BRAIN LOCALIZATION AND FEATURE EXTRACTION USING COMPUTATIONAL INTELLIGENCE TECHNIQUES

A Thesis
submitted in partial fulfillment of the requirement for the Degree of
Master of Electronics and Telecommunication Engineering

Jadavpur University
May 2016

By

Rimita Lahiri

Registration No: 128930 of 2014-15
Examination Roll No: M4ETC1615

Under the Guidance of

Prof. Amit Konar

Department of Electronics & Telecommunication Engineering
Jadavpur University, Kolkata-700032
India

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

CERTIFICATE

This is to certify that the dissertation entitled "**Brain Localization and Feature Extraction Using Computational Intelligence Techniques**" has been carried out by RIMITA LAHIRI(University Registration No.:128930 of 2014-15) under my guidance and supervision and be accepted in partial fulfillment of the requirement for the Degree of Master of Electronics & Telecommunication Engineering. The research results presented in the thesis have not been included in any other paper submitted for the award of any degree to any other University or Institute.

Prof. Amit Konar (Thesis Supervisor)
Dept. of Electronics & Telecommunication Engineering
Jadavpur University

Prof. Palaniandavar Venkateswaran
Head of the Department
Electronics & Telecommunication Engineering
Jadavpur University

Prof. Sivaji Bandyopadhyay
Dean
Faculty of Engineering and Technology
Jadavpur University

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

CERTIFICATE OF APPROVAL*

The forgoing thesis is hereby approved as a creditable study of an engineering subject and presented in a manner satisfactory to warrant acceptance as prerequisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn there in but approve the thesis only for which it is submitted.

Committee on final examination for the evaluation of the thesis

---
Signature of the Examiner

Signature of the Supervisor

*Only in the case the thesis is approved

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC THESIS

I hereby declare that this thesis entitled "**Brain Localization and Feature Extraction Using Computational Intelligence Techniques"** contains literature survey and original research work by the undersigned candidate, as part of her Degree of Master of Electronics & Telecommunication Engineering.

All information have been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name: Rimita Lahiri

Examination Roll No.: M4ETC1615

Thesis Title: Brain Localization and Feature Extraction Using Computational Intelligence Techniques

---
Signature of the candidate

ACKNOWLEDGEMENTS

I would like to express my earnest gratitude and sincere thanks to my thesis supervisor Prof. Amit Konar, Department of Electronics and Telecommunication Engineering, Jadavpur University, for giving me the opportunity to work under him and inspiring me to explore the interesting fields of Machine Intelligence and Human Computer Interfaces. I am indebted to him for his patient guidance, critical and constructive views and untiring support that shaped my work. The past year has been a remarkable experience in terms of gaining knowledge and skill that I hope to carry on and develop further.

I am thankful to Prof. Palaniandavar Venkateswaran and Prof. Iti Saha Mishra who have acted as the Heads of the Department of Electronics and Telecommunication Engineering during the tenure of my Master of Engineering course for their guidance and support. I would also like to show my gratitude to the respected professors of the Department of Electronics and Telecommunication Engineering for their constant guidance and valuable advices.

I am grateful to my seniors Anuradha Saha, Archana Chowdhury, Arup Kumar Sadhu, Pratyusha Rakshit, Reshma Kar and Sriparna Saha, for their constant support and encouragement. I am indebted to my classmates for their constant support and good wishes with a special mention to Arnab Rakshit, Mainak Dan and Snehalika Lall who supported me in a number of ways. I am extremely thankful to my friends Ranita Saha, Ranodip Das and Srishti Srivastava for always helping me in every possible way throughout the tenure of two years. Lastly, I would like to thank my parents and my entire family for their love, support and guidance through the course work.

Date:

Place:

Rimita Lahiri

Dept. of Electronics and Telecommunication Engineering

Examination Roll No. : M4ETC1615

Jadavpur University

LIST OF PUBLICATIONS BY THE AUTHOR

- Rimita Lahiri, Pratyusha Rakshit, Amit Konar, Atulya. K. Nagar, "Evolutionary Approach for Selection of Optimal EEG Electrode Positions and Features for Classification of Cognitive Tasks", *IEEE Congress on Evolutionary Computation (CEC)*, WCCI 2016, (Accepted, To be published).

- Sriparna Saha, Rimita Lahiri, Amit Konar, Bonny Bannerjee, Atulya K. Nagar, "Human Skeleton Matching for E-Learning of Dance Using Probabilistic Neural Network", *IEEE International Joint Conference on Neural Networks (IJCNN)*, WCCI 2016, (Accepted, To be published).

- Arnab Rakshit, Rimita Lahiri, Sayan Ghosal, Abhirup Sarkar, Sanchita Ghosh, Amit Konar, "Robotic Link Position Control Using Brain-Computer Interface", *IEEE International Conference on Microelectronics, Computing and Communication (Microcom)*, 2016, (Accepted, To be published).

- Snehalika Lall, Rimita Lahiri, Amit Konar, Sanchita Ghosh, "An Improved Measure for Data Clustering in High Dimensional Space", *IEEE International Conference on Microelectronics, Computing and Communication (Microcom)*, 2016, (Accepted, To be published).

PREFACE

The primary motif behind the formulation and development of Brain-Computer Interfaces is to provide aid to the patients suffering from several motor disabilities and thus cannot interact with the surrounding environment; hence the only possible way of survival for them is to instruct an assistive device that is capable of reading and decoding oscillatory activities occurring inside the cognitively intact brain. The thesis is focused on the application of different computationally intelligent tools that enhances the system performance in the concerned research domain. Two major problems rather two phases of a problem have been mainly addressed in the thesis, firstly, Brain Localization and the other one is Feature Extraction. There are different modalities for acquisition of brain signal waves, among them EEG is found to be the most suitable for the applications discussed in the thesis.

The first segment proposes a novel framework for selection of optimal EEG electrodes and feature set as well. A variant of the traditional Firefly Algorithm, formally termed as Self Adaptive Firefly Algorithm (SAFA) has been employed as an optimization tool to serve the purpose. From elementary neurophysiological knowledge it is well known that not all brain regions contribute equally for all sort of cognitive tasks, hence it is extremely important to find out the specific regions of human brain that is activated for a specific task. Again, redundant information is also undesirable for any problem as that leads to a huge computational expense, so selection of features also plays a key role in EEG based BCI systems.

The second segment designs an EEG based BCI system from pattern recognition perspective and emphasizes on the impact of feature extraction in such a system. Due to the poor spatial resolution of EEG signals spatial filtering has been emerged as one of the most efficient ways of spatially representing EEG signals. In this case, CSP has outperformed the other basic spatial filters in terms of most of the performance metrics. CSP basically projects the raw EEG signals into another subspace such that the projected signals are maximally discriminate. A regularized framework has been developed to compensate for the drawback of overfitting and lack of robustness. The efficacy of the designed framework has been tested by implementing for classification of EEG motor imagery paradigms. Two distinct methodologies have been adopted stressing on different phases of a pattern recognition based approach and in both the cases CSP has been utilized as a feature extracting tool. With respect to EEG based BCI system design the novelty in the present research lies in the process of bypassing peripheral muscular activities with the help of an assistive machine that can detect mental states easily. Sufficient experimental data acquisition has been carried out in laboratory and the efficiency of the proposed scheme has been validated by conducting standard statistical tests.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

This chapter primarily emphasizes upon the main objective of the thesis as a whole by introducing the necessary preliminary concepts. It attempts to provide an insight to the rudimentary concepts of man-machine interaction along with its scopes of possible applications of BCI to develop an intelligent interface to cater to the human needs. Section 1.1 recapitulates the steps of evolution of a basic BCI device and states its attributes with a functional definition. Section 1.2 introduces Computational Intelligence (CI) as a new research discipline and establishes a link of application of CI techniques in the present work. Section 1.3 delineates the prerequisite concepts of brain localization and reports the recent trends of research in the domain. Section 1.4 describes different brain signal measuring modalities and justifies the reason of choosing EEG as the preferred technique for the present work by reporting its superiority over the existing methodologies. Section 1.5 highlights one of the most popular EEG paradigms, termed as Motor Imagery (MI) and presents the related works in the domain of MI detection. Section 1.6 provides a rough outline regarding the possible applications of the thesis. Finally section 1.7 and 1.8 presents the objective and organization of the thesis briefly.

1.1 INTRODUCTION TO BRAIN COMPUTER INTERFACE (BCI)

A BCI [1] system primarily aims at establishing a platform for successful communications between a human being and an assistive machine, preferably a computer and is employed in a large number of applications. The preliminary concept of BCI [2] has been evolved through well known American TV series Star Trek that was aired long back in 1966. The series presented a character that suffers from locked in syndrome where the body is paralyzed after losing the voluntary control of the muscles but the person is cognitively intact, that is a functionally active brain is trapped in an inactive body that is not able move its organs. In such a scenario, the only possible modality to communicate with the environment for survival is to employ an assistive device that can read brain signals and decode them into executing necessary control commands. Such a device is termed as BCI [3].

Back in 1960, such an idea of controlling a device using brain signal patterns seemed to be entirely based on fantasy, despite the successful attempts of German scientist Hans Berger to record and measure brain electrical potentials from human scalp in 1929. Moreover, the technology available to measure and process of brain signals for analysing user intensions was very limited. The first BCI was introduced by Dr. Grey Walter in 1964. Unfortunately, Dr. Walter did not involve himself in publishing this revolutionary development; instead he simply discussed the basic matters while presenting a talk to Ostler Group, London. During those days and even since the turn of the century, there were very limited number of research enthusiasts who invested their valuable intellectual resources in this particular research domain and it is needless to say that there were very few labs actively dedicated for research conducted in the field of BCI. But the situation has changed these days, extensive research carried out in the field of neuroscience over the last decades has led to much better understanding about the working principle of human brain. Now a day, with the help of novel signal processing algorithms and the rapid development of computing power have enabled successful implementation of BCI systems in real time applications [4]. Unlike earlier days, today thousands of research groups are participating in active BCI research and as an obvious consequence; every day there is more BCI related papers, conferences, public talks, media articles and so on. More importantly, BCI is no longer considered as science fiction and thus BCI has succeeded in achieving its initial goal of proving the worth of rehabilitative applications of BCI for patients suffering from motor function disabilities.

Normally for performing a task involving control or communication, firstly the process is initiated with the user's intent which in turn triggers a complex operation of activating

specific brain areas that contribute for the specific control task and next signals are transported through the motor pathways to the corresponding muscles that perform the physical movements required to complete the task. A BCI system provides an alternative path way by recording the brain signal pattern according to a user's intent and then translating the recorded brain activity into effective control instruction for real time applications. Deciphering these brain signal patterns into active commands require the tools of pattern recognition and signal processing which can be employed by using a computer. Since the processed signals originate from the human brain and because of the involvement of a computer in the procedure, the system is aptly termed as "Brain Computer Interface". A BCI system possesses three mandatory attributes,

- It records signal rhythms directly from human brain (invasively or non-invasively)
- It must send feedback to the user for real time signal processing
- While recording brain signal potentials the user must involve himself in intentional control.

