$\psi_{s,r}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-\tau}{s})$$

where $x(t)$ is the original signal and * denotes its complex conjugation. $\omega(t,f)$ shows how the signal $x(t)$ is translated into a set of wavelet basis functions $\psi_{s,r}(t)$ at scale and translation dimensions s and τ. ψ is the mother wavelet function with which the signal is convolved.

2. Next, it groups same decomposition level coefficients from each channel into sets of coefficients, $A_n = c_j^i \epsilon C \mid \forall_i \epsilon K, j = n$, where K is the set of channels and n denotes the decomposition level.

3. Estimate an ICA demixing matrix to separate the coefficients into maximally statistically independent components (ICs) from the set of approximation coefficients (A_i).

4. Multiply demixing matrix with the set of approximation coefficients.

5. Identify ICs with artifacts and remove them.

6. To obtain an estimate of the cleaned approximation coefficient set, invert the ICA decomposition.

7. Identify spike zones in both the approximation and detail coefficient sets and then apply soft thresholding to reduce magnitude.

8. Reconstruct the cleaned EEG from the wavelet approximation and detail coefficient sets.

Sample 14 Channel Raw and Clean EEG signal after application of FORCe algorithm is shown in Fig 4.3 and Fig 4.4. It can be seen clearly that the atrifacts present in the 14 Channel Raw EEG signal has been eliminated after application of this novel algorithm.

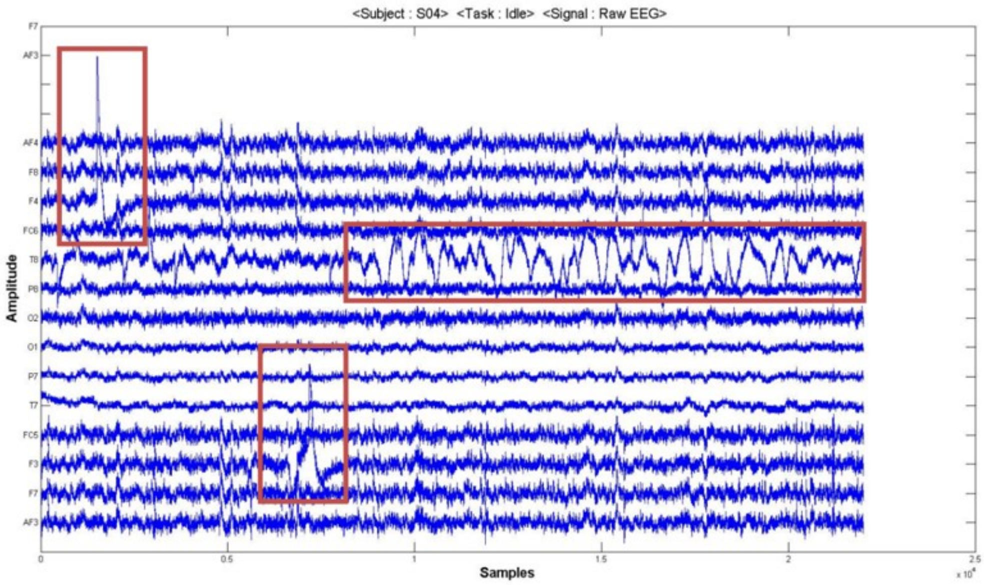


Figure 4.3: Raw EEG signal.

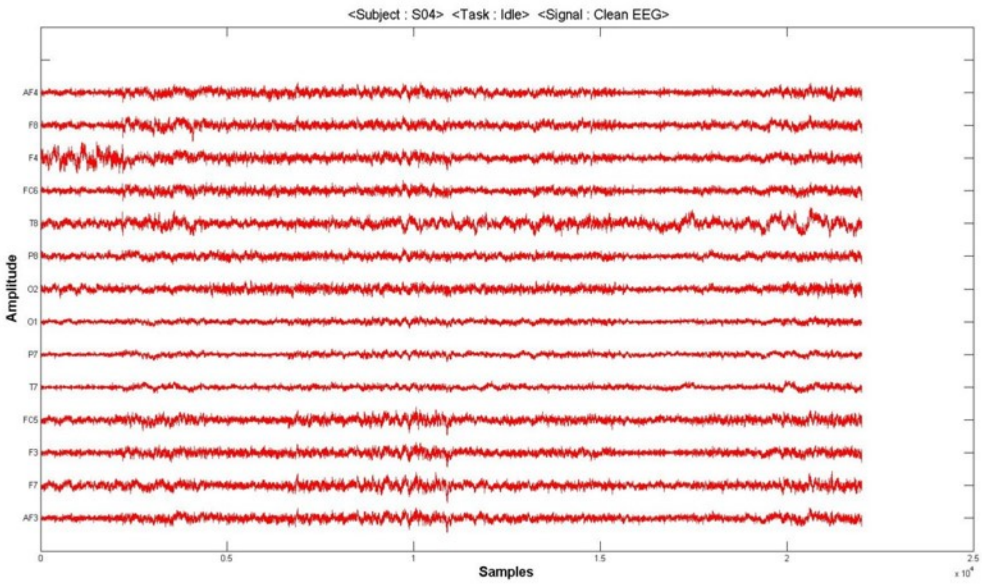


Figure 4.4: Clean EEG signal.

The clean data after artifact removal is fed for feature extraction. Baseline removal of each EEG time-series data (the .MAT extension files) was done before doing the further processing. The signal preprocessing was done in MATLAB. After the signal preprocessing done in this chapter, the processed data in .MAT (MATLAB file) format was converted to .CSV (the universal Comma Separated Values format) for further processing.

4.2 Channel Selection

Various channel selection techniques used have already been discussed in Chapter 2. In our work, we use a very simple technique of channel selection which is based on non-linear approach called Mutual Information (MI) as described below.

4.2.1 Mutual Information

To evaluating nonlinear dependencies between random variables, MI has been used in probability theory [40] . Let X and Y be two random variables. Then, the MI between X and Y is the measure of amount of knowledge on Y provided by X and vice-versa. The MI between above stated two random variable X and Y can be defined as:-

$$I(X;Y) = H(X) - H(X|Y)$$
$$I(X;Y) = H(Y) - H(Y|X)$$
$$I(X;Y) = H(X) + H(Y) - H(X;Y)$$

where $H(X)$ and $H(Y)$ are the entropies of random variables X and Y, and $H(X;Y)$ is their joint entropy.

$$H(X) = -\int_X p_X(x) \log p_X(x) dx$$
$$H(Y) = -\int_Y p_Y(y) \log p_Y(y) dy$$
$$H(X;Y) = -\int_X \int_Y p_{X,Y}(x,y) \log p_{X,Y}(y) dx dy$$

If $H(X)$ and $H(Y)$ are to be independent, then it means that X contains no information about random variable Y and vice versa; implying the MI between the two, to be zero.

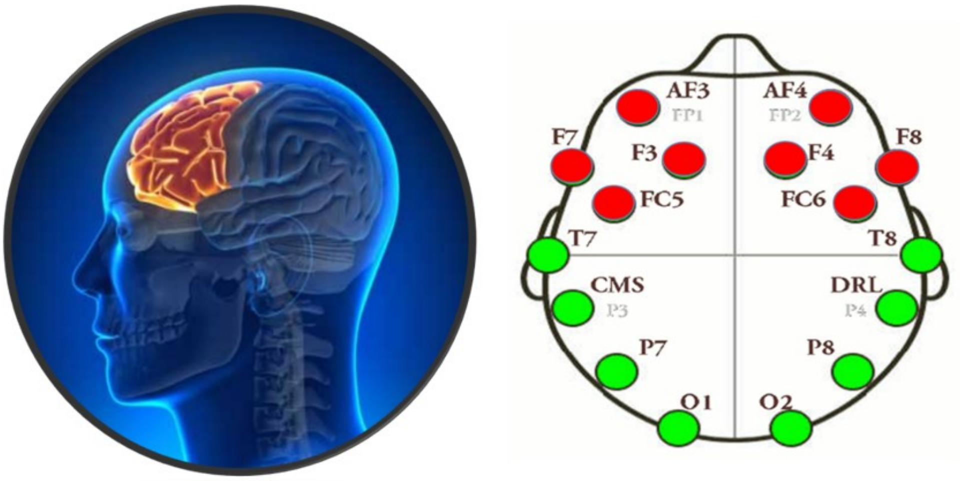


Figure 4.5: Channel selection.

Based on this MI, as discusses above, various channels are either selected or rejected. As shown clearly in Fig above, the selected channels are confined to frontal lobe, which verifies the theory of cognition which clearly describes that the cognitive workload neuron activity is observed in the forntal lobe of human brain. Above mentioned popular channel selection technique based on MI was applied to entire set of 14 channel clean input EEG signals and following 8 channels were selected after exhaustively running the algorithm on entire datasets from various sets of experiments: AF3, F3, FC5, F7, F8, FC6, F4 and AF4. The algorithm was run in Python 2.7 on Linux OS.

4.3 Feature Extraction

After signal preprocessing and channel selection, the information needed for classification into individual load levels is extracted. This process is called feature extraction, and represents an important step in EEG data processing. Feature extraction can be defined as automated recognition of various descriptive features of signals. Each segment obtained by signal segmentation can be represented by its extracted features. A good feature should remain unchanged if variations take place within a class, and it should reveal important differences when discriminating between patterns of different classes. It is necessary not only to compute appropriate features but also to save them into a unified data structure. In our case, we store a list of computed features and the matrix of their values into the .csv file structure on a local linux machine. Various types and categories of features calculated in this thesis are discussed below in brief:-

Statistical Features

Statistical features are best suited for a time-series signal like EEG. An EEG signal can be characterized by the distribution of the amplitude and its moments. For each epoch of an EEG signal, following features were calculated.

Table 4.1: Statistical features

FEATURE	DESCRIPTION
MEAN	Mean value
STD	Standard deviation
MAX_VALUE	Maximum positive amplitudes
MIN_VALUE	Maximum negative amplitudes
SKEWNESS	A measure of asymmetry of the distribution
KURTOSIS	A measure of flatness of the distribution
MEDIAN	The middle value of a set of ordered data
FD	Fractal Dimension
AR	Auto Regression

Derivative Features

Calculating the first and second derivative of mean and max of series signals gives a very meaningful feature value.

Table 4.2: Derivative features

FEATURE	DESCRIPTION
1^{st} DIFF_MEAN	Mean value of the first derivative of the signal
1^{st} DIFF_MAX	Maximum value of the first derivative of the signal
2^{nd} DIFF_MEAN	Mean value of the second derivative of the signal
2^{nd} DIFF MAX	Maximum value of the second derivative of the signal

Interval or Period Features

EEG signals can also be analyzed based on measurement of distribution of the intervals between zero and other level crossings or between maximum and minimum. We have calculated the following features.

Table 4.3: Interval or period features

FEATURE	DESCRIPTION
LINE_LENGTH	Line length
MEAN_VV_AMPL	Mean of vertex to vertex amplitudes
VAR_VV_AMPL	Variance of vertex to vertex amplitudes
MEAN_VV_TIME	Mean of vertex to vertex times
VAR_VV_TIME	Variance of vertex to vertex times
MEAN_VV_SLOPE	Mean of vertex to vertex slope
VAR_VV_SLOPE	Variance of vertex to vertex slope
ZERO_CROSSING	Number of zero crossings in a signal
MIN_MAX_NUMBER	Number of local minima and maxima
COEFF_OF_VARIATION	A statistical measure of the deviation of a variable from its mean, standard deviation divided by mean
AMPL_RANGE	The difference between the maximum positive and maximum negative Amplitude values

Hjorth Parameters

Hjorth parameters gives the complexity of a time-series EEG signal. These values are vary useful in EEG analysis and are a very useful tools for the its quantitative description.

Table 4.4: Hjorth parameters

FEATURE	DESCRIPTION
HJORTH1	Ability
HJORTH2	Mobility $(\sigma_{x'}/\sigma_x)$
HJORTH3	Complexity $(\frac{\sigma_{x''}/\sigma_{x'}}{\sigma_{x'}/\sigma_x})$

Frequency Features

These features are the most important features for analysis of EEG Signal. Based on the frequency content of EEG signals, following variations of EEG signals frequency features were calculated. Fast Fourier Transform was applied to various EEG wave bands to obtain values. We also calculate useful ratios of FFT in various bands. These are the most promising features of EEG signal.

Table 4.5: Frequency features

FEATURE	DESCRIPTION
FFT_DELTA	0.1 - 4 Hz
FFT_THETA	4 - 8 Hz
FFT_ALPHA	8 - 13 Hz
FFT_BETA	13 - 30 Hz
FFT_GAMMA	30 - 40 Hz
FFT_WHOLE	0.1 - 40 Hz
FFT_DT_RATIO	DELTA / THETA
FFT_DA_RATIO	DELTA / ALPHA
FFT_TA_RATIO	THETA / ALPHA
FFT_DTA_RATIO	(DELTA + THETA) / ALPHA
FFT_SEF	Spectral edge frequency
FFT_SP-ROLL_OFF	Below which 85 % of the total spectral power resides

Wavelet Features

The wavelet transform (WT) is capable of distinguishing very small and delicate differences between time-series signals even from short epoch of signal. It can describe highly irregular and non-stationary signals easily. WT based methods can localize the signal components in time-frequency space better than FFT.

Table 4.6: Wavelet features

FEATURE	DESCRIPTION
MIN_WAV_VALUE	Minimum value
MAX_WAV_VALUE	Maximum value
MEAN_WAV_VALUE	Mean value
MEDIAN_WAV_VALUE	Median value
STD_WAV_VALUE	Standard deviation
SKEWNESS_WAV_VALUE	Skewness
KURTOSIS_WAV_VALUE	Kurtosis
WAV_BAND	Relative energy
ENTROPY_SPECTRAL_WAV	The spectral entropy
1st DIFF_WAV_MEAN	Mean value of the 1st derivative
1^{st}_DIFF_WAV_MAX	Maximum value of the 1^{st} derivative
2^{nd}_DIFF_WAV_MEAN	Mean value of the 2nd derivative
2^{nd}_DIFF_WAV_MAX	Maximum value of the 2nd derivative
ENERGY_PERCENT_WAV	Percentage of the total energy of a detail/approximation
WAV_ZERO_CROSSING	Zero crossing
WAV_COEFF_OF_VARIATION	Coefficient of variation
WAV_TOTAL_ENERGY	Total Energy

4.4 Feature Optimization

As discussed in Chapter 2, aim of feature optimization is, simplification of models to make them easier to interpret by researchers/users, shorter training times, avoid the curse of dimensionality and enhanced generalization by reducing overfitting.

4.4.1 Minimal Redundancy and Maximum Relevance (mRMR)

In feature optimization/selection we identifies subsets of data that are relevant to the parameters selected and is known as *Maximum Relevance.* These parameters are sometimes redundant and mRMR removes this redundancy of subsets. mRMR has a variety of applications speech recognition, a time-series indeed.

Features that correlate strongest to the classification variable are selected. This is known as *maximum-relevance* selection. Many algorithms can be used to calculate this, like, the sequential forward, backward or floating selections. Also, features which are mutually far away from each but still having high degree of correlation to the classification variable are selected. This is known as Minimum Redundancy Maximum Relevance (mRMR). Here, the correlation could be the statistical dependency between variables. *Mutual information*, as explaine earlier also, can also be used to quantify the dependency between variables. It is shown that mRMR is an approximation to maximizing the dependency between the joint distribution of the selected features and the classification variable.

Mutual information

Let x and y be two random variables, the MI, in terms of probabilistic density functions is defined as $p(x),\ p(y)$ and $p(x,y)$:-

$$I(x,y) = \int\int p(x,y)\log\frac{p(x,y)}{p(x)p(y)}dxdy$$

Maximal relevance

Selecting the features with has highest relevance to target class c, based on MI, without considering relation between features is called maximal relevance. In terms of mutual information, the purpose of feature selection is to find a feature set S with m features x_i, which jointly have the largest dependency on the target class c

$$minR(S), R = \frac{1}{|S|^2}\sum_{x_i,x_j\ \epsilon\ S} I(x_i,x_j)$$

Combining the above two constraints we get what is called mRMR. We define the operator $max\Phi(D,R)$ to combine D and R and consider the following simplest form to optimize D and R simultaneously:-

$$max\Phi(D,R), \Phi = D - R$$

Incremental search methods can be used to find the near-optimal features defined by above equation. Suppose we already have S_{m-1}, the feature set with $m-1$ features. The task is to select the m^th feature from the set XS_{m-1}. This is done by selecting the feature that maximizes above equation. The respective incremental algorithm optimizes the following condition:-

$$\frac{max}{x_j\epsilon X - S_{m-1}}[I(x_j,c) - \frac{1}{m-1}\Sigma_{x_i\epsilon S_{m-1}}I(x_j,x_i)]$$

Epoching EEG Time Series

Before calculating features, we do epoching of EEG time series for every 3 seconds. Hence, we get an epoch of 14 $channels \times 128\ Hz \times 3\ seconds = 14 \times 382$ samples. We get 60 feature vectors per channel and total 840 feature vector per experiment. Size of feature matrix is 850 $\times$ number of features. After applying above algorithm to feature vector, it is observed that following optimal features were selected giving better or at par accuracy.

Table 4.7: Optimized features

FEATURE	**DESCRIPTION**
FD	Fractal Dimension
AR	Auto regression
1^{st}_DIFF_MEAN	Mean value of the first derivative of the signal
1^{st}_DIFF_MAX	Maximum value of the first derivative of the signal
2^{nd}_DIFF_MEAN	Mean value of the second derivative of the signal
2^{nd}_DIFF_MAX	Maximum value of the second derivative of the signal
HJORTH1	Ability
HJORTH2	Mobility $(\sigma_{x'}/\sigma_x)$
HJORTH3	Complexity $(\frac{\sigma_{x''}/\sigma_{x'}}{\sigma_{x'}/\sigma_x})$
FFT_DT_RATIO	DELTA / THETA
FFT_DA_RATIO	DELTA / ALPHA
FFT_TA_RATIO	THETA / ALPHA
FFT_DTA_RATIO	(DELTA + THETA) / ALPHA
WAV_COEFF_OF_VARIATION	Coefficient of variation
WAV_TOTAL_ENERGY	Total Energy

Chapter 5

Modelling for Cognitive Workload Measurement

5.1 Machine Learning

Machine learning is a branch of computer science that deals with the design of machines that learn to make predictions on data without being explicitly programmed. The formal definition of machine learning due to Mitchell 1998 is "*A* computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E". Machine learning algorithms are divided into three categories; supervised, semi-supervised and un-supervised. We have explored following machine learning algorithms for CWM estimation in our thesis:-

- k-Nearest Neighbors (k-NN)
- Random Forest
- Decision Tree Classifier
- Support Vector Machine (SVM)
- Multi Layer Perceptron (MLP)
- Linear Discriminant Analysis (LDA)

All above mentioned algorithms are supervised machine learning algorithms that learn from labeled data <**data,label**> pair. The problem of CWM is posed as a multi-class classification problem. In our experiments we classify 5 levels of cognitive workload described in previous chapters. We observe that machine learning methods are effective in distinguishing different levels of cognitive workload consistently.

We have used *sklearn* open library for executing our machine learning algorithms. Our labeled *feature vector* is fed as .CSV extension files obtained for various class of experiments conducted to all of classifiers above. In supervised learning, the data is organized in <**data,label**> format. Here,

$$\textbf{data} \Rightarrow Optimized\ feature\ vector\ set$$

$$\textbf{label} \Rightarrow Cognitive\ workload\ class$$

We label our various cognitive workload level as Ci, where $i = 1,\ 2,\ 3,\ 4,\ 5$. For example, for Dual *1*-Back task the class label is C4.

5.2 Modelling

Classification involves assigning a class to an unknown object. In the case of EEG signal processing, the objects are segments described by vectors of optimized features. Various algorithms for classifying brain states have been reported. We have explored following supervised learning algorithms. In all ML algorithms, we divide the complete dataset into 80% training set and remaining 20% as test set. Tuning of parameters for each classifier has been done carefully.

5.2.1 k-Nearest Neighbors (k-NN)

k-NN is a simple and robust classifier. The classifier works by comparing testing data with training data. The classifier finds the K neighborhood in the training data and assigns the class that appears most frequently in the neighbourhood of k. The default value of k is 1, and the default neighborhood object similarity setting is the Euclidean distance. In this experiment, the value of k is set to 5.

$$d(X_i, X_j) = \sqrt{\sum_i (X_i - X_j)^2}$$

5.2.2 Random Forest

Random forest (or random forests) is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the class's output by individual trees. The term came from random decision forests that was first proposed by Tin Kam

Ho of Bell Labs in 1995. The method combines Breiman's "bagging" idea and the random selection of features. Each tree is constructed using the following algorithm:-

1. Let the number of training cases be N, and the number of variables in the classifier be M.

2. We are told the number m of input variables to be used to determine the decision at a node of the tree; m should be much less than M.

3. Choose a training set for this tree by choosing n times with replacement from all N available training cases (i.e. take a bootstrap sample). Use the rest of the cases to estimate the error of the tree, by predicting their classes.

4. For each node of the tree, randomly choose m variables on which to base the decision at that node. Calculate the best split based on these m variables in the training set.

5. Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier). For prediction a new sample is pushed down the tree. It is assigned the label of the training sample in the terminal node it ends up in. This procedure is iterated over all trees in the ensemble, and the average vote of all trees is reported as random forest prediction.