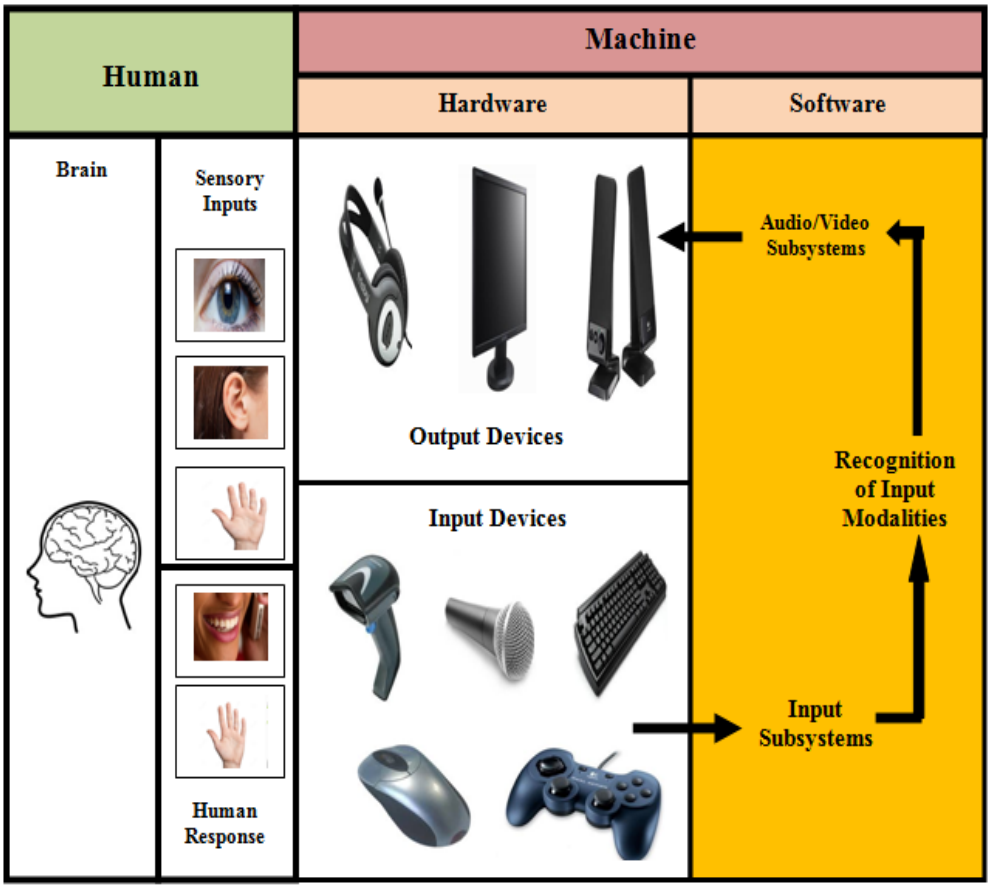


Fig 1.1 Overview of BCI System

Fig 1.1 depicts a detailed description of the fundamental concepts of BCI through different interfaces using a diagram. As shown in the diagram, a user can interact with the hardware directly through an interface. A man-machine interface environment always has a feedback component incorporated in the input-output chain that evaluates and controls the data which traverses from the user to an assistive computer and comes back to the user as well in the form of a recurrent loop.

Although most researchers have accepted the formal terminology and the relevant concepts of BCI, but there is no universally accepted formal definition of BCI, in fact many variants of the same are found in the BCI literature. According to Wolpaw, "*A direct brain-computer interface is a device that provides the brain with a new, non-muscular communication and control channel*", in the words of Donoghue, BCI can be defined as, "*A major goal of a BMI (brain-machine interface) is to provide a command signal from the cortex. This command serves as a new functional output to control disabled body parts or physical devices, such as computers or robotic limbs".*

1.2 Computational Intelligence

Computational Intelligence (CI) [5] is originally an invention of Professor Lotfi A. Zadeh. Since its initiation the discipline has undergone numerous alterations both in its content as well as organization. The elementary definition of CI emphasizes on fuzzy logic, genetic algorithm, neural networks and probabilistic reasoning along with the study of their intersecting regions. Gradually, the periphery of CI expanded and today's definition of CI is greatly motivated by the biologically inspired models of machine intelligence. Modern concepts of CI are mainly concerned with granular computing, neural computing, evolutionary computing and their impact and interaction with artificial life, chaos theory and others, as shown in Fig. 1.2.

According to Prof. James Bezdek CI can be defined as, "*A system is computationally intelligent when it: deals with only numerical (low level) data, has pattern recognition components, does not use knowledge in the AI sense; and additionally when it (begins to) exhibits i) computational adaptivity, ii) computational fault tolerance, iii) speed approaching human like turnaround and iv) error rates that approximate human performance."*

In the present context, computational adaptation typically refers to the ability of a system to tune its parameters following certain optimizing criterion and depending upon the temporal changes in its input and output variables. Most of the Artificial Neural Networks (ANN) follows this attribute. Further, computational fault tolerance is more or less common

characteristic in parallel and distributed environment since computational resources are replicated at each of the distributed unit in such a way that even if a few units are damaged or malfunctioning then that would not cause the whole system to shut down, as resources are available at each unit. It is important to note that while ANN and fuzzy logic is controlled by their inherent fault tolerance mechanisms, but belief networks or genetic algorithms can be configured in this way to get the benefit of computational fault tolerance.

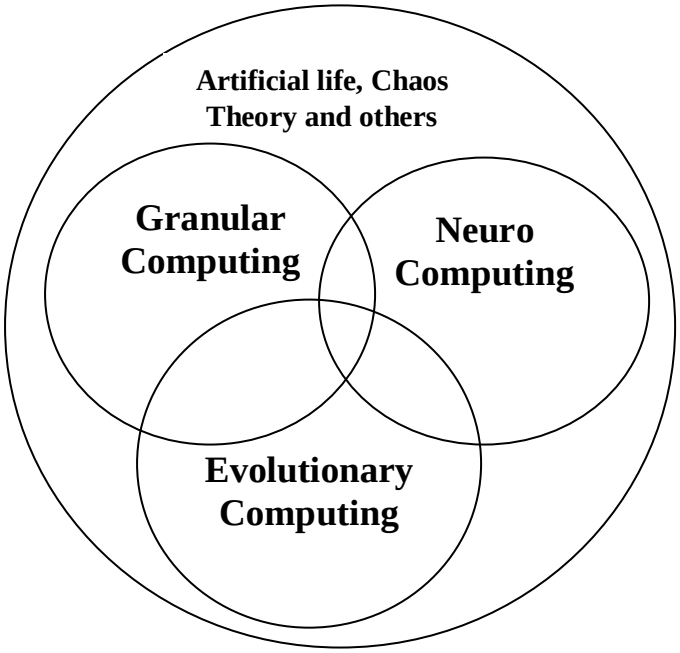


Fig. 1.2 Computational Intelligence Family

The other two important aspects of Bezdek's definition is related to computational speed and error rates, as it is often seen that accuracy is sacrificed to achieve higher speed which is extremely undesirable in this context. Both ANN and fuzzy logic produce considerably quicker responses against any input excitation, moreover unlike traditional systems fuzzy logic has provisions for firing multiple rules ensuring the partial matching of the available facts with antecedent clauses of those rules. Hence fuzzy logic is less susceptible to error and has got high computational speed as well. Similarly, ANN has provisions for updating multiple neurons concurrently yielding high computational speed and the parameters are altered wisely at each step satisfying certain constraints such that error rate is minimized. The parallelly constructed structure of Genetic Algorithms(GA) and belief networks increases the computational speed and their inherent filtering component takes care of the accuracy issues and enhances the same to the greatest possible extent. To justify the first clause of

Bezdek's definition it is important to note that all of fuzzy logic, GA and ANN deals with numerical data, they have certain pattern recognition components and none of them use traditional AI concepts, hence from this point of view fuzzy logic, GA and ANN are reliable members of CI family.

Basically, computational intelligence emerged as an alternative to deal with the shortcomings of traditional Artificial Intelligence (AI). Gradually, the limitations of AI became more and more pronounced over the decades when AI became incompetent to serve the demand of search, optimization and machine learning in i) information systems with large biological and commercial databases, ii) factory automation for steel, aerospace, power and pharmaceutical industries. Almost at the same time, the contemporary model of non traditional machine intelligence including rough sets, fuzzy logic, chaos theory, artificial neural networks, genetic algorithms etc. proved their worth and efficiency to deal with the drawbacks of AI and thus the limitations of traditional AI opened up new avenues for the non conventional model of various engineering applications. This new discipline is termed as **Computational Intelligence**.

1.3 BRAIN LOCALIZATION FOR BCI

EEG signals despite having an excellent temporal resolution suffers from the drawback of poor spatial resolution because of the limited number of recording sites over the human scalp and also due to the shielding effect of the skull. Due to volume conduction effect, the source signals inside the brain often gets scattered and as a result many electrodes placed in the closed proximity of that specific scalp region records components of the concerned source, as shown in Fig. 1.3. It is considered as a challenging problem in the domain of neuroscience to predict the active regions of brain from the recorded potential distributions. Estimation of original sources is apparently a tricky problem because there can be an infinite combinations of internal currents that can result in creating such potential distributions over human scalp.

A BCI can be briefly defined as a machine learning device that can detect and decode a certain set of patterns from the control signals directly acquired from a user's brain. In the context of BCI, the fundamental concept behind source localization lies in the mapping of multichannel EEG signal into a higher dimensional subspace that is comprised of multiple sources that are modelled either as current dipoles or monopoles. Further, the source localization can be proved as an efficient tool for classifying EEG signals as well, instead of employing the traditional machine learning algorithms utilizing source localization a user can