5.2.3 Decision Tree Classifier

Decision tree learning is one of the most widely used and practical inductive inference method. The decision tree classifier is a method for approximating discrete-valued functions, in which the learned function is represented by a decision tree. Learned trees can also be represented as sets of if-then rules that may improve human readability. In addition, the results of decision tree classifiers can be directly compared with the rules used by physicians when evaluating EEG.

5.2.4 Support Vector Machine (SVM)

SVMs, also called support vector networks are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. We have used both linear and RBF kernel in our algorithm.

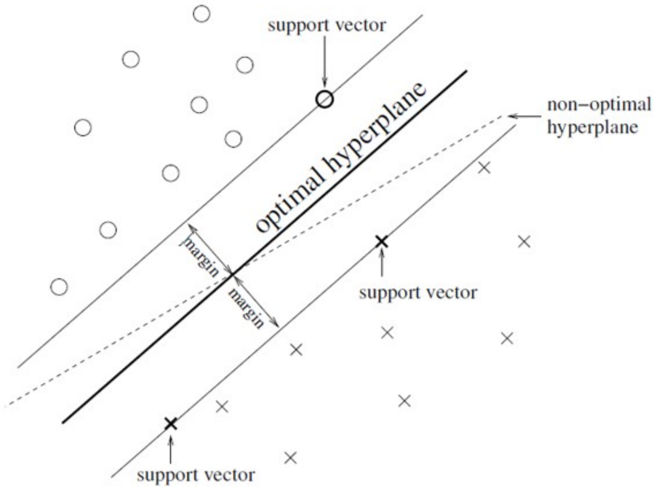


Figure 5.1: Support vector machine.

5.2.5 Multi Layer Perceptron (MLP)

The most commonly used representative of ANNs (Artificial Neural Network) is the MLP. MLP is the most commonly used feedforward ANN, due to ease of implementation and smaller training set requirements. The MLP model maps sets of input data on to a set of appropriate output.

MLP can solve complex classification tasks, but it risks overfitting the training data and it takes a lot of time. The diagram shown in Fig illustrates a MLP network with hidden layers. Neural Networks and thus MLP, are universal approximators, i.e., when composed of enough neurons and layers, they can approximate any continuous function. Added to the fact that they can classify any number of classes, this makes NN very flexible classifiers that can adapt to a great variety of problems. A MLP without hidden layers is known as a perceptron. Interestingly enough, a perceptron is equivalent to LDA and, as such, has been sometimes used for BCI applications.

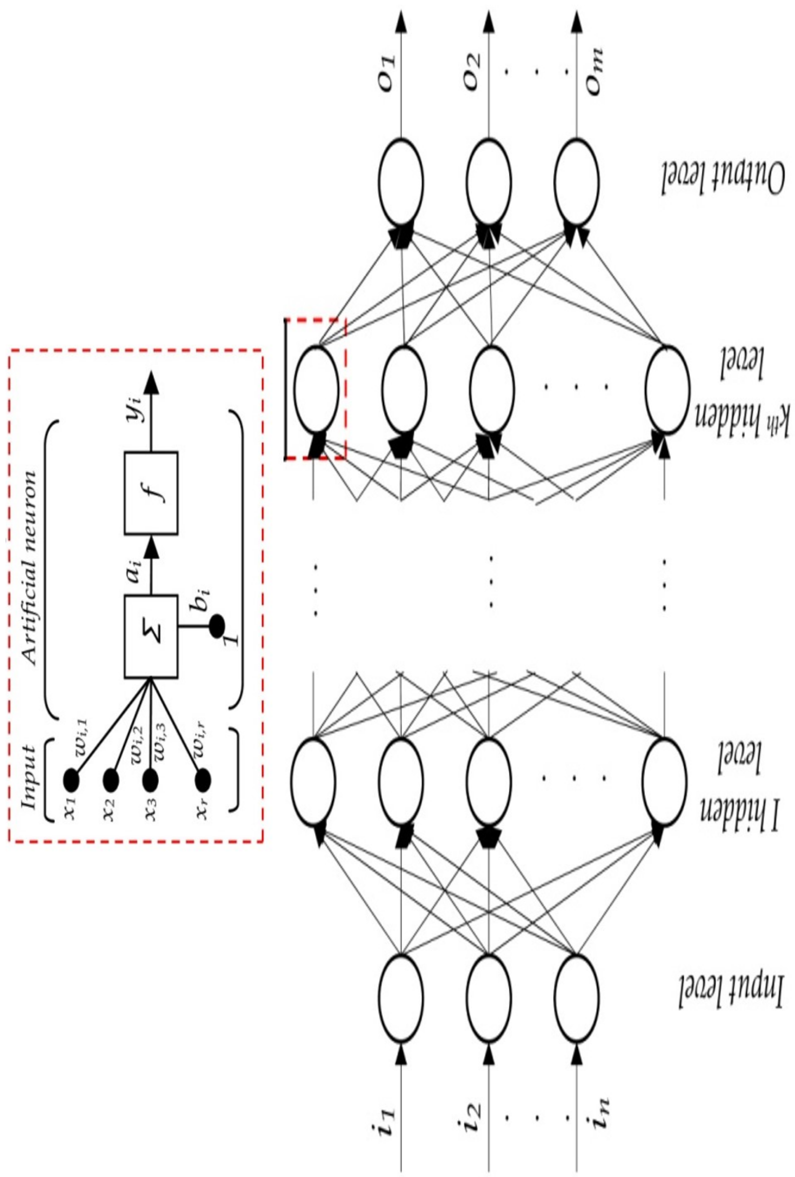


Figure 5.2: Multi layer perceptron.

5.2.6 Linear Discriminant Analysis (LDA)

LDA also known as Fisher's linear discriminant, main purpose of LDA is to employ hyperplanes to distinguish the data representing various classes. For a problem consisting of two-class, the class of a feature vector relays on which side of the hyperplane the vector is.

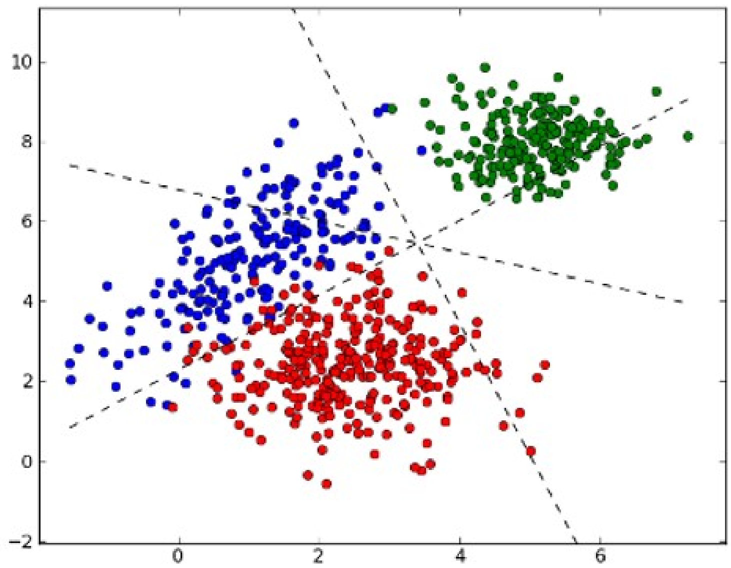


Figure 5.3: Linear discriminant analysis.

This technique has very less mathematical calculations which makes it suitable for online BCI systems. They are simple to use and generally provides good results. The main drawback of LDA is their poor results on complex nonlinear EEG signals. In a regularized Fisher's LDA (RFLDA), a regularization parameter C is introduced. This parameter can penalize classification errors that tend to occur on training set. The resulting can accommodate outliers. They are less used than LDA for BCI applications.

Chapter 6

Results, Conclusion and Future Work

This chapter includes exhaustive results for various permutation and combinations of classes for cognitive task and type of subjects. We conclude our thesis with the observations noticed and future plan of work.

6.1 Results before Channel Selection and Feature Optimization

The experimental results are calculated for male and female subjects separately and together as well. The idea behind doing this is to establish some facts and conclusions for better understanding the cognitive workload generation in human beings.

6.1.1 Male Subjects

Classifier results for EEG signals collected from five male subjects as discussed in Chapter 3 are presented in the form of *table* and *bar-chart* for ease of understanding the data. The tables and bar-charts have been divided in three categories:- Firstly for various combinations of *two-class*, secondly for *three-class* and thirdly for *four-class* and *five-class*. The summary of *Maximum* and *Minimum* accuracies achieved are given along with tables.

Table 6.1: % Accuracy : 2 Class (Male Subjects)

CLASS	CLASSIFIER						
Two Class	**k-NN**	**R Forest**	**SVM (RBF)**	**SVM (LINEAR)**	**MLP**	**D Tree**	**LDA**
C1-C2	88.00	99.19	92.00	89.00	96.79	95.99	94.39
C1-C3	92.00	95.19	91.00	95.00	94.39	88.00	98.39
C1-C4	90.40	92.00	92.00	96.00	96.79	85.59	99.19
C1-C5	88.70	94.35	93.00	92.00	94.35	87.09	90.32
C2-C3	76.00	90.40	88.00	87.00	90.40	80.80	83.19
C2-C4	84.79	94.39	85.00	86.00	88.80	91.20	86.39
C2-C5	81.45	93.54	83.00	86.00	87.90	88.70	86.29
C3-C4	72.79	76.80	61.00	67.00	71.99	64.80	66.40
C3-C5	75.80	70.16	70.00	65.00	71.77	66.93	68.54
C4-C5	70.96	81.45	76.00	74.00	77.41	72.58	73.38
						Max	99.19
						Min	61.00

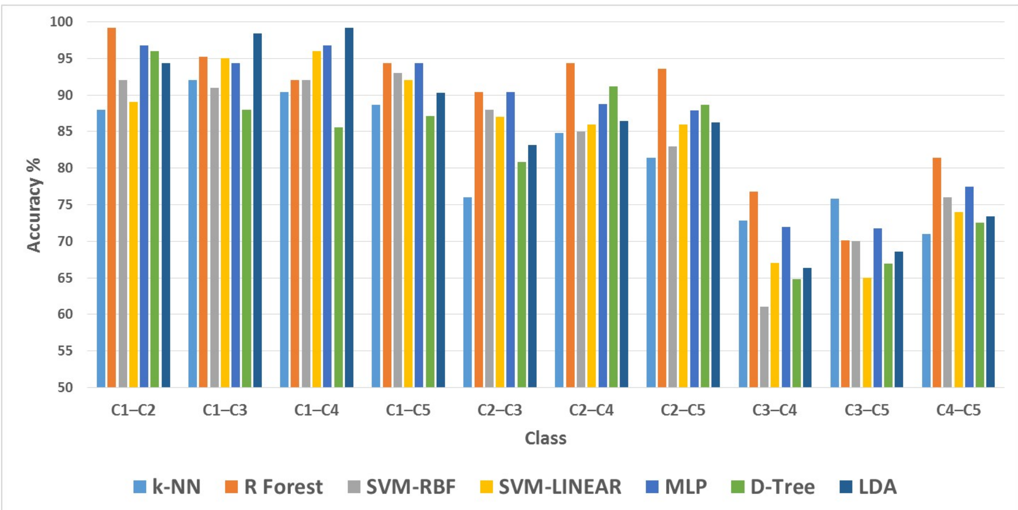


Figure 6.1: % Accuracy chart : 2 Class (Male subjects).