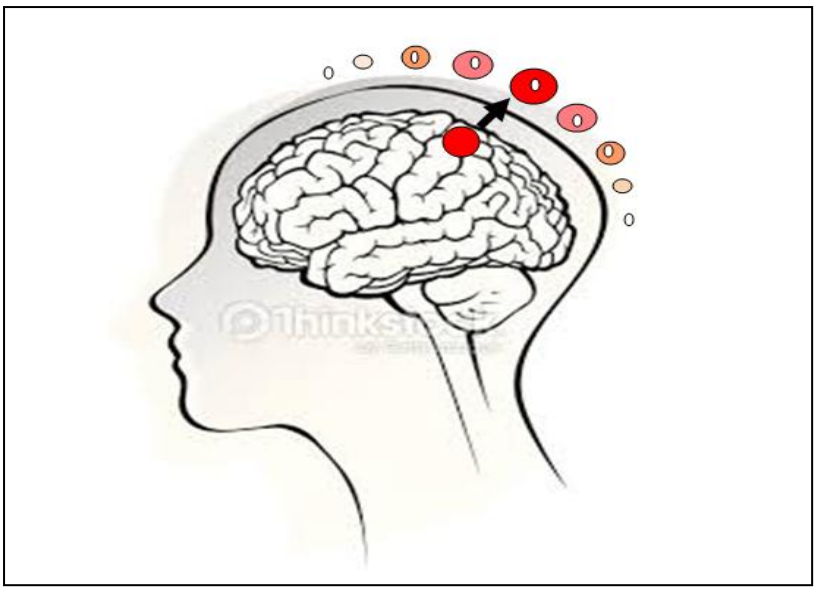


Fig 1.3 Volume Conduction effect of Human Brain

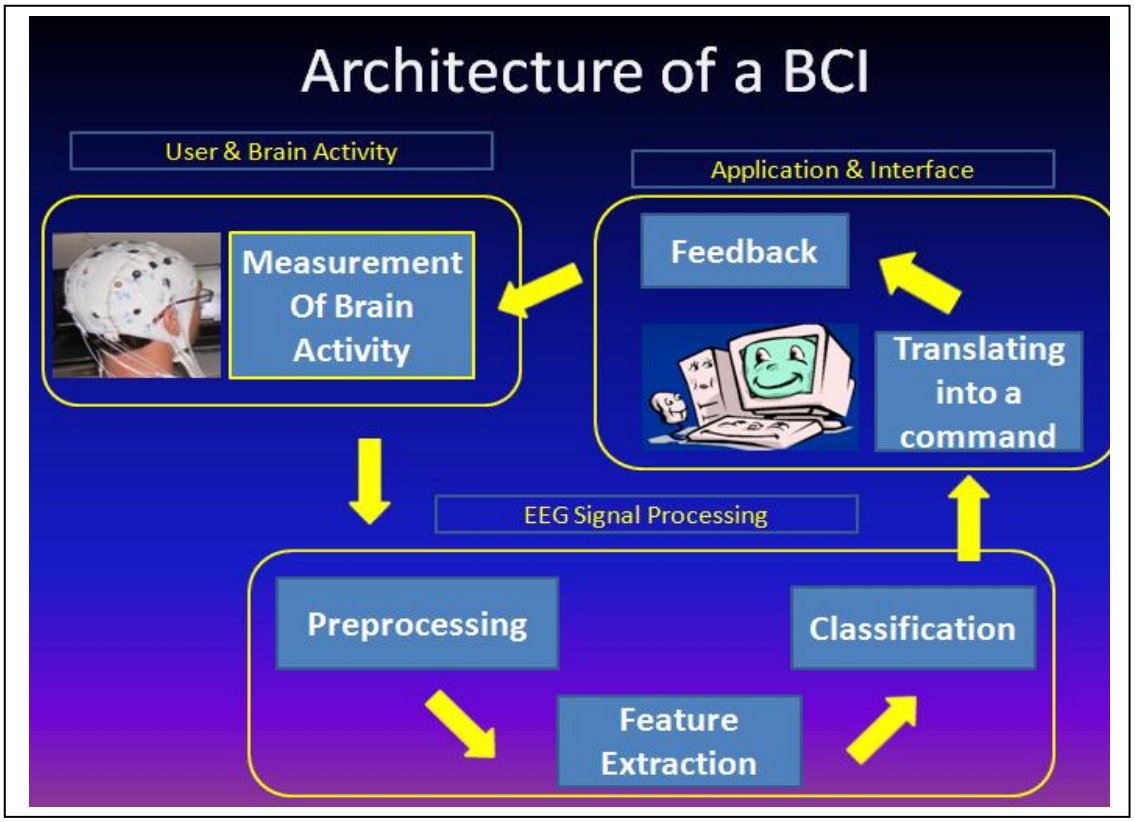


Fig 1.4 Overview of EEG Based BCI System

exploit two neuro physiological postulates by reconstructing the sources of recorded scalp potentials,

Table 1.1 HIGHLIGHTS OF THE TRENDS IN EEG SOURCE LOCALIZATION RESEARCH

Author and Year of Publication	**Highlights**
Zarghami *et al*.(2015)	Transfer-Function-Based Calibration of Sparse EEG Systems for Brain Source Localization
Shirvany *et al.* (2014)	Particle Swarm Optimization Applied to EEG Source Localization of Somatosensory Evoked Potentials
Hammond *et al.* (2013)	Cortical Graph Smoothing: A Novel Method for Exploiting DWI-Derived Anatomical Brain Connectivity to Improve EEG Source Estimation
Shirvany *et al.* (2012)	Investigation of Brain Tissue Segmentation Error and Its Effect on EEG Source Localization
Yitembe *et al.* (2011)	Reduced Conductivity Dependence Method for Increase of Dipole Localization Accuracy in the EEG Inverse Problem
Antelis *et al*. (2010)	DYNAMO: Dynamic Multi Model Source Localization Method for EEG and/or MEG
Barton *et al.* (2009)	Evaluating the Performance of Kalman-Filter-Based EEG Source Lcalization
Noirhomme *et al.*(2008)	Single-trial EEG Source Reconstruction for Brain-Computer Interface
Xu *et al*. (2007)	Lp Norm Iterative Sparse Solution for EEG Source Localization
Rodriguez-Rivera *et al.* (2006)	MEG and EEG Source Localization in Beamspace
Qiu *et al.* (2005)	A Feasibility Study of EEG Dipole Source Localization Using Particle Swarm Optimization
Gutierrez *et al*. (2004)	Estimating Brain Conductivities and Dipole Source Signals with EEG Arrays
Rodriguez-Rivera *et al*.(2003)	Statistical Performance Analysis of Signal Variance Based Dipole Models for MEG/EEG Source Localization and Detection
Schimpf *et al.* (2002)	Dipole Models for the EEG and MEG
Sclabassi *et al*. (2001)	EEG Source Localization: a Neural Network Approach

.

- Lateralization of electrocortical activities (for example, a stronger arousal of left sensorimotor cortex due to movement of right hand because of the contra lateral working mechanism of human brain)
- Spatially representative distributions of different extremities in motor as well as sensorimotor cortex region.

In the later stage, this phenomenon yields activations that are spatially discriminate during real as well as imaginary movements of various extremities and thus source localization can be used as an efficient tool for differentiating brain signals which in turn can enable a BCI system to classify multitude of conditions.

Further, BCI systems based on the concept of source localization are resistant to the local variations of EEG signals and hence can be chosen as a robust alternative to deal with such situations. To use the traditional machine learning algorithms it is often required to train the algorithms a large number of times because of the variations in EEG signal patterns over time instances as well as different subjects, which can result in a high run time complexity as well.

Contrarily, as long as the spatial locations of different electrodes remain constant, such trivial variations won't affect BCI systems designed for source localization. In the present work, we have proposed an extension of the source localization based upon Independent Component Analysis (ICA) [6]. By employing ICA, we can decompose the EEG signal recorded into *N* maximally independent components where *N* is the maximum number of electrodes.

1.4 BRAIN SIGNAL MEASUREMENT TECHNIQUES

In order to measure the neuronal activity occurring inside human brain, several devices [26] with different working mechanisms are available [7].

Firstly, Electroencephalogram (EEG) [8] devices are used to detect electrical signals originating within human brain. Due to neuronal firing, an ionic decomposition occurs that yields a state leading to the formation of dipoles and as a result negative charges are found to be accumulated over the motor cortex. Due to the attractive and repulsive forces that act in between these ions, volume conduction occurs as a result of which information is transported across the brain to maintain resting potential and to propagate action potential, as shown in Fig. 1.5. There are a variety of electrode options that are available in the market which are capable of recording these signals from human scalp. For example, tin (Sn) electrodes, gold plated (Au) electrodes, silver or silver chloride (Ag/AgCl) electrodes etc. can serve the

purpose. But in most of the cases, Ag/AgCl electrodes are preferred over others due to its longevity, cost effectiveness, precision and ease of use. More importantly, the signal recorded from human scalp has its amplitude in the order of microvolts (µV), to deal with signals at this amplitude level becomes extremely difficult. Hence, a mandatory operation performed by every EEG device is the amplification of the original scalp recording to a certain threshold limit. Further, to employ and analyse these signals using a digital computer requires digitization of the same that is an analog to digital converter is also embedded as a part of EEG acquisition device. Finally, a band pass filter is employed to extract the signals within the range of 0-70 Hz, which is usually considered as the maximum range of EEG [9] recordings.

Functional Magnetic Resonance Imaging (fMRI) [10] is another modality that utilizes the difference in magnetization occurring due to oxygenation and de-oxygenation of blood as the prime measure. Hence, the coupling between cerebral blood flow and neuronal arousal has a major influence in this method. Despite its precision of locating sources of neuronal firing inside human brain there remains a time lag between actual firing and the detection of the concerned source. This temporal delay does not allow these FMRI systems to be employed in real time on line applications.

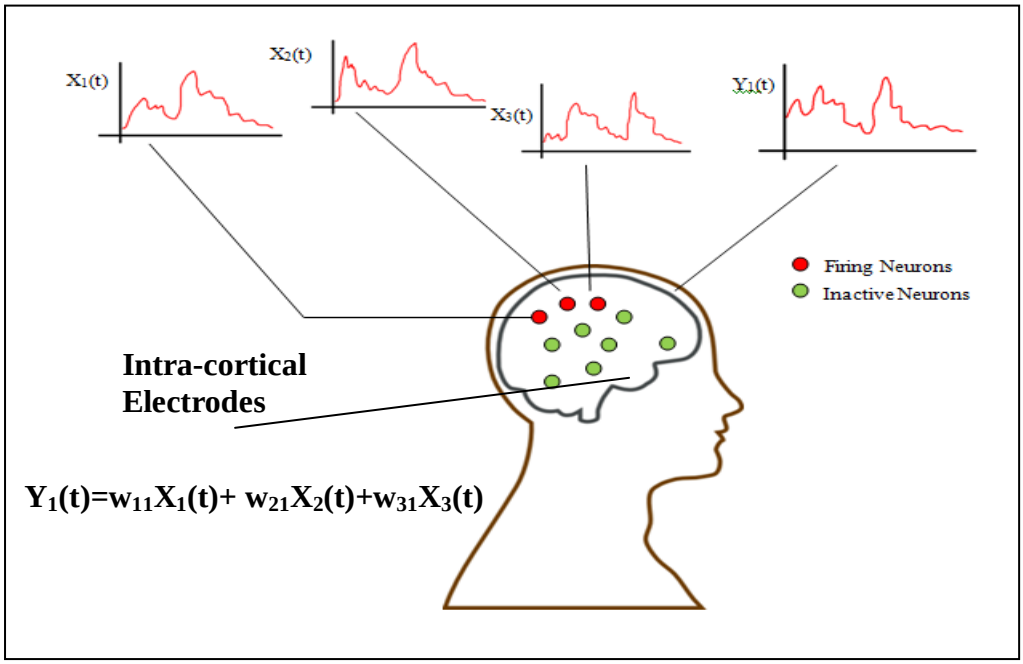


Fig. 1.5 Neurophysiological Working Mechanism of EEG

Magnetoencephalography (MEG) is another brain rhythm recording device that considers the magnetic fields arising out of the electrical signals originating within human

brain as their primary metric. Because of the sensitivity involved with these kinds of signal, extremely expensive high quality equipments are recommended to carry out these experiments successfully. So, high cost is the reason which stops it from gaining wide popularity.

Electrocorticography (ECoG) [11] is another device which has maximum of the attributes similar to EEG in terms of working mechanism, but the difference lies in the procedure of implanting the electrodes on/inside human scalp. Unlike EEG, in ECoG systems the electrodes are placed inside the cortex that is over the exposed area of human brain, to enhance the information content. Since, for such electrode placements a user needs to undergo craniotomy (surgery on skull), it is not always a feasible method for real time applications.

For similar reasons, in spite of having excellent temporal and spatial resolutions of the brain signals acquired through the electrodes placed deep inside human brain, intra-cortical electrodes [12] are not always considered as the preferred alternative for most of the lab based experiments. In fact, the invasiveness greatly limits the applicability of these electrodes.

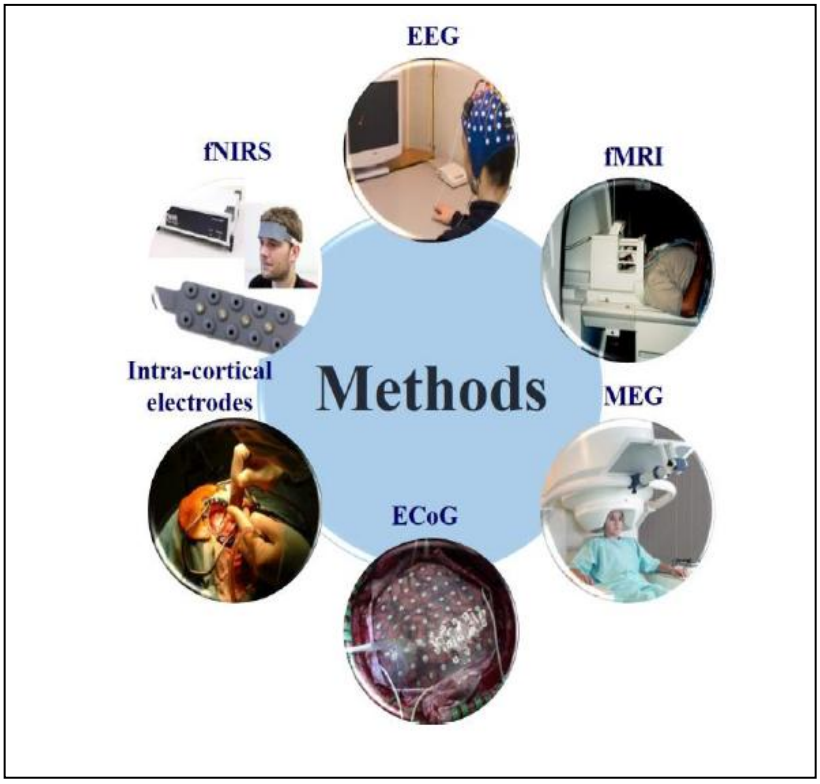


Fig. 1.6 Brain Signal Measuring Interfaces (Invasive & Non invasive)

Functional Near-Infrared Spectroscopy (fNIRS) [13] is a hemodynamic technique based functional neuroimaging method where the transmission/scattering characteristic of the near Infrared (NIR) radiation has been employed and it is inferred that oxygenated

haemoglobin molecules and deoxygenated haemoglobin molecules absorb and scatter the light at different portions of NIR spectrum [26]. Hence, the degree of attenuation or the signal strength of the light received back from the target tissue measures of the oxygen concentration in different regions of brain and thereby locates the source of neuronal firing.