Table 6.2: % Accuracy : 3 Class (Male Subjects)

CLASS	CLASSIFIER						
Three Class	**k-NN**	**R Forest**	**SVM (RBF)**	**SVM (LINEAR)**	**MLP**	**D Tree**	**LDA**
C1-C2-C3	75.00	88.29	80.00	85.00	86.70	79.78	86.17
C1-C3-C5	77.00	77.54	75.00	74.00	80.74	70.58	74.33
C2-C3-C4	77.00	78.07	75.00	74.00	80.74	69.51	74.33
C3-C4-C5	60.42	64.70	55.00	55.00	57.21	56.14	57.75
						Max	88.29
						Min	55.00

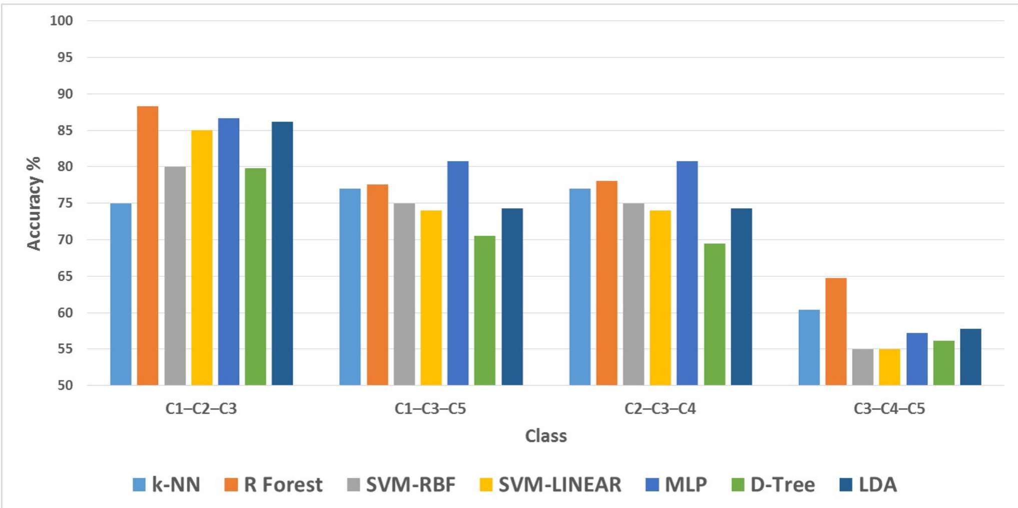


Figure 6.2: % Accuracy chart : 3 Class (Male subjects).

Table 6.3: % Accuracy : 4 and 5 Class (Male Subjects).

CLASS	CLASSIFIER						
Four Class	**k-NN**	**R Forest**	**SVM (RBF)**	**SVM (LINEAR)**	**MLP**	**D Tree**	**LDA**
C1-C2-C3-C4	64.00	75.20	63.00	72.00	78.00	66.00	75.20
C2-C3-C4-C5	57.42	70.28	53.00	63.00	63.45	58.23	60.24
C1-C2-C4-C5	67.06	78.71	71.00	77.00	80.32	68.67	72.28
						Max	80.32
						Min	53.00
Five Class							
C1-C2-C3-C4-C5	56.08	73.07	57.00	62.00	65.06	60.89	62.50
						Max	73.07
						Min	56.08

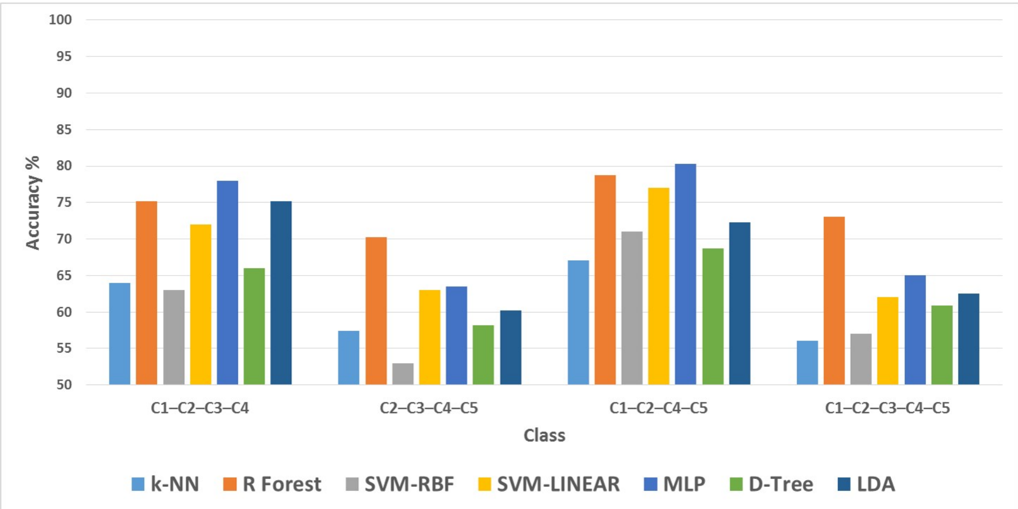


Figure 6.3: % Accuracy chart : 4 and 5 Class (Male subjects).

6.1.2 Female Subjects

Classifier results for EEG signals collected from five female subjects as discussed below:-

Table 6.4: % Accuracy : 2 Class (Female Subjects)

CLASS	**CLASSIFIER**						
Two Class	**k-NN**	**R Forest**	**SVM (RBF)**	**SVM (LINEAR)**	**MLP**	**D Tree**	**LDA**
C1-C2	85.84	96.46	92.00	95.00	94.69	88.49	86.72
C1-C3	89.60	88.80	90.00	90.00	94.39	82.39	91.20
C1-C4	85.59	93.60	92.00	92.00	93.60	80.00	90.40
C1-C5	77.60	94.39	78.00	79.00	80.80	84.79	79.20
C2-C3	76.99	80.53	77.00	82.00	84.07	74.33	86.72
C2-C4	82.30	91.15	85.00	89.00	93.80	77.87	90.26
C2-C5	76.99	93.80	80.00	89.00	89.38	83.18	85.84
C3-C4	80.00	84.79	75.00	76.00	83.19	75.20	80.80
C3-C5	85.59	87.20	81.00	83.00	83.99	83.19	86.39
C4-C5	82.39	86.39	76.00	82.00	86.39	71.19	80.80
						Max	96.46
						Min	71.19

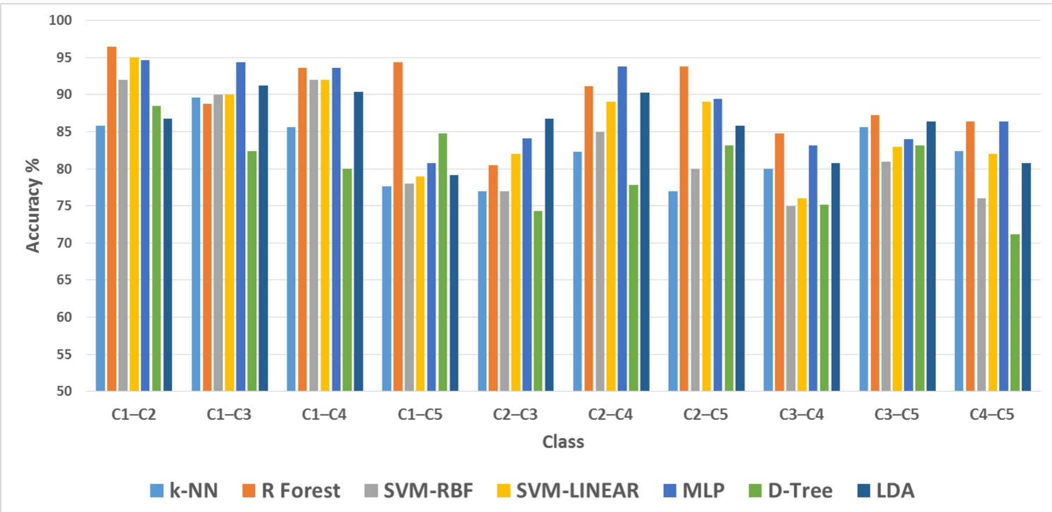


Figure 6.4: % Accuracy chart : 2 Class (Female subjects).

Table 6.5: % Accuracy : 3 Class (Female Subjects)

CLASS	**CLASSIFIER**						
Three Class	**k-NN**	**R Forest**	**SVM (RBF)**	**SVM (LINEAR)**	**MLP**	**D Tree**	**LDA**
C1-C2-C3	76.70	86.36	81.00	79.00	86.93	73.86	83.52
C1-C3-C5	71.65	83.95	75.00	73.00	81.81	70.05	75.93
C2-C3-C4	71.02	81.25	71.00	76.00	76.70	63.06	81.25
C3-C4-C5	71.02	82.95	71.00	76.00	76.70	55.68	81.25
						Max	86.93
						Min	55.68

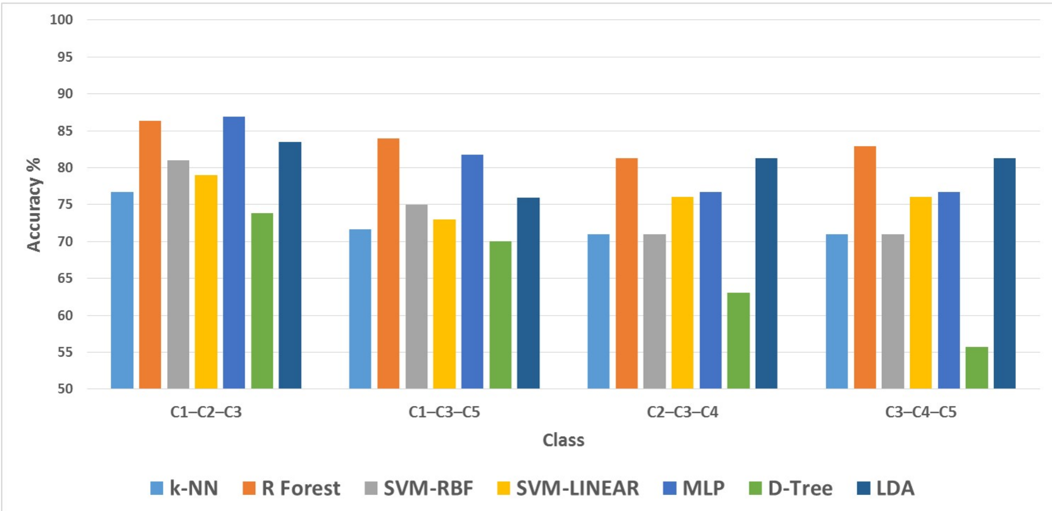


Figure 6.5: % Accuracy chart : 3 Class (Female subjects).