Fig. 1.6 [26] illustrates the different equipments that are popularly used to record brain signal rhythms. From the above discussion, it can be inferred that invasive recording techniques combine excellent signal quality, very good spatial resolution, and a higher frequency range; moreover these recordings are less vulnerable to artifacts. Further, the cumbersome application and re-application of electrodes is unnecessary for invasive approaches. Despite having these advantages, invasive methods have a serious limitation of surgery requirement. From ethical and financial perspective, neuro surgery is often considered as an impractical solution other than the users who depend upon BCI for communication. In fact, in certain cases, these needs can be fulfilled by non invasive techniques thus avoiding the complexity of risky surgeries. More importantly, long term stability of these approaches is also doubtful which motivated us to consider EEG as the most suitable bio-modality to carry out lab based experiments for analysing brain activities.

TABLE 1.2 RELATIVE DRAWBACKS OF OTHER BRAIN SIGNAL RECORDING INTERFACES OTHER THAN EEG

Measuring Interface	**Limitations**
fMRI	Exorbitant, large sized and immobile interface, aggravates claustrophobia in the users while recording
MEG	Exposure to high intensity (> 1 Tesla) magnetic fields, bulky and stationary device
ECoG	Invasive technique, requirement of risky surgery for implantation of electrodes
fNIRS	Poor temporal resolution, less depth of penetration and recording of limited sites over the human scalp

1.4.1 Electroencephalogram (EEG)

Electroencephalography [14] refers to the well established procedure of recording EEG signals from human scalp using electrodes that has been used in the clinical and research

setup since last few decades. EEG signal acquisition system is comparatively light weight, cost efficient and easy to apply. EEG based BCI systems (Fig. 1.4) possess exceptional temporal resolution that is those systems can detect changes within a predefined interval of time with precision. On the other hand, EEG signals suffer from poor topographic resolution and more importantly they are susceptible to be disturbed by noise and artifacts (for example, bioelectrical signals originated because of the eye movement or eye blink (Electrooculographic activity or EOG), or from body muscles (Electromyographic activity or EMG)).

The primary attributes of EEG signals can be listed as,

- EEG signals are extremely non-stationary in nature because of the asynchronous neuronal firing over certain period of time.
- EEG signals are non-periodic and non-Gaussian in nature.
- The amplitudes of EEG signals lie within the range of [-10μV, 10μV].
- Usually, the frequency band in the range 0.5-70 Hz is referred to as EEG bandwidth.(Table 1.3)
- EEG signals are primarily determined by the arousal of different brain regions. The cerebral cortex which covers the outer part of the cerebrum, which is responsible for major part of information processing inclusive of thoughts and action, is primarily segmented into four components termed as "lobes". Each lobe is responsible for carrying distinct set of functions; Fig 1.7 depicts a clear picture of brain partitioned into functional components. To conduct an EEG based experiment, it is extremely important to have a prior knowledge about the operations of the different regions of the cerebral cortex in order to acquire relevant signals corresponding to a particular cognitive task.

- *Frontal lobe*: Creative thought, problem solving, intellect, judgment, attention, abstract thinking, physical reactions, muscle movements, coordinated movements and personality. The prefrontal region is responsible for behaviour, planning and short term memory.
- *Parietal lobe:* Visual functions, language, reading, internal stimuli, tactile sensation and sensory comprehension, visuospatial processing
- *Temporal lobe:* Senses of smell and sound, processing of complex stimuli like faces and scenes
- *Somatosensory Cortex*: Perception of somatosensations
- *Occipital lobe*: Sense of sight or vision

- *Motor Cortex*: Movement co-ordination

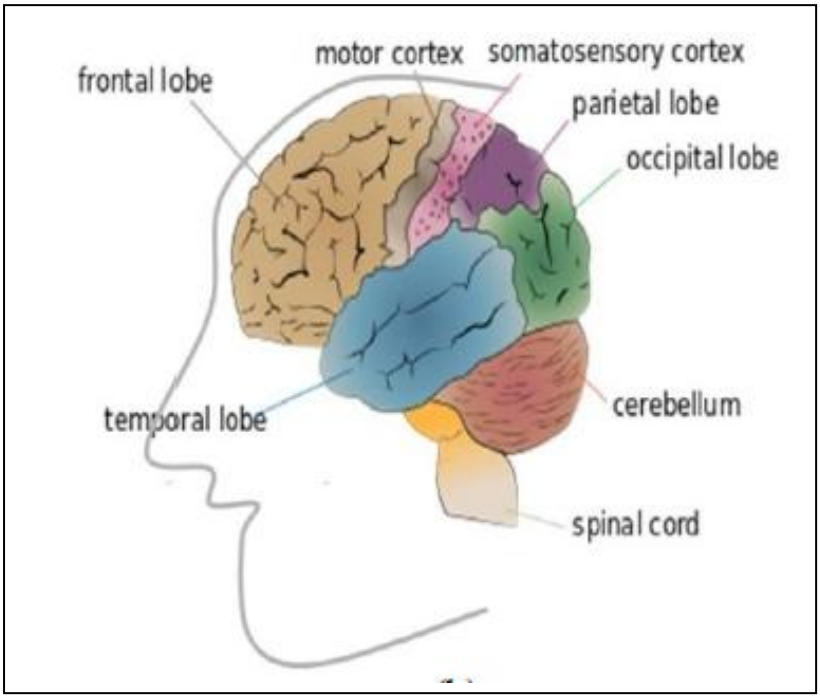


Fig. 1.7 Functional Areas of Human Brain

TABLE 1.3 EEG FREQUENCY BANDS

Band	Frequency Range (Hz)	Implications
Delta	0.5-4	Sleep, fatigue, severe slowing of mental processes
Theta	4-8	Meditation, attention lapses, slowed processing, memory consolidation
Alpha	8-13	Relaxation, readiness, inactive cognitive processing
Beta	13-30	Focus, active concentration, alertness
Gamma	30-70	Cross-modal sensory processing, short-term memory matching of recognized objects

A number of EEG signal modalities are used for BCI systems depending upon the stimulus presented to a user and requirement of the problem. The most commonly used

modalities include P300 potential, slow cortical potential (SCP), Steady State Visually Evoked Potential (SSVEP), Event Related Desynchronization/ Synchronization (ERD/ERS).

A P300 based BCI system depends upon a stimulus that flashes in succession. Practically, stimuli are comprised of letters, but at times it can be goal directed symbols for practical applications such as robot movement, cursor control etc. Selective concentration and attention to a specific flashing symbol yields a brain pattern formally termed as P300, which is originated from the centro parietal brain region at around 300ms after the presentation of the stimulus.

SCP signifies the change in the membrane potentials of the cortical dendrites that lasts typically upto several seconds starting from 300ms after the stimulus presentation. This response is usually observed in the motor cortex region.

An SSVEP based BCI system requires a number of stimuli where each distinct stimulus corresponds to an output command, which in turn is associated with a task that a BCI based system can execute. Unlike P300 systems, instead of flashing in succession the stimuli continues to flicker at different frequencies in the range of 6-30 Hz. Paying attention to a flicker elicits an SSVEP activity in the visual cortex with the same frequency as that of the target flicker. So SSVEP based BCI systems are capable of determining which flicker has occupied a user's attention simply by checking the SSVEP rhythm in the visual cortex at a specific frequency.

ERD/ERS [15] refers to the decrease/increase in EEG signal amplitudes in specific frequency bands during execution of dynamic cognitive processes as a response to excitation invoked by stimuli presentation. Motor Imagery tasks produce ERD/ERS in motor cortex while tactile stimulations give rise to alpha band ERD/ERS in somatosensory cortex.

1.5 MOTOR IMAGERY

Not only actual task execution, but the preparation of movement or imagination of task execution elicits a change in the Sensorimotor Rhythms (SMR). An SMR typically signifies the oscillations in the brain activity recorded from the somatosensory and motor cortex region. As mentioned previously, ERD/ERS patterns can be volitionally produced by motor imagery signals which refer to the imagination of the movement without actually performing it. Topographically, ERD/ERS patterns follow a homuncular mechanism that is the oscillatory activity invoked by right hand motor imagery is most prominent over the electrodes (C_3) placed over the left hemisphere and vice versa for the left hand motor imagery case [16]. So, the activity invoked by hand motor imagery is located on the contralateral side [17]. Since the

cortical areas corresponding to left and right leg motor imagery resides too close, hence it is not possible to discriminate between left foot and right foot in EEG. For similar reasons, detection of ERD/ERS patterns corresponding to individual fingers is not feasible. Since, hand areas, feet areas and tongue areas are topographically different and large enough hence in the present works, these four motor imagery signals are considered to control BCI systems. Unlike selective attention based BCI, motor imagery based system do not require any external stimulus, but a user requires to adapt the skills of motor imagery to perform it with efficiency. For this reason, BCI systems based upon motor imagery do not produce satisfactory results in the first session itself instead some training can enhance the result. Again the performance and the training time also vary amongst the subjects, however longer training is recommended at times for sufficient control. Thus, training is an important aspect in most of the BCI systems; in such systems users undergo a process called operant conditioning. During BCI learning, operant conditioning incorporates a feedback that is displayed on the computer screen and formally known as "neurofeedback". Usually, the feedback is in the visual form, but it can be presented in tactile as well as audio form also. The feedback basically evaluates the user's performance that whether he has done well or failed, a user can exploit this feedback to optimize his performance and thus enhance the BCI system outcome.

1.6 PRINCIPLES OF COMPUTATIONAL INTELLIGENCE

The principles of CI that have been employed in the present work include two major components,

- Pattern Recognition
- Optimization

1.6.1 Pattern Recognition

Pattern Recognition is a sub branch of CI that mainly emphasizes on discriminating patterns or regularities in a dataset based on certain attributes. Although the possible approaches of utilization of this tool include supervised, unsupervised and semisupervised methodologies, the present work has been designed based upon supervised algorithms only. Any Pattern Recognition system follows the three fundamental steps of,

- Feature Extraction
- Feature Selection
- Classification [18]

Table 1.4 RELATED WORKS IN MOTOR IMAGERY DETECTION

Author and Year of Publication	Highlights
Xu *et al.*(2014)	Enhanced Low-Latency Detection of Motor Intention from EEG for Closed-Loop Brain-Computer Interface Applications
Park *et al*. (2013)	Classification of Motor Imagery BCI Using Multivariate Empirical Mode Decomposition
Bamdadian *et al.* (2012)	Online Semi-Supervised Learning with KL Distance Weighting for Motor Imagery Based BCI
Ang *et al.* (2011)	Calibrating EEG Based Motor Imagery Brain-Computer Interface from Passive Movement
Li *et al*. (2010)	An EEG Basrd BCI System for 2-D Cursor Control by Combining Mu/Beta Rhythm and P300 Potential
Ang *et al.* (2009)	A Clinical Study of Motor Imagery Based Brain Computer Interface for Upper Limb Robotic Rehabilitation
Wu *et al*. (2008)	Classifying Single-Trial EEG During Motor Imagery by Iterative Spatio-Spectral Patterns Learning (ISSPL)
Sadeghian *et al.*(2007)	Continuous Detection of Motor Imagery in a Four Class Asynchronous BCI
Wang *et al.* (2006)	Common Spatial Pattern Method for Channel Selection in Motor Imagery Based Brain-Computer Interface
Xiaomei *et al.* (2005)	Adaboost for Improving Classification of Left and Right Hand Motor Imagery Tasks
Townsend *et al.* (2004)	Continuous EEG Classification During Motor Imagery-Simulation of an Asynchronous BCI
Cincotti *et al.* (2003)	The Use of EEG Modifications Due to Motor Imagery for Brain-Computer Interfaces
Dornhege *et al.*(2002)	Combining Features for BCI
Pfurtscheller *et al.* (2001)	Motor Imagery and Direct Brain-Computer Communication
Ramoser *et al*. (2000)	Optimal Spatial Filtering of Single Trial EEG During Imagined Hand Movement
Castellano *et al.* (1999)	Moving Target Detection in Infrared Imagery Using a Regularized CDWT Optical Flow

1.6.2 Optimization

Optimization signifies the mathematical steps that are involved in formulation of the procedure of making an entity or a solution as efficient or perfect as possible. In the present context, the term "Optimization" refers to the method of selecting the optimal solution from a pool of candidate solutions satisfying a predefined constraint with a goal of either maximizing or minimizing an objective function. Among the various methods of optimization, evolutionary techniques [19] are those which typically follow the postulate of "survival of the fittest". In this work Firefly Algorithm [20] has been employed for solving a specific problem related to EEG signal modality.

1.7 OBJECTIVE OF THE THESIS

The primary objective of the thesis is to assemble the existing CI techniques and analyze their scopes of application in the source localization and feature extraction paradigm. The formulation of modern day intelligent man machine interactive devices majorly depends on the judicious analysis of existing CI modalities and proficient development of the concerned interface. These CI based BCI modalities find extensive use in rehabilitation, tele-operation, tele-navigation as well as context awareness.

Amongst the brain signal measuring techniques, EEG is considered to be the most efficient modality for BCI applications [21]. Literature show that already many interesting applications have been implemented using EEG, still there exists ample opportunity to improvise the existing applications by introducing innovative concepts. The entire work has been partitioned into two components. The first part presents an evolutionary perspective for selection of optimal electrode positions and EEG features for a specific cognitive task. This part is mainly concerned with the feature extraction and features as well as electrode selection phase and justifies the usage of an optimization algorithm to reach closer towards relatively simpler solutions of such complex problems. The second segment explores the signal processing methods and emphasizes on the impact of spatial filtering [22] on EEG signals. This part reiterates the existing variants of the well known CSP algorithm and analyses the outcomes of the application of newly formulated regularized variants of CSP for motor imagery detection [23] purpose. All computations are performed in a MatlabR2012b environment in an Intel Core i3 processor running at 2.30 GHz. All human subjects on whom experiments are conducted sign consent forms prior to providing data.

The possible application areas of the present thesis include,

- BCI systems
- Rehabilitative systems [24]
- Robotic control [25]
- Context Awareness

1.8 ORGANIZATION OF THE THESIS

The thesis is segmented into six distinct chapters. Chapter 2 provides a detailed illustration regarding the various machine learning tools and techniques that have been employed in the numerous experiments that have been carried out during the works included in the present thesis. Chapter 3 describes the proposed scheme of EEG optimum electrodes and feature selection from the evolutionary perspective. Further Chapter 4 narrates the needs of spatial filtering in EEG based works and introduces RCSP as the solution to deal with the drawbacks of traditional CSP. This chapter also mentions further scopes of future work by mentioning the limitations of the existing frameworks and opens new avenue for improvising the same. Chapter 5 aims to show the impact of employing RCSP as a feature extraction tool in motor imagery detection purpose for online BCI applications. Finally in Chapter 7 the conclusions are drawn and future scopes of work are stated. All chapters are provided with necessary bibliography. Appendix A provides the detailed descriptions of the performance metrics and statistical methods used to evaluate results and Appendix B presents a step wise description of running the source codes.

REFERENCES

[1] B. Graimann, B. Allison, and G. Pfurtscheller, "Brain – Computer Interfaces : A Gentle Introduction," pp. 1–28, 2010.

[2] J. M. Carroll, "Human–Computer Interaction," *Encycl. Cogn. Sci.*, 2009.

[3] A. Dix, *Human-computer interaction*. Springer, 2009.

[4] T. O. Zander, C. Kothe, S. Jatzev, and M. Gaertner, "Enhancing human-computer interaction with input from active and passive brain-computer interfaces," in *Brain-computer interfaces*, Springer, 2010, pp. 181–199.

[5] A. Konar, *Computational intelligence: principles, techniques and applications*. Springer Science & Business Media, 2006.

[6] L. Zhukov, D. Weinstein, and C. Johnson, "Independent component analysis for EEG source localization," *Eng. Med. Biol. Mag. IEEE*, vol. 19, no. 3, pp. 87–96, 2000.

[7] S. Sen Purkayastha, V. K. Jain, and H. K. Sardana, "Topical Review: A Review of Various Techniques Used for Measuring Brain Activity in Brain Computer Interfaces."

[8] E. Niedermeyer and F. H. L. da Silva, *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[9] M. Teplan, "Fundamentals of EEG measurement," *Meas. Sci. Rev.*, vol. 2, no. 2, pp. 1–11, 2002.

[10] L. I. Kuncheva, J. J. Rodríguez, C. O. Plumpton, D. E. J. Linden, and S. J. Johnston, "Random

subspace ensembles for fMRI classification," *Med. Imaging, IEEE Trans.*, vol. 29, no. 2, pp. 531–542, 2010.

[11] E. C. Leuthardt, K. J. Miller, G. Schalk, R. P. N. Rao, and J. G. Ojemann, "Electrocorticography-based brain computer interface-the Seattle experience," *Neural Syst. Rehabil. Eng. IEEE Trans.*, vol. 14, no. 2, pp. 194–198, 2006.

[12] D. T. Westwick, E. A. Pohlmeyer, S. A. Solla, L. E. Miller, and E. J. Perreault, "Identification of multiple-input systems with highly coupled inputs: application to EMG prediction from multiple intracortical electrodes," *Neural Comput.*, vol. 18, no. 2, pp. 329–355, 2006.

[13] S. C. Bunce, M. Izzetoglu, K. Izzetoglu, B. Onaral, and K. Pourrezaei, "Functional near-infrared spectroscopy," *Eng. Med. Biol. Mag. IEEE*, vol. 25, no. 4, pp. 54–62, 2006.

[14] S. Sanei and J. A. Chambers, *EEG signal processing*. John Wiley & Sons, 2013.

[15] C. Qiang, P. Hu, and F. Huanqing, "Experiment study of the relation between motion complexity and event-related desynchronization/synchronization," in *Neural Interface and Control, 2005. Proceedings. 2005 First International Conference on*, 2005, pp. 14–16.

[16] G. Pfurtscheller and C. Neuper, "Motor Imagery and Direct Brain–Computer Communication," vol. 89, no. 7, pp. 1123–1134, 2001.

[17] B. Xu, A. Song, and J. Wu, "Algorithm of imagined left-right hand movement classification based on wavelet transform and AR parameter model," in *Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on*, 2007, pp. 539–542.

[18] S. Bhattacharyya, A. Khasnobish, S. Chatterjee, A. Konar, and D. N. Tibarewala, "Performance analysis of LDA, QDA and KNN algorithms in left-right limb movement classification from EEG data," in *Systems in Medicine and Biology (ICSMB), 2010 International Conference on*, 2010, pp. 126–131.

[19] A. Atyabi, M. Luerssen, S. Fitzgibbon, and D. M. W. Powers, "Evolutionary feature selection and electrode reduction for EEG classification," in *Evolutionary Computation (CEC), 2012 IEEE Congress on*, 2012, pp. 1–8.

[20] S. Bhattacharyya, P. Rakshit, A. Konar, D. N. Tibarewala, and R. Janarthanan, "Feature selection of motor imagery EEG signals using firefly temporal difference Q-Learning and support vector machine," in *Swarm, Evolutionary, and Memetic Computing*, Springer, 2013, pp. 534–545.

[21] A. Vallabhaneni, T. Wang, and B. He, "Brain—computer interface," in *Neural engineering*, Springer, 2005, pp. 85–121.

[22] D. J. McFarland, L. M. McCane, S. V David, and J. R. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 3, pp. 386–394, 1997.

[23] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain-computer interfacing.," *J. Neural Eng.*, vol. 9, no. 2, p. 026013, 2012.

[24] J. J. Daly and J. R. Wolpaw, "Brain–computer interfaces in neurological rehabilitation," *Lancet Neurol.*, vol. 7, no. 11, pp. 1032–1043, 2008.

[25] C. J. Bell, P. Shenoy, R. Chalodhorn, and R. P. N. Rao, "Control of a humanoid robot by a noninvasive brain–computer interface in humans," *J. Neural Eng.*, vol. 5, no. 2, p. 214, 2008.

[26] S. Datta, "COGNITIVE AND HAPTIC MODALITIES OF PERCEPTUAL ABILITY ANALYSIS FOR NEXT GENERATION HUMAN COMPUTER INTERFACES," 2014.

[27] M. Pal, "DESIGN OF AN EEG-DRIVEN ARTIFICIAL LIMB SYSTEM," 2014.

Chapter 2

Standard Tools and Techniques

This chapter aims to revisit the standard tools and techniques that have been employed in the works described throughout the thesis. These methodologies contain the detailed description of different feature extraction and classification algorithms. Section 2.1 illustrates the feature extracting algorithms with required diagrams including autoregressive parameters, adaptive autoregressive parameters, Hjorth parameters, power spectral density estimates and wavelet based features. Section 2.2 describes the classification algorithms with required diagrams including SVM classifier, k-NN classifier, Naïve Bayesian classifier and Ensemble classifiers.

2.1 FEATURE EXTRACTION METHODS FOR EEG SIGNAL PROCESSING

From pattern recognition perspective, feature extraction can be defined as the procedure to derive a significant set of values (termed as "features") from an initial set of measured datapoints, which are intended to be informative and non-reduntant and to facilitate the subsequent steps of learning, generalization and enhanced human interpretations as well. This section provides an insight to the EEG feature extraction methods that include time domain features [1], [2] like Autoregressive Parameters (AR), Adaptive Autoregressive Parameters (AAR) and Hjorth Parameters, frequency domain features [3] like Power Spectral Density (PSD)/ Band Power Estimates and time-frequency correlated features [4] like Wavelet Decomposition [5] Based Features.

2.1.1 Autoregressive Parameters

The AR parameter model is considered to be one of the most well known feature extracting techniques for EEG based BCI paradigm. The primary reason of this popularity is the efficiency of this technique to represent the randomness of any signal accurately because AR parameters are estimated using methodical algorithms. Moreover, AR parameter model provides 'maximum entropy spectral estimation' which ensures that only reduced numbers of parameters are sufficient for accurate representation of a stochastic signal removing the need for signal averaging. Formally, this model is a basic parametric model for a time series signal.