Table 6.6: % Accuracy : 4 and 5 Class (Female Subjects)

CLASS	CLASSIFIER						
Four Class	**k-NN**	**R Forest**	**SVM (RBF)**	**SVM (LINEAR)**	**MLP**	**D Tree**	**LDA**
C1-C2-C3-C4	68.06	77.73	71.00	73.00	75.63	65.12	79.83
C2-C3-C4-C5	71.42	83.61	71.00	78.00	80.67	70.16	80.25
C1-C2-C4-C5	65.96	82.35	63.00	70.00	73.10	65.54	74.78
						Max	83.61
						Min	63.00
Five Class							
C1-C2-C3-C4-C5	61.12	76.07	67.00	71.00	75.74	57.80	74.75
						Max	76.07
						Min	57.80

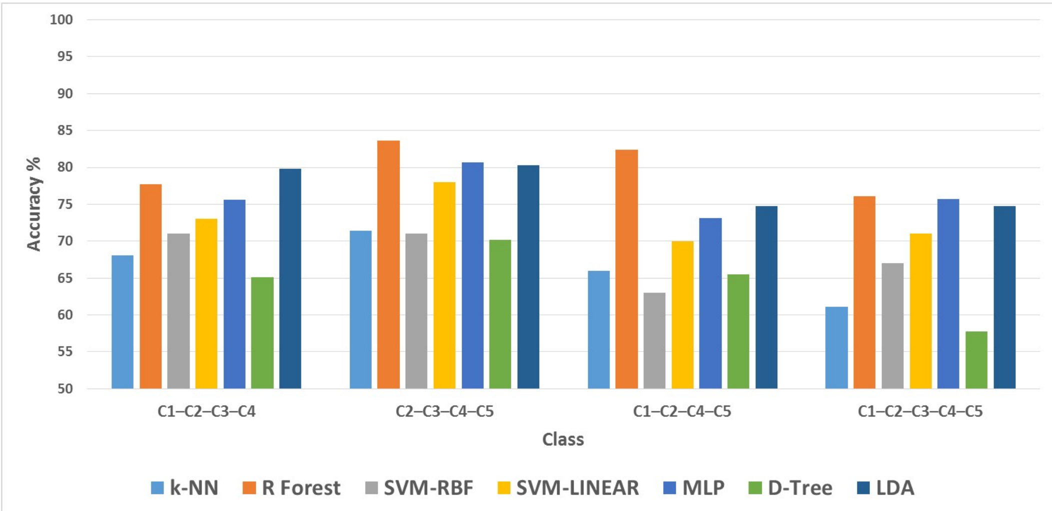


Figure 6.6: % Accuracy chart : 4 and 5 Class (Female subjects).

It is strictly observed that the % accuracy of all classifiers is lower as compared to those of male subjects in respective classes. This fact is established as, while recording the data, it was noticed that female subjects could only score 80% score in Dual *1*-Back and Dual *2*-Back tasks, where as, all male subjects scored 100% on all combinations of n-Back and Dual n-Back task.

6.1.3 Male and Female Subjects

Classifier results for EEG signals collected from five male and female subjects as discussed below:-

Table 6.7: % Accuracy : 2 Class (Male and Female Subjects)

CLASS	CLASSIFIER						
Two Class	**k-NN**	**R Forest**	**SVM (RBF)**	**SVM (LINEAR)**	**MLP**	**D Tree**	**LDA**
C1-C2	86.55	97.47	89.00	94.00	94.95	87.39	92.01
C1-C3	90.80	97.99	94.00	89.00	96.39	90.80	90.40
C1-C4	92.40	96.39	94.00	92.00	94.39	90.00	91.20
C1-C5	89.55	95.58	91.00	86.00	93.57	91.16	85.14
C2-C3	81.09	89.91	83.00	86.00	89.70	77.31	81.51
C2-C4	92.85	95.37	88.00	89.00	93.69	86.55	92.43
C2-C5	84.81	92.40	84.00	91.00	92.82	81.85	88.60
C3-C4	77.60	84.39	74.00	72.00	82.39	73.99	78.40
C3-C5	81.92	86.74	75.00	75.00	82.32	73.49	79.91
C4-C5	83.13	86.34	80.00	84.00	88.35	77.91	83.93
						Max	97.99
						Min	72.00

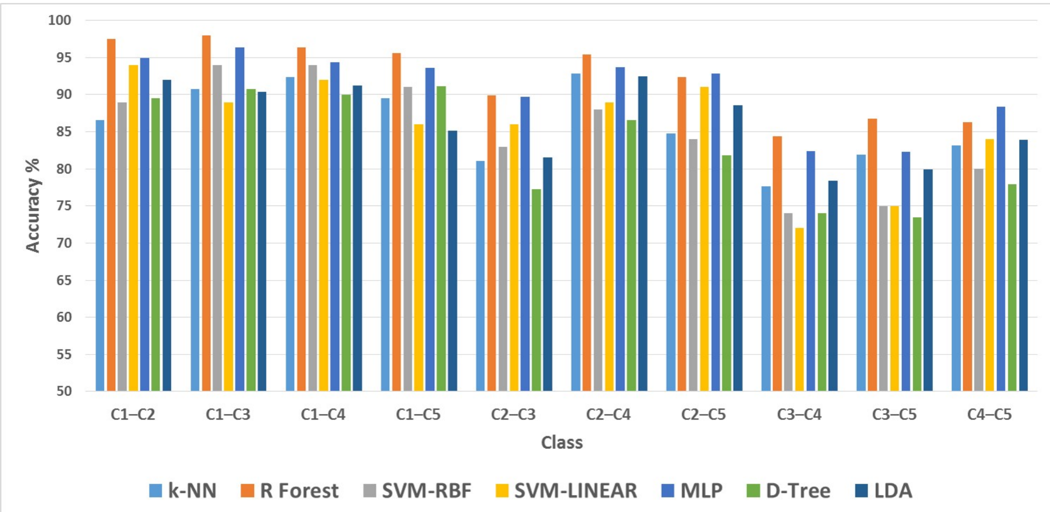


Figure 6.7: % Accuracy chart : 2 Class (Male and Female subjects).

Table 6.8: % Accuracy : 3 Class (Male and Female Subjects)

CLASS	CLASSIFIER						
Three Class	**k-NN**	**R Forest**	**SVM (RBF)**	**SVM (LINEAR)**	**MLP**	**D Tree**	**LDA**
C1-C2-C3	77.41	91.46	73.00	82.00	84.29	79.06	82.09
C1-C3-C5	76.47	87.96	73.00	75.00	82.08	68.71	71.92
C2-C3-C4	66.94	79.33	62.00	70.00	75.75	67.49	70.79
C3-C4-C5	66.84	82.62	65.00	66.00	71.65	66.04	67.11
						Max	91.46
						Min	62.00

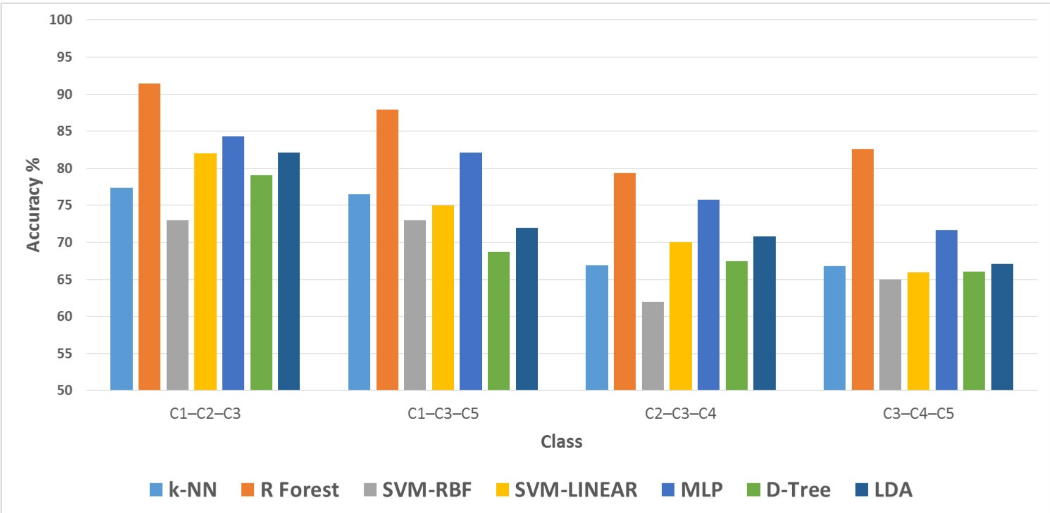


Figure 6.8: % Accuracy chart : 3 Class (Male and Female subjects).

Table 6.9: % Accuracy : 4 and 5 Class (Male and Female Subjects)

CLASS	CLASSIFIER						
Four Class	**k-NN**	**R Forest**	**SVM (RBF)**	**SVM (LINEAR)**	**MLP**	**D Tree**	**LDA**
C1-C2-C3-C4	65.98	81.55	64.00	68.00	76.02	67.82	71.31
C2-C3-C4-C5	67.14	79.87	60.00	63.00	72.07	60.36	62.42
C1-C2-C4-C5	71.66	86.44	70.00	75.00	82.54	68.58	76.79
						Max	86.44
						Min	60.00
Five Class							
C1-C2-C3-C4-C5	61.43	80.22	57.00	62.00	68.13	58.33	63.39
						Max	80.22
						Min	57.00

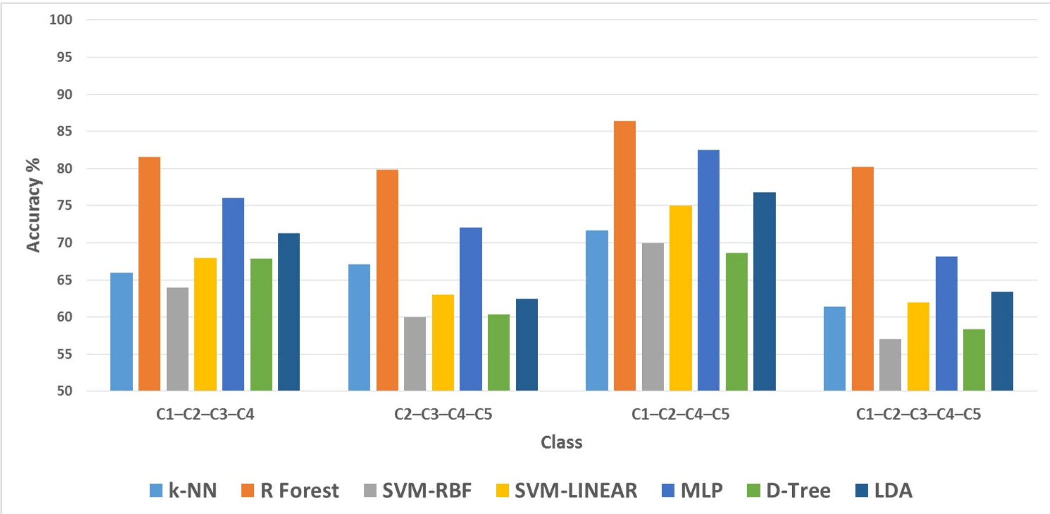


Figure 6.9: % Accuracy chart: 4 and 5 Class (Male and Female subjects).