Mathematically, the model can be illustrated using (2.1),

$$y_k = a_1 y_{k-1} + a_2 y_{k-2} + ...a_p y_{k-p} + x_k \tag{2.1}$$

where x_k denotes a zero mean Gaussian noise process calculated as $N\{0, \sigma_x^2\}$, p is the order of the AR model, y_{k-i} denotes the previous samples, a_i signifies the coefficients and i is an integer that can vary [0, p], and $y_{k\backslash}$ is the estimated output while index k is used to refer to distinct equidistant time instances. In the present works, we have used sixth order AR model, that is p=6. Usually, AR model is employed to represent any wide sense stationary stochastic time series that is characterized as,

$$< x_k > = \text{constant} \qquad \text{and} \qquad < x_k x_{k+d} >= r_d \tag{2.2}$$

To compute the AR model coefficients two methodologies have been adopted,

- Yule-Walker Method [6]

  For an AR model of order p, (2.1) can be expressed as,

$$y_k = \sum_{i=1}^{p} a_i y_{k-i} + x_k \tag{2.3}$$

By convention, y_k is assumed to have zero mean too. The estimation method includes multiplying both sides of (2.3) by x_{k-d} where d signifies a time delay, averaging and normalizing the obtained outcome. Since d can vary from 1 to p, repeating the above mentioned steps for p number of times yields a set of equations, known as Yule-Walker equations, which can be presented as matrix form as,

$$\begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & 1 & r_1 & \cdots & r_{p-2} \\ r_2 & r_1 & 1 & \cdots & r_{p-3} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_p \end{bmatrix} \tag{2.4}$$

- Burg's Method

  In contrast to Yule-Walker method, instead of estimating the AR model parameters directly, Burg's approach [7] computes the reflection coefficient at first and these reflection coefficients are treated as the last AR parameter estimate for each model of order p. From these, the AR parameter estimates are calculated using Levinson-Durbin algorithm. Given N discrete values, it is possible to compute the values of k coefficients using forward or backward linear prediction method,

$$f_n = -\sum_{i=1}^{k} a_i y_{n-i} \tag{2.5}$$

$$b_n = -\sum_{i=1}^{k} a_i y_{n+i} \tag{2.6}$$

Using the above two equations a_i is chosen in such a way that the forward error F_k and backward error B_k are minimized,

$$F_k = \sum_{n=k}^{N} (y_n - f_n)^2 = \sum_{n=k}^{N} (y_n - (-\sum_{i=1}^{k} a_i y_{n-i}))^2 \tag{2.7}$$

$$B_k = \sum_{n=0}^{N-k} (y_n - b_n)^2 = \sum_{n=0}^{N-k} (y_n - (-\sum_{i=1}^{k} a_i y_{n+i}))^2 \tag{2.8}$$

So basically this technique attempts to find μ such that the forward and backward errors are minimized while constraining the AR parameters to satisfy the Levinson-Durbin equation, given by,

$$A_{k+1} = A_k + \mu V_k \tag{2.9}$$

Here, initial estimate $A_0=1$ and $A_{k+1}= [1\ a_1\ a_2\ a_3.....a_k\ 0]$ and $V_{k+1}= [0\ a_k...\ a_3\ a_2\ a_1\ 1]$.

2.1.2 Adaptive Autoregressive Parameters

Because of the random behavior of EEG signals it is not wise to use stationary processes to estimate the model coefficients. To model a stochastic natured signal using AR model one can segment the entire data and then obtain AR model parameters for each segment. However, there also lies one tradeoff between the segment length and the error. The shorter is the length the resolution is more improved but it comes with a large error as well. Employing a sliding window for segmentation can solve the problem, but because of the computational overhead the method cannot be employed for online studies. In case of AAR parameter model[8], the randomness is incorporated by varying the AR parameters obtained with time. The coefficients are predicted in an adaptive way and it is assumed that in each iteration the AAR parameters are altered by an amount less than the prediction error, otherwise it becomes impossible to represent such non-stationary processes using AAR model. A number of AAR prediction algorithms are found in literature like Least Mean Squares (LMS), Recursive Least Squares (RLS), Recursive AR (RAR), Kalman Filtering and so on. Mathematically, it can be expressed as,

$$y_k = a_{1,k} y_{k-1} + a_{2,k} y_{k-2} + ...a_{p,k} y_{k-p} + x_k \tag{2.10}$$

Here x_k is computed as $N\{0, \sigma^2_{x,k}\}$ and $\sigma^2_{x,k}$ denotes the time varying variance of the noise process. Basically, it is a two step process,

- Calculation of the error coefficient
- Updating of the estimates using error coefficient.

In AAR model, the adaptation rate, smoothening of the parameters and time resolution is administered by the update coefficient, whose value is predetermined along with the order. The AAR model is chosen to meet the issues created by AR parameters,

- A stochastic model describes well the random behavior of the EEG.
- An adaptive method provides parameters with a high time resolution.
- As side-effect no frequency band has to be selected.

2.1.3 Hjorth Parameters

Hjorth parameters [9] indicate statistical properties for the analysis of EEG signals for feature extraction phase. This technique basically has three embedded parameters namely, activity, mobility and complexity. Activity signifies the variance of the signal that is basically a measure of the signal power. Mobility provides an outline about the measure of the signal's mean frequency and finally complexity generates the measure of the change in frequency. Apart from providing useful information about the frequency spectrum, this technique enables the users to analyze the signal in time domain as well; moreover this method allows a user to deal with a computationally less expensive framework. For a signal *y(n)* having length of *N*, activity *(A(y))*, mobility *(M(y))* and Complexity *(C(y))* can be defined as,

$$A(y) = \frac{\sum_{n=1}^{N} (y(n) - \bar{y})^2}{N-1} \tag{2.11}$$

$$M(y) = \sqrt{\frac{A(y')}{A(y)}} \tag{2.12}$$

$$C(y) = \frac{M(y')}{M(y)} \tag{2.13}$$

Here, $\bar{y}$ and y' refers to the mean and first order derivative respectively. In the present work, Hjorth parameters were calculated for each channel yielding total three features corresponding to activity, mobility and complexity for each channel. It is preferable to employ small windows in case of Hjorth parameters to accurately represent the non-stationarities associated with EEG signals.

2.1.4 Power Spectral Density

Spectral estimation is done to get the distribution of signal power [10] over the frequency range. For a finite data set $a(n)$ and its autocorrelation sequence X_{aa}, the Power Spectral Density (PSD) can be estimated as below,

$$P_{aa}(\omega) = \frac{1}{2\pi} \sum_{l=-\infty}^{+\infty} X_{aa}(l) e^{-j\omega l} \tag{2.14}$$

If the sampling frequency is denoted as f_s, then ω can be replaced as $\frac{2\pi f}{f_s}$ and above equation can be written as,

$$P_{aa}(f) = \frac{1}{f_s} \sum_{l=-\infty}^{+\infty} X_{aa}(l) e^{-j2\pi fl / f_s} \tag{2.15}$$

According to Nyquist criteria, the maximum frequency present in the system is half of the sampling frequency (f_s), the average power for entire signal over the entire Nyquist range is described as follows,

$$\int_{-f_s/2}^{f_s/2} P_{aa}(f) df \tag{2.16}$$

where $P_{aa}(f)$ represents power in an infinitesimal bandwidth, so it is termed as PSD.

- *Welch Method*

  Welch method [11] belongs to the category of non parametric PSD estimation methods where PSD of a signal is estimated from the signal itself, in other words using Welch method PSD estimate is computed using the Fourier Transform of the signal or the autocorrelation function of the signal.

  To compute the complete PSD estimate from a time varying EEG signal, the following steps are followed,

  First the signal is split up into overlapping time sequences. Then the segmented signals are passed through a suitably chosen window function; basically a window is applied to each of the segmented signals and further, Discrete Fourier Transform is computed and the result obtained is squared in terms of magnitude to obtain the periodograms and the individual periodograms thus obtained are time averaged to yield a final PSD measure.

  Basically, the Welch method comes up with two modifications in the traditional Bertlett's method that the subsequences formed are overlapping and instead of time averaging the periodograms, the modified periodograms serve the purpose.

  Mathematically, the signal can be split in *L* overlapping sequences of length *M*. The i^{th} signal segment can be expressed as,

$$a_i(n) = a(n + iD) \tag{2.17}$$

  where *n=0, 1, 2, …., M-1* and *i=0, 1, 2, …., L-1*.

Next, the signals are required to be windowed before computing the periodograms. Further, after the windowing functions are employed, the periodograms are calculated using,

$$\hat{P}_{aa}^{i}(f) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} a_i(n)w(n)e^{-j2\Pi fn}\right|^2 \tag{2.18}$$

where *w(n)* signifies the windowing function and U is a normalization factor defined as,

$$U = \frac{1}{M}\sum_{n=0}^{M-1} w^2(n) \tag{2.19}$$

Finally, Welch PSD estimate is calculated as the time average of the modified periodogram,

$$P_{aa}^{w} = \frac{1}{L}\sum_{i=0}^{L-1} \hat{P}_{aa}^{i}(f) \tag{2.20}$$

2.1.5 Wavelet Decomposition Based Features

Wavelet Transform [12] is an efficient mathematical tool for EEG signal analysis having a wide variety of applications. Classical Fourier Transform method has been successfully employed for stationary signal, but because of the stochastic nature of the EEG signals it is not always wise to use Fourier Transform (FT) directly for these EEG signals. FT of a signal provides localization in frequency domain while compromising time domain localization. Basically, there exists a tradeoff between time and frequency domain localization depending upon the size of the window. To deal with this dilemma, Discrete Wavelet Transform (DWT) [13] provides a flexible way of signal representation in frequency as well as temporal domain by varying the size of the windows. Moreover, DWT is more precise in localizing artifacts and transients and apart from that this technique provides multi resolution analysis by decomposing the signal into fine and coarse elements. The Continuous Wavelet Transform (CWT) of a signal *x(t)* is computed as an integral of the signal multiplied by scaled by (*s= 1/frequency*) and shifted version (by τ) of a wavelet function $\psi(t)$, which is also termed as the mother wavelet [14].

$$X_{CWT}(s,\tau) = \frac{1}{\sqrt{|s|}}\int_{-\infty}^{\infty} x(t)\psi^{*}\left(\frac{t-\tau}{s}\right)dt \tag{2.21}$$

Calculating CWT in this way for every possible scale is a computationally hectic work, instead representing s and τ in terms of power of two results in an easier analysis. Such analysis leads to DWT which can be expressed as,

$$X_{DWT}(j,k) = \frac{1}{\sqrt{|2^j|}} \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t-2^j k}{2^j}\right)dt \qquad (2.22)$$

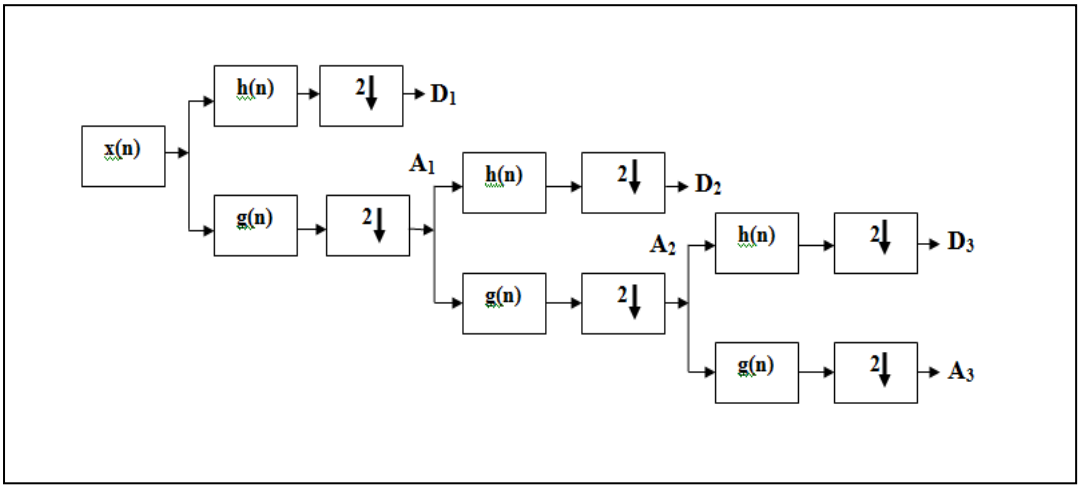


Fig. 2.1 Working Principle of DWT Decomposition

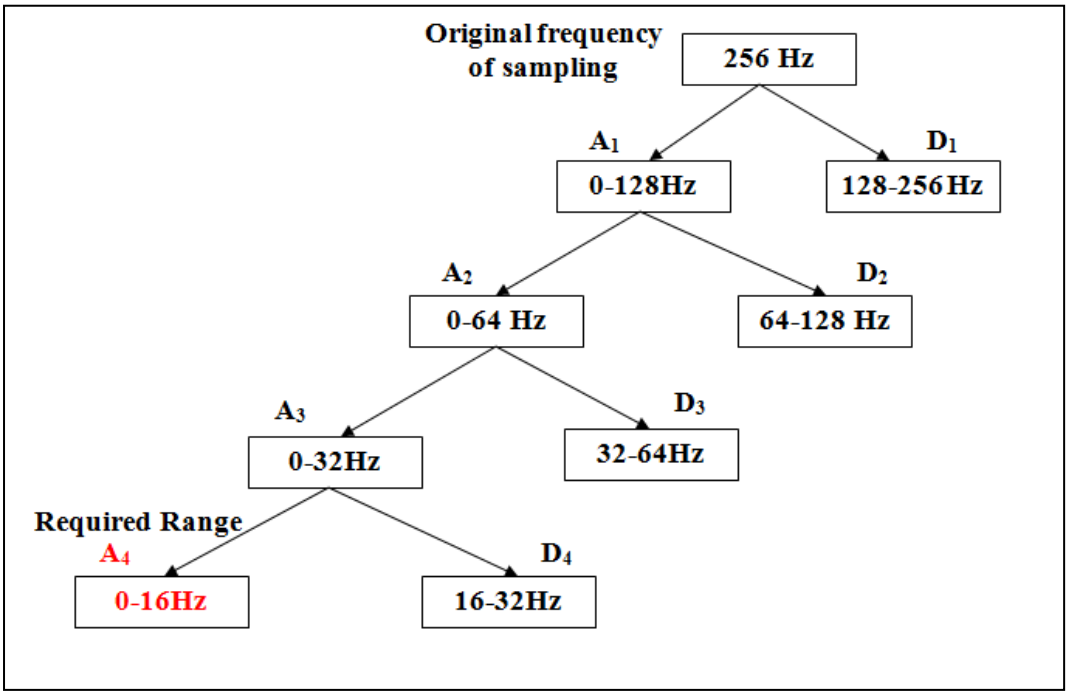


Fig 2.2 Signal Decomposition Using DWT

where s and τ are replaced by 2^j and $k\,2^j$ respectively. As shown in Fig. 2.1, the input signal is passed through a high pass filter (*h[.]*) and a low pass filter (*g[.]*) and the corresponding

filtered outputs (2.23) and (2.24) are down sampled by 2 at each level to generate the Approximate Coefficient (A_i) and the Detail Coefficient (D_i) from the low pass and high pass filter respectively.

Further, the Approximate Coefficient thus obtained is further decomposed in the similar procedure to get the Approximate and Detail Coefficients of the subsequent stages. The limit of decomposition is determined by the required frequency span of signals, for example in this case, the signals are recorded at 256 Hz frequency (Fig. 2.2), but only (0-16Hz) is likely to contribute relevant oscillatory components for motor imagery detection tasks, in this scenario, the required component will be A_4.The mother wavelet chosen in this work is Daubechies which has been found suitable for EEG recordings. The order is chosen to be 4.

$$y_{low}(n) = \sum_{k=-\infty}^{\infty} x(k)h(2n-k) \tag{2.23}$$

$$y_{high}(n) = \sum_{k=-\infty}^{\infty} x(k)g(2n-k) \tag{2.24}$$

2.2 CLASSIFICATION ALGORITHMS

Classification [15] is the next step that comes after Feature Extraction and performance of a motor imagery based BCI system depends almost entirely on the performance of the classifier used for the purpose. The main objective of an BCI system [16] lays mainly in the classification of the mental states from the raw EEG signals and differentiating those classified movements for actuation purpose. There are popular classifiers [17] that have beeen used for this purpose in the later chapters, like SVM (Support Vector Machine) [18], BPNN (Back Propagation Neural Network), k-NN (k-Nearest Neighbour), Naïve Bayes Classifier and Ensemble classifier.

2.2.1 SVM classifier

Primarily developed based on the concept of hyperplanes SVM is inherently a non-probabilistic linear classifier, while it can be extended for non-linear problems [19] as well. An SVM classifies the data samples in two different classes by constructing an N-dimensional hyperplane, in other words SVM searches for that direction of the hyperplane which leaves the maximum possible margin for both the classes [20]. In this case, the complexity does not depend on the dimension of the feature space, so SVM can be used for high dimensional feature spaces.

Let us consider the mathematical aspects of linear SVM. Suppose, x_i, i=1, 2,….,N denote the feature vectors of training set X, now the goal is to construct an optimal hyperplane (2.25),

$$g(x) = \omega^T x + \omega_0 \tag{2.25}$$

such that the training vectors are classified correctly in the two linearly separable classes ω_1, ω_2. In the above equation w indicates the direction of the hyperplane and ω_0 indicates the actual position. Now, SVM will search for the direction of the hyperplane leaving maximum margin from both the classes. The distance of a particular point from the hyperplane $g(x)$ can be expressed as, $z = \frac{|g(x)|}{\|\omega\|}$. Now, w and ω_0 can be scaled so that the value of $g(x)$ is +1 in case of ω_1 and -1 in case of ω_2. This is equivalent to,

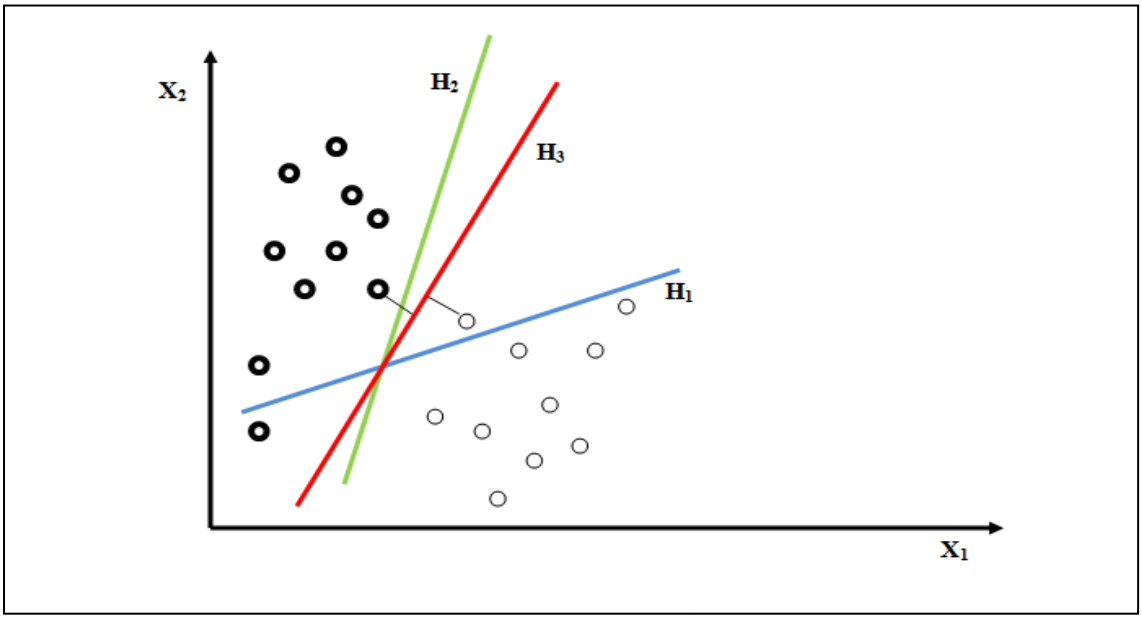


Figure 2.3 Linear SVM Classification, H_1 does not classify the data, H_2 classifies with very small margin but H_3 classifies with maximum margin

- Having a margin of $\frac{1}{\|\omega\|} + \frac{1}{\|\omega\|} = \frac{2}{\|\omega\|}$
- Requiring that $\omega^T x + \omega_0 \geq 1, \forall x \in \omega_1$

  $\omega^T x + \omega_0 \leq -1, \forall x \in \omega_2$

(2.26)

Let, y_i denote the class indicators such that $y_i = +1$ for ω_1 and $y_i = -1$ for ω_2. So, now the task is basically a quadratic optimization problem of computing the parameters *w* and ω_0 to minimize the objective function,

$$J(\omega, \omega_0) = \frac{1}{2}\|\omega\|^2 \text{ subject to the constraint } y_i(\omega^T x_i + \omega_0) \geq 1 \text{ for } i=1, 2, \ldots, N \quad (2.27)$$

The above problem can be solved by using the method of Karush-Kuhn-Tucker (KKT) conditions [21], after solving, since the Lagrange multipliers considered can either be zero or positive, the vector parameter *w* of the optimal solution comes out to be a linear combination of N_f number of feature vectors corresponding to $\lambda_i \neq 0$ such that $N_f \leq N$. These vectors are known as Support Vectors [22] and the corresponding classifier is termed as Support Vector Machine (SVM). Basically the datapoints can be broadly categorized into three sections,

- Datapoints lying beyond the margin (2.26) and correctly classified using the constraint part of (2.27)
- Datapoints lying within the margin (2.26) and correctly classified using (2.28)
- Datapoints lying within the margin (2.26) and incorrectly classified using (2.29)

$$0 \leq y_i(\omega^T x_i + \omega_0) < 1, i = [1, N] \quad (2.28)$$

$$y_i(\omega^T x_i + \omega_0) < 0, i = [1, N] \quad (2.29)$$

These issues can be addressed by introducing a new variable ξ_i, termed as slack variable with values ξ_i <0,0< $\xi_i \leq$1 and ξ_i >1 for the three cases respectively as expressed in (2.30),

$$y_i(\omega^T x_i + \omega_0) \geq 1 - \xi_i, i = [1, N] \quad (2.30)$$

The goal of the optimization therefore is now to maximize the margin while minimizing the number of points with ξ_i > 0. This can be stated by the minimization of $J(\boldsymbol{w}, \boldsymbol{\xi})$ given by (2.31) subject to the condition (2.30), where *C* is a positive constant that controls the relative influence of both the effects and $\xi_i \geq 0$ for *i=[1,N]*.

$$J(\omega, \omega_0) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}\xi_i \quad (2.31)$$

The nonlinear extension of SVM [23] is described in details in Chapter 3.

2.2.2 k-NN classifier

k-Nearest Neighbour classifier [24] belongs to the family of instance based classifiers where classification of an unknown data point can be done by relating the unknown point to certain points with known class labels based on some similarity measures/distance measures. This is a non-parametric supervised algorithm that does not require extensive training like other classifiers. Let us consider a data set consisting of samples where each sample is represented by an associated position vector and each component of the concerned vector corresponds to a distinct attribute characterizing the specific sample. Since it is a supervised algorithm [25] so the sample space must comprise of known class labels corresponding to each sample. To find the class of an unknown test sample the following two steps are carried out,

- Find out the *k* nearest neighbor of the test sample
- Determine the class of the sample from the *k* nearest neighbours.

Fig. 2.4 depicts the working principle of k-NN classifier with $k=3$ having attributes A_1 and A_2.The choice of the neighbours [26] is controlled by the selection of a specific similarity measure from a pool of similarity measures including Euclidean Distance, City Block Distance, Hamming Distance, Correlation values etc. The process of determination of the class label of the query sample is dependent on the choice of voting mechanism as well. Usually, majority voting technique is followed where the query sample is assigned to a class having maximum number of votes from the *k* nearest neighbours.