6.2 Results after Channel Selection and Feature Optimization

Results after applying *Channel Selection* and *Feature Optimization* to EEG data for only male subjects are presented below:-

Table 6.10: % Accuracy : 2 Class (Male Subjects)

CLASS	**CLASSIFIER**						
Two Class	**k-NN**	**R Forest**	**SVM (RBF)**	**SVM (LINEAR)**	**MLP**	**D Tree**	**LDA**
C1-C2	95.19	100	95.00	88.00	95.99	92.00	95.19
C1-C3	93.60	95.19	96.00	95.00	94.39	91.20	96.79
C1-C4	95.19	93.60	97.00	97.00	99.19	88.00	97.59
C1-C5	94.35	95.96	94.00	92.00	96.77	87.90	91.12
C2-C3	83.19	92.80	82.00	94.00	86.39	85.59	91.20
C2-C4	95.19	96.79	90.00	92.00	92.00	91.20	90.40
C2-C5	86.29	94.35	88.00	95.00	93.54	84.67	91.12
C3-C4	81.59	83.99	75.00	76.00	80.00	74.39	80.80
C3-C5	79.03	88.70	78.00	76.00	76.61	75.00	76.61
C4-C5	80.64	87.90	82.00	87.00	83.06	81.45	83.87
						Max	100
						Min	74.39

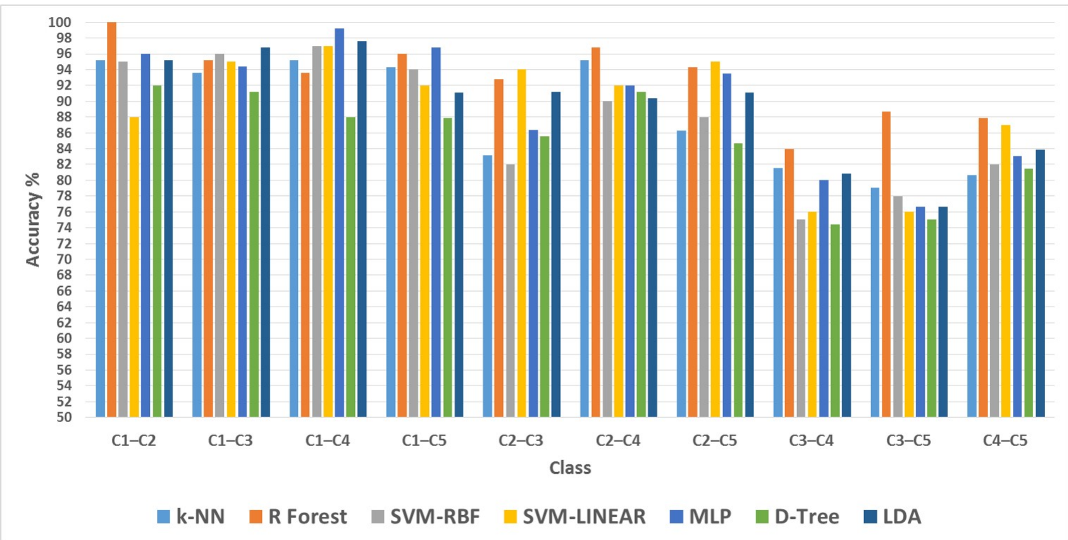


Figure 6.10: % Accuracy chart : 2 Class (Male subjects)

Table 6.11: % Accuracy : 3 Class (Male Subjects)

CLASS	CLASSIFIER						
Four Class	**k-NN**	**R Forest**	**SVM (RBF)**	**SVM (LINEAR)**	**MLP**	**D Tree**	**LDA**
C1-C2-C3	83.51	91.48	86.00	89.00	88.82	82.97	86.70
C1-C3-C5	81.28	86.63	83.00	83.00	87.16	75.40	86.09
C2-C3-C4	75.00	91.48	71.00	78.00	80.85	82.44	77.65
C3-C4-C5	66.31	79.67	66.00	69.00	76.47	65.77	74.33
						Max	91.48
						Min	65.77

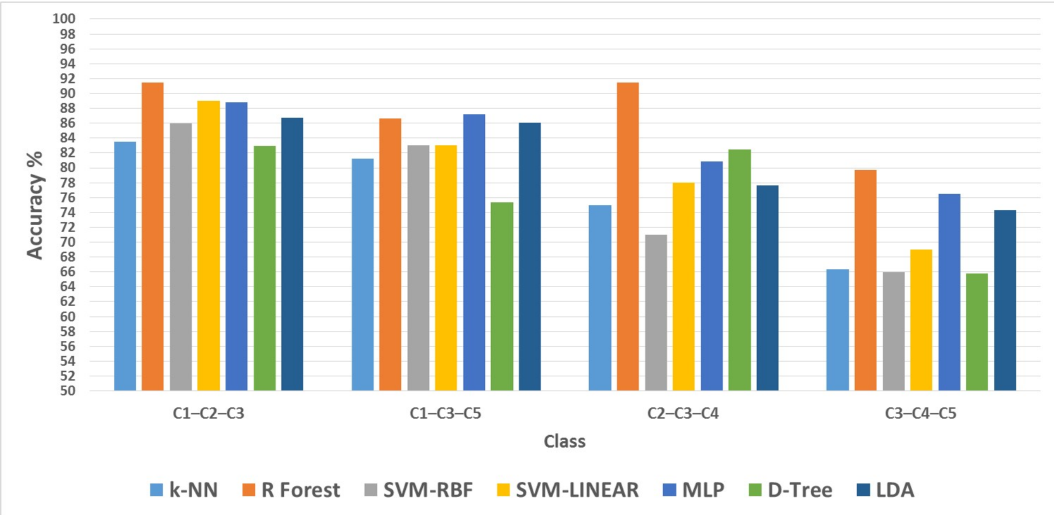


Figure 6.11: % Accuracy chart : 3 Class (Male subjects).

Table 6.12: % Accuracy : 4 and 5 Class (Male Subjects)

CLASS	CLASSIFIER						
Four Class	**k-NN**	**R Forest**	**SVM (RBF)**	**SVM (LINEAR)**	**MLP**	**D Tree**	**LDA**
C1-C2-C3-C4	71.59	87.60	73.00	77.00	85.19	77.20	79.20
C2-C3-C4-C5	67.87	82.73	64.00	73.00	73.09	69.07	78.71
C1-C2-C4-C5	74.29	90.76	74.00	84.00	89.15	73.89	82.32
						Max	90.76
						Min	64.00
Five Class							
C1-C2-C3-C4-C5	64.42	84.61	63.00	73.00	79.80	69.55	78.20
						Max	84.61
						Min	63.00

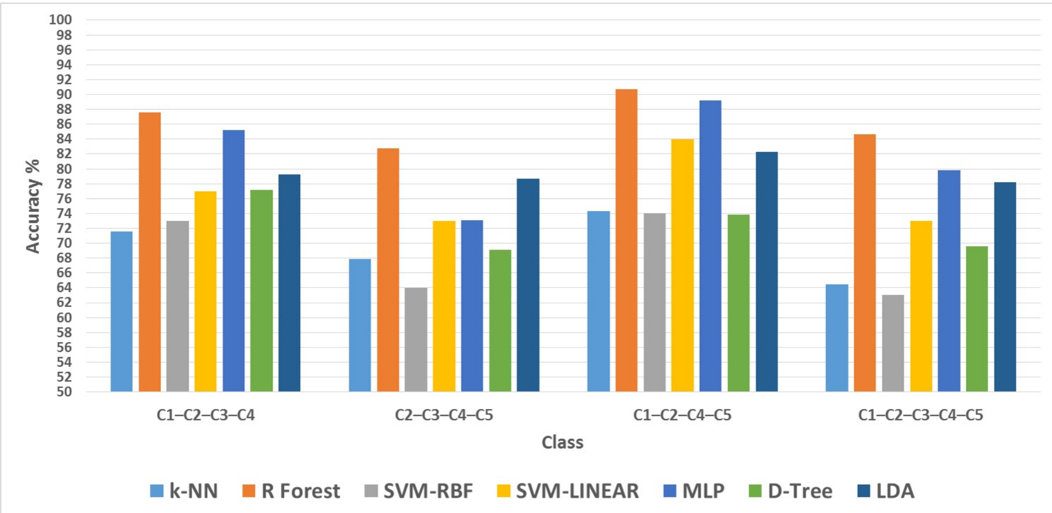


Figure 6.12: % Accuracy chart : 4 and 5 Class (Male subjects).

6.3 Confusion Matrix

In ML, a confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one. Each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class (or vice versa). The diagonal elements represent the number of points for which the predicted label is equal to the true label, while off-diagonal elements are those that are mislabeled by the classifier. The higher the diagonal values of the confusion matrix the better, indicating many correct predictions. We present confusion matrix here for selected classes of cognitive workload for male subjects in two category:-

Confusion matrix before channel selection and feature optimization:-

Confusion matrix for male subjects before *channel selection* and *feature optimization* techniques are applied to EEG data for selected cognitive workload classes as shown in Fig below.

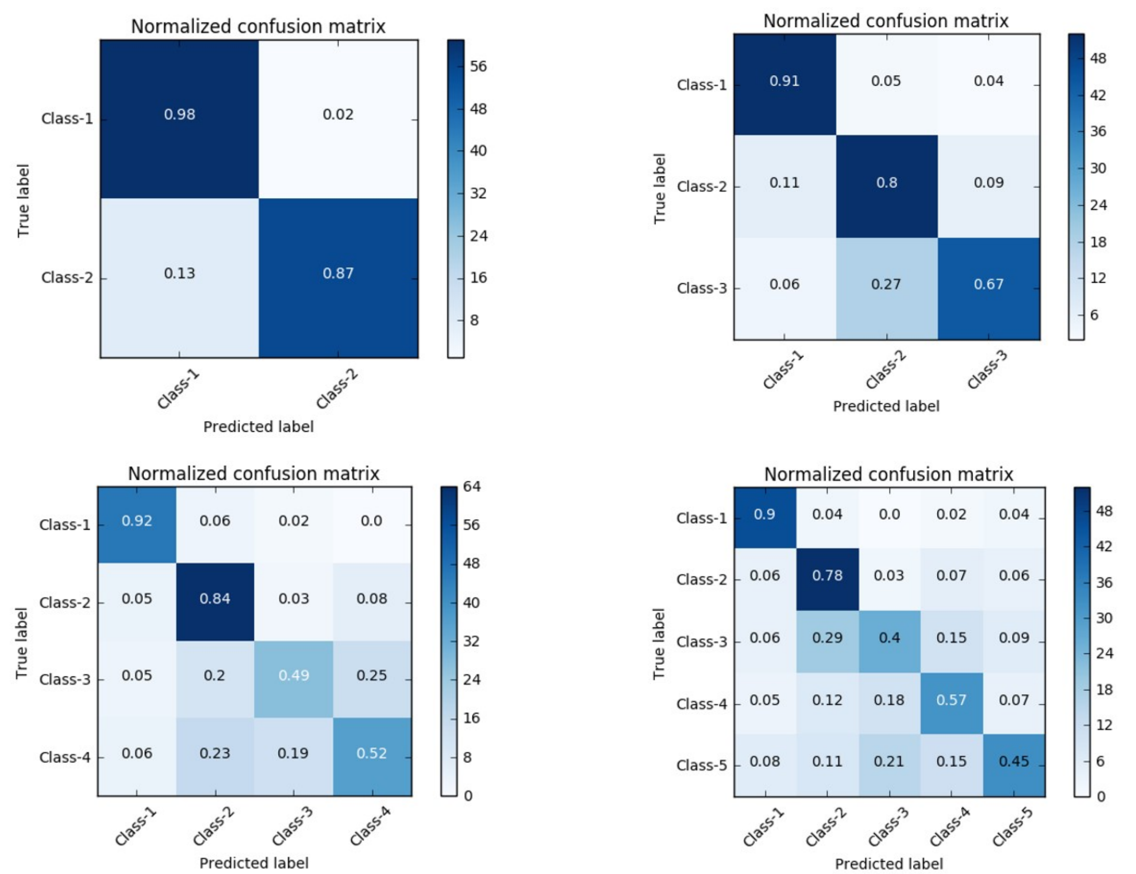


Figure 6.13: Confusion matrix before channel selection and feature optimization.

Confusion matrix after channel selection and feature optimization:-

Confusion matrix for male subjects after *channel selection* and *feature optimization* techniques are applied to EEG data for selected cognitive workload classes as shown in Fig below.

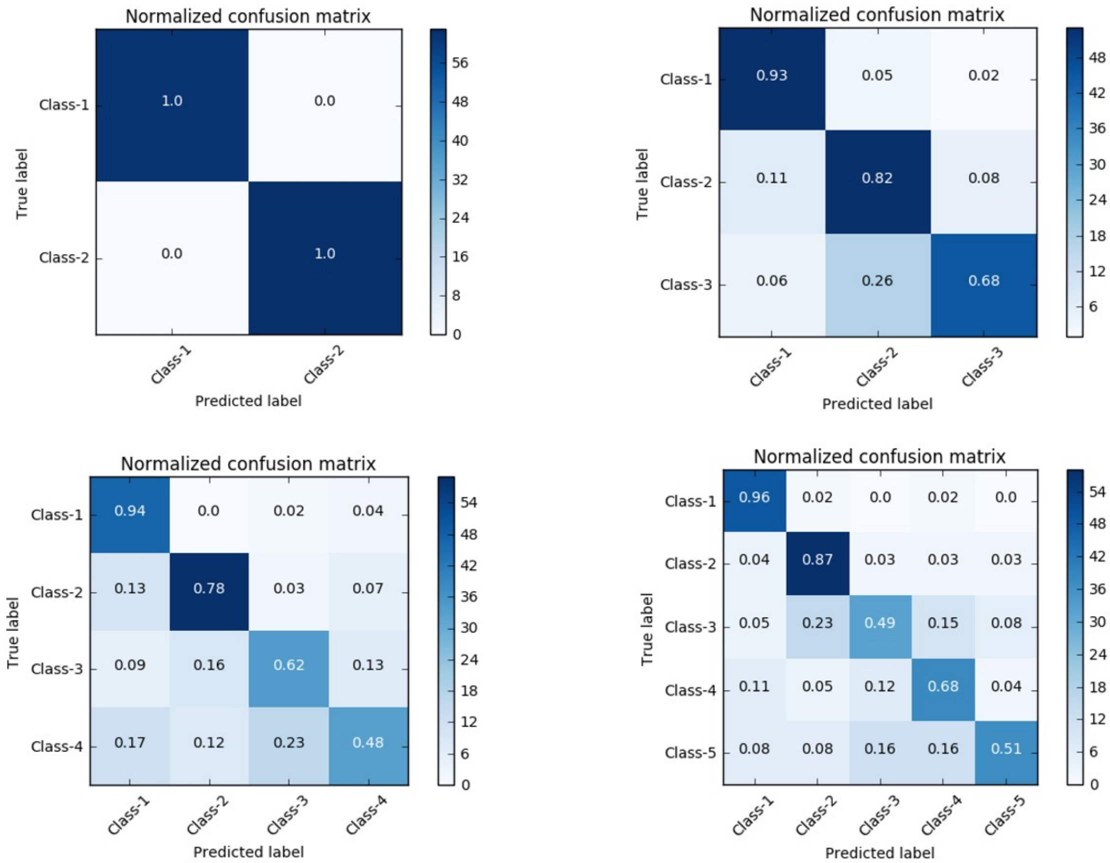


Figure 6.14: Confusion matrix after channel selection and feature optimization.

6.4 Spectrogram

We plot spectrogram for all EEG wave bands for subject-M05. Spectrogram is plotted for all five level of cognitive task. It can be clearly seen that theta and alpha waves activity are the most dominant. Most of the power is confined to theta and alpha band of EEG signal. However, due to increase in cognitive workload, the beta band activity rises as clearly seen in the spectrogram for Dual *1*-Back and Dual *2*-Back task.

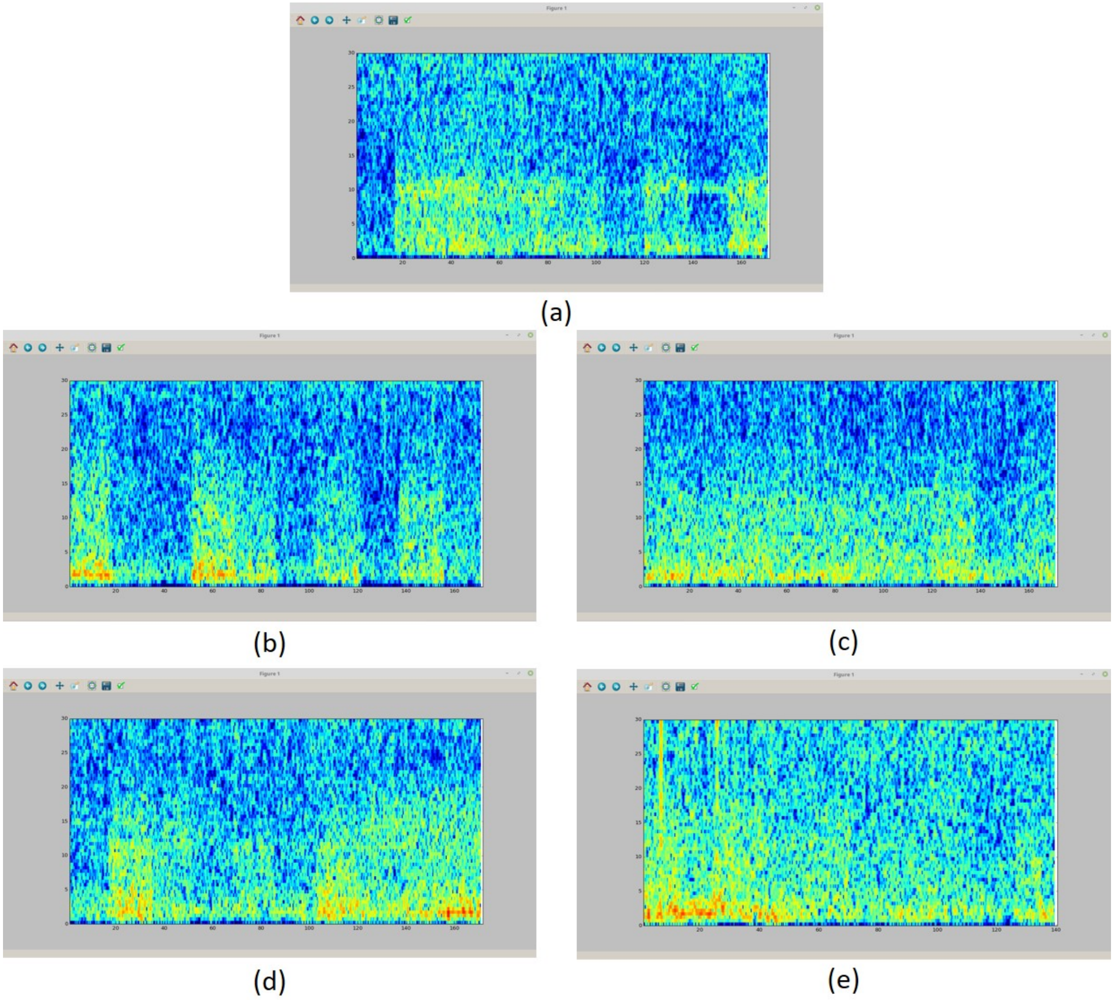


Figure 6.15: Spectrogram : (a) Ideal, (b) *1*-Back, (c) *2*-Back, (d) Dual *1*-Back, (d) Dual *2*-Back tasks.

6.5 Conclusion and Future Work

Exhaustive results, both in form of tables and charts, as shown in previous section gives us a detailed intuition of cognitive workload and how it increases as the demand of cognitive resources increases as we migrate from Ideal task up till Dual *2*-Back task. Our thesis has successfully established the cognitive workload load theory by measuring cognitive load generated in various experiments designed in Chapter 3. This study marks a humble step in understanding the human cognition. Considering the task demand in terms of cognitive resources is gaining important in various critical mission related to Armed Forces and Space technologies.

We have studied the gradual increase in cognitive workload as per the experiment designed in Chapter 3. However, we can extend the research, by replacing the experiments conducted in controlled laboratory condition with practical tasks. For example, we can expose a combatant to practical military training and have the EEG signal recorded through EEG enabled helmets. We can do the online analysis of the cognition, which can be very useful for various kinds of applications as discussed next.

6.5.1 Proposed Applications for Cognitive Workload Measurement

Following are few important military and non-military applications which can use CWM study in this thesis.

Military Applications

- Selection of Combatant for Critical Mission
- Monitoring Combatant in Critical Mission
- Cognitive Workload Management of Pilot

Non-Military Applications

- HR Management using EEG
- Improving working Environment based on Workload Measurement
- Monitoring Employees in a Company

6.5.2 Proposal for Future Work

In the field of Deep Learning (DL), Convolutional Neural Network (CNN) has proved its worth in image recognition. We propose to use DL for CWM. Since CNN can handle two dimensional images precisely, it is proposed to convert the time-series EEG signal to two dimensional EEG images.

We can plot 2D colored heat map of brain by calculating the Power Spectral Density (PSD) in various bands for every channel. This way, we can convert 14 channel EEG data into plenty of images. After this, the database of 2D EEG images can be fed to CNN directly without calculating the features and hence saving effort and time to calculate features and then do optimization. RNNs has memory and hence they may be used in conjunction with CNN to retain the time dependence of continuous EEG signal. Of course, the final layer consists of fully connected layer as per the class of cognitive workload desired. The entire idea is proposed in Fig 6.16.

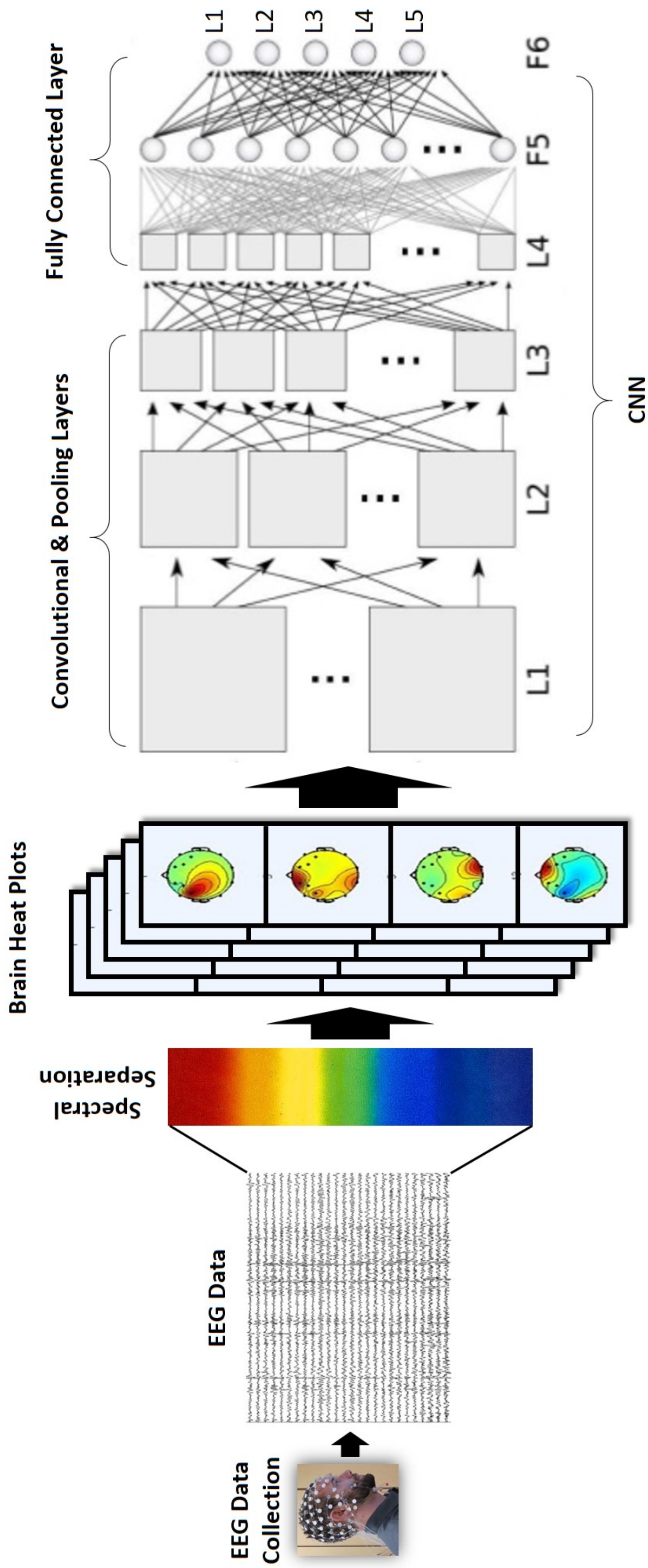


Figure 6.16: Proposed deep-learning architecture.

Publications

1. **Published:-**

   - **Title** : "BCI Augmented Text Entry Mechanism for People with Special Needs".
   - **Conference** : "Intelligent Human Computer Interaction", IHCI 2016
   - **Publisher** : Springer
   - **DOI** : 10.1007/978-3-319-52503-7_7
   - **Link** : https://link.springer.com/chapter/10.1007/978-3-319-52503-7_7

2. **To be communicated:-**

   - **Title** : "Cognitive Workload Measurement using Brain Computer Interface (BCI)".

Bibliography

[1] "http://serendip.brynmawr.edu/bb/kinser/structure1.html."

[2] "https://en.wikipedia.org/wiki/electroencephalography."

[3] Loft S., Sanderson P., Neal A., Mooij M., "Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications," in *Human Factors*, 2007, pp. 376–399.

[4] Paas F., Renkl A., Sweller J., "Cognitive load theory and instructional design: Recent developments," in *Educational Psychologist*, 2003a, pp. 1–4.

[5] Sweller J., "Cognitive load during problem solving: Effects on learning," in *Cognitive Science*, 1988, pp. 257–285.

[6] Schnotz W., Kurschner C., "A Reconsideration of Cognitive Load Theory," in *Educational Psychology Review*, 2007, pp. 469–508.

[7] Edith Galya, Magali Carioub, Claudine Melanb, "What is the relationship between mental workload factors and cognitive load types?" in *International Journal of Psychophysiology*, 2012, pp. 269–275.

[8] Sandra G., Hart Lowell E., Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Advances in Psychology*, 1988, pp. 139–183.

[9] Reid G. B., Nygren T.E., "The subjective workload assessment technique: A scaling procedure for measuring mental workload," in *Human mental workload*, 1988, pp. 185–218.

[10] C. F. Chi, F. T. Lin, "A new method for describing search patterns and quantifying visual load using eye movement data," in *International Journal of Industrial Ergonomics*, 1997, pp. 249–257.

[11] Shouyi Wang, Jacek Gwizdka, W. Art Chaovalitwongse, "Using Wireless EEG Signals to Assess Memory Workload in the n-Back Task," in *IEEE Transactions on Human-Machine Systems*, Jun 2016.

[12] "https://en.wikipedia.org/wiki/electroencephalography."

[13] Gevins A, Smith ME, McEvoy L, Yu D, "High resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice," in *Cereb Cortex*, 1997, pp. 374–385.

[14] "https://en.wikipedia.org/wiki/brain_computer_interface."

[15] "https://en.wikipedia.org/wiki/Eriksen_flanker_task."

[16] Gianclaudio Casutt, Mike Martin, Lutz Jancke, "Driving Simulator Training Is Associated with Reduced Inhibitory Workload in Older Drivers," in *MDPI*, 2016.

[17] Farshid Anvari, Hien Minh Thi Tran, Manolya Kavakli, "Using Cognitive Load Measurement and Spatial Ability Test to Identify Talented Students in Three-Dimensional Computer Graphics Programming," in *International Journal of Information and Education Technology*, 2013, pp. Vol. 3, No. 1.

[18] Gwizdka J., "Using Stroop Task to Access Cognitive Load," in *Proceedings of the 2010 European Conference on Cognitive Ergonomics*, 2010.

[19] "https://en.wikipedia.org/wiki/N-back."

[20] T. S. Braver, J. D. Cohen, L. E. Nystrom, J. Jonides, E. E. Smith, D. C. Noll, "A parametric study of prefrontal cortex involvement in human working memory," in *Neuroimage*, 1997, pp. 4–62.

[21] S. Ravizza, M. Behrmann, J. Fiez, "Right parietal contributions to verbal working memory: Spatial or executive?" in *Neuropsychologia*, 2005, pp. 2057–2067.

[22] Ali Bashashati, Mehrdad, Fatourechi, Rabab K Ward and Gary E Birch, "A survey of signal processing algorithms in brain?computer interfaces based on electrical brain signals," in *Journal of Neural Engineering*, 2007, p. Vol. 4.

[23] Arnaud Delorme, Scott Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," in *Journal of Neuroscience Methods*, 2005, p. Vol. 134.

[24] Josep Mourino, Jose del R. Millan, Febo Cincotti, Silvia Chiappa, Raimon Jane, Fabio Babiloni, "Spatial Filtering in the Training Process of A Brain Computer Interface," in *IEEE Engineering in Medicine and Biology*, 2001.

[25] M. Rajya Lakshmi, Dr. T. V. Prasad, Dr. V. Chandra Prakash, "Survey on EEG Signal Processing Methods," in *International Journal of Advanced Research in Computer Science and Software Engineering*, Jan 2014.

[26] Claudio G. Carvalhaes, Patrick Suppes, "A Spline Framework for Estimating the EEG Surface Laplacian Using the Euclidean Metric," in *Neural Computation*, 2010.

[27] "http://en.wikipedia.org/wiki/principal_component_analysis."

[28] Tomas Zeman, "BSS - Preprocessing Steps for Separation Improvement," 2000.

[29] Dieter Devlaminck, BartWyns, Moritz Grosse-Wentrup,Georges Otte, Patrick Santens, "Multisubject Learning for Common Spatial Patterns in Motor-Imagery BCI," in *Computational Intelligence and Neuroscience*, 2011.

[30] Chinmay Chandrakar, M.K. Kowar, "De-noising ECG Signals Using Adaptive Filter Algorithm," in *Int. J. of Soft Computing and Engg. (IJSCE)*, 2012, pp. Vol. 2, Issue 1.

[31] Turky Alotaiby, Fathi E Abd El-Samie, Saleh A Alshebeili, Ishtiaq Ahmad, "A review of channel selection algorithms for EEG signal processing," in *EURASIP Journal on Advances in Signal Processing*, 2015.

[32] M Dash, H Liu, "Feature selection for classification," in *Intelligent Data Analytics*, 1997, pp. 131–156.

[33] H Liu, L Yu, "Toward integrating feature selection algorithms for classification and clustering," in *IEEE Transaction, Knowledge Data Engineering*, 2005, pp. 491–502.

[34] Vidya Devi M., Thippeswamy G., "A Survey on Signal Acquistion and Feature Extraction Techniques," in *National Conference on ?Recent Trends in Information Technology*, 2016.

[35] Garrett D., Peterson D. A., Anderson, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," in *IEEE Transaction On Rehabilitation Engineering*, 2003, pp. 141–144.

[36] Fabien Lotte, “A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain-Computer Interfaces,” in *Guide to Brain-Computer Music Interfacing*, 2014.

[37] “https://www.emotiv.com.”

[38] “http://brainworkshop.sourceforge.net.”

[39] Ian Daly, Reinhold Scherer, Martin Billinger, Gernot Muller-Putz, “FORCe: Fully Online and Automated Artifact Removal for Brain-Computer Interfacing,” in *IEEE Transactions on neural systems and rehabiliation engineering*, Sep 2015.

[40] “https://en.wikipedia.org/wiki/mutual_information.”