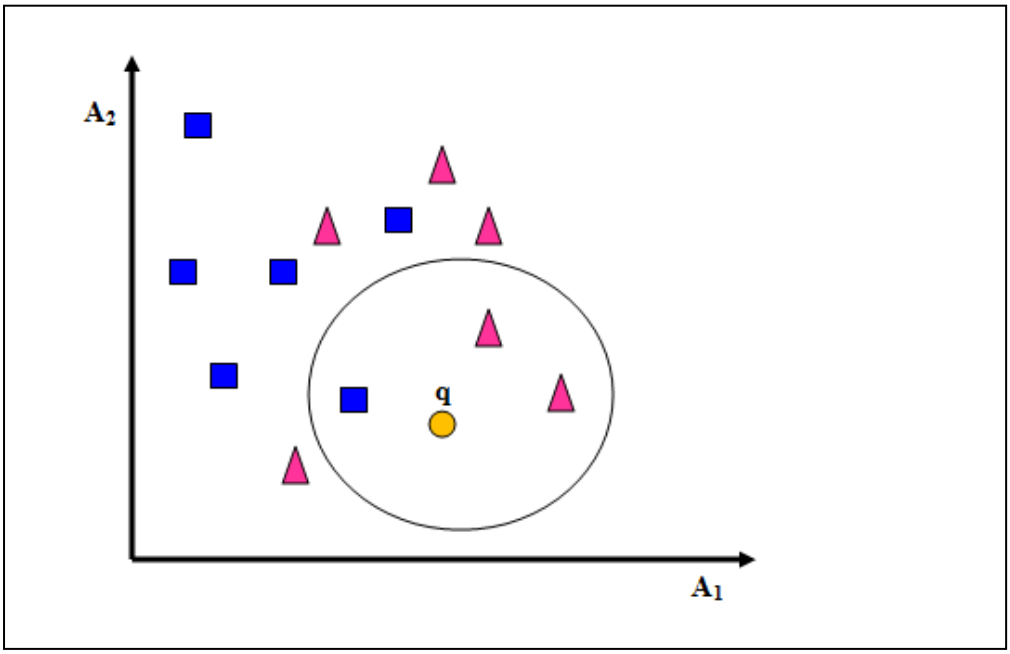


Fig. 2.4 k-NN Classification with *k*=3

Another commonly used approach is the inverse distance-weighted voting, where, closer the neighbours, the higher weight is associated with their votes which means that the

weights corresponding to the votes vary with the inverse of the distance or the inverse of the squared distance. The weighted votes are then summed up and the class with the highest vote is returned.

2.2.3 Naïve Bayesian Classifier

A Naïve Bayesian classifier [27] is a probabilistic classifier [28] that has been developed based upon the Bayes' theorem [29] with an assumption of independence among predictors. In other words, this method assumes that the occurrence of a feature under a particular class is independent of any other feature belonging to the same class. Apparently, this technique seems to be a simple one but has the capability of outperforming typically standard algorithms in terms of performance especially in case of high dimensional data. Given a variable *Y*, dependent upon features x_1, x_2, ...x_n, Bayes' theorem states,

$$posterior = \frac{prior \times likelihood}{evidence}$$

$$p(Y \mid x_1,...,x_n) = \frac{p(Y).p(x_1,...,x_n \mid Y)}{p(x_1,x_2,...,x_n)} \tag{2.32}$$

Using the assumption, the numerator is equivalent to the joint probability, which can be expanded using chain rule as,

$$\begin{aligned} p(Y,x_1,...,x_n) &= p(Y)\,p(x_1,...,x_n \mid Y) \\ &\approx p(Y).p(x_1 \mid Y).p(x_2 \mid Y,x_1).....p(x_n \mid Y,x_1,x_2,...x_{n-1}) \end{aligned} \tag{2.33}$$

To simplify the complicated problem, it is assumed that each feature x_i is independent of every other feature x_j such that $i \neq j$. Thus it is obtained,

$$\left.\begin{aligned} p(Y \mid x_1,...,x_n) &\propto p(Y,x_1,...,x_n) \\ &\propto p(Y).p(x_1 \mid Y).p(x_2 \mid Y)...p(x_n \mid Y) \\ &\propto p(Y).\prod_{i=1}^{n} p(x_i \mid Y) \end{aligned}\right\} \tag{2.34}$$

This model is combined with the Maximum a Priority (MAP) rule to produce the Naïve Bayesian classifier which classifies according to the *classfunc* described as,

$$classfunc(X_1,X_2,...X_n) = \arg\max\, p(Y=y)\prod_{i=1}^{n} p(x_i = X_i \mid Y = y) \tag{2.35}$$

The commonly used probability distribution of the features include normal distribution, multinomial or multivariate distributions while the class priors may be calculated by

assuming either equiprobable classes or from the estimation of the fraction of the total number of samples that belong to a particular class.

2.2.4 Ensemble Classifiers

Ensemble classifiers [30] refer to the group of individual distinct classifiers that are cooperatively trained on a data set for a supervised classification problem to enhance the performance of individual classifiers. This technique is often referred to as a "multiclassifier" technique and the component classifiers are known as *base classifiers* or *weak learners*. Each of the base classifiers performs the classification task separately and the final output of the ensemble classifier is obtained after incorporating the decisions of all the classifiers on the test data set. Because of the fusion of so many classifiers it is likely that the concerned ensemble classifier should provide superior performance in comparison to the individual component classifiers. The two well known principles for implementation of ensemble classifiers include bagging and boosting [31].

In bagging the training subsets are randomly drawn (with replacement) from the training set. Homogeneous base classifiers are trained on the subsets. The class chosen by most base classifiers is the considered to be the final verdict of the ensemble classifier. There are a number of variants of bagging and aggregation approaches including random forests and large scale bagging [32]. Boosting [33] creates data subsets for base classifier training by re-sampling the training patterns, however, by providing the most informative training pattern for each consecutive classifier. Each of the training patterns is assigned a weight that determines how well the instance was classified in the previous iteration. The training data that are wrongly classified is included in the training subset for the next iteration. AdaBoost is a more generalized version of boosting.

2.2.5 Artificial Neural Network

A typical artificial neuron [34] is a real world electrical analogue of biological neurons performing similar functions. Mathematically, it is represented by two sections, i) a linear activation/inhibition module and ii) a non-linearity that limits the signal level within a finite band as shown in Fig. 2.5 Here, the weighted summer plays the role of a cell body by representing the linear combiner module producing *Net* from the inputs. The synapse in an artificial neuron is modelled as a nonlinear function like step function, signum function, sigmoid function, *tanh* function etc. that is implemented in the phase of obtaining *Out* from *Net*.

$$Net = \sum_{i=1}^{n} w_i x_i \tag{2.36}$$

$$Out = u(Net - Threshold) \tag{2.37}$$

$$Out = \text{sgn}(Net) \tag{2.38}$$

$$Out = \frac{1}{1 + \exp(-Net)} \tag{2.39}$$

Learning or adaptation of weights in an artificial neural network can be realized in four distinct methods, 1) Supervised Learning [35], 2) Unsupervised Learning [36], 3) Reinforcement Learning and 4) Competitive Learning. In the present work mainly supervised learning procedures have been adopted. In case of supervised learning, it always requires a trainer who would feed the network with output-input instances [37]. To illustrate further, an input vector *X* and a target vector *T* is provided and the objective is to adapt the weights in such a way that the output is *T*. Usually, in the first iteration the weights are adjusted randomly and then the error *E=T-X* is calculated and the weights are updated at the successive iterations to move closest to the output *T*.

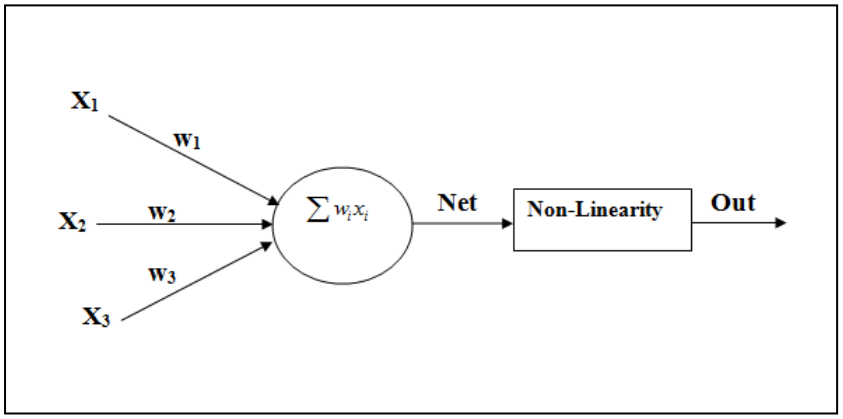


Fig. 2.5 Artificial Neural Network

Amongst classical supervised learning algorithms, Back Propagation [38] is considered as one of the most efficient and well known algorithm. It is formulated by employing a feed-forward architecture of neurons, having a number of layers (input, output, intermediate) and each layer containing a number of neurons. Each neuron is responsible to produce a weighted sum of the inputs associated with it from the previous layer and finally passing on the summation of all these results to the non-linear segment. The error computed at the output layer can be expressed as (2.40), where T_i and Out_i signifies the target and output produced at *i*-th neuron of the output layer.

$$E = \frac{1}{2}\sum_{\forall i}(T_i - Out_i)^2 \tag{2.40}$$

Steps of Back Propagation algorithm can be listed as follows,

1. Initialize instance *i*=1.
2. Supply input components for the *i*-th instance to the input of the neural net; make a forward pass and compute the outputs.
3. Calculate the corresponding error vector E_i by considering the componentwise difference between the output vector and the target vector as $E_{ij} = T_{ij} - O_{ij}, \forall j$, where E_{ij}, T_{ij} and O_{ij} denote the *j*-th component of *i*-th error vector, target vector and output vector respectively.
4. Repeat steps 2 and 3 for *i*=1 to *n*.
5. Determine the root mean square value of error, denoted by **Error**, whose *j*-th component is defined as,

$$Error_j = \sqrt{\frac{\sum_{i=1}^{n} E_{ij}^2}{n}}$$

6. Back propagate the error from the RMS value of the error components of the last layer to the preceding layers and adapt the weights of the network layerwise starting from the last layer.
7. Repeat steps 2 to 6 until $\sum_{\forall j}(Error_j)^2$ is negligibly small or the maximum iteration limit is reached.

Weight adaptation is also an important aspect in case of artificial neural networks, among numerous methods the traditional Gradient Descent Learning has been utilized here.

- Gradient Descent Search (GDS)

According to Gradient Descent Learning [39] the weights are adapted [40] using the following rule,

$$w_{new} = w + \triangle w \tag{2.41}$$

where, $\triangle w = -\eta \frac{\partial E}{\partial w}$. Here, *E* is the error, *w* denotes the weights and η is the learning rate ($0 < \eta < 1$). Employing this rule, there is an extreme tendency of getting trapped at local minima. To avoid dealing with this limitation, a momentum term (*mc*) has been incorporation in the equation which helps the solutions to slide through the local minimum points on the

error surface and reach a global minimum. The modified weight adaptation rule is formally known as Gradient Descent Search with Momentum (GDS-M) and can be expressed as,

$$\triangle w = mc.\triangle w_{prev} + (1-mc)\eta \frac{\partial E}{\partial w} \tag{2.42}$$

where $\triangle w_{prev}$ denotes the previous weight alteration.

REFERENCES

[1] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, 1970.

[2] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl, "Time domain parameters as a feature for EEG-based brain–computer interfaces," *Neural Networks*, vol. 22, no. 9, pp. 1313–1319, 2009.

[3] A. Alkan and A. S. Yilmaz, "Frequency domain analysis of power system transients using Welch and Yule–Walker AR methods," *Energy Convers. Manag.*, vol. 48, no. 7, pp. 2129–2135, 2007.

[4] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in EEGs using time–frequency analysis," *Inf. Technol. Biomed. IEEE Trans.*, vol. 13, no. 5, pp. 703–710, 2009.

[5] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Syst. Appl.*, vol. 32, no. 4, pp. 1084–1093, 2007.

[6] G. Eshel, "The yule walker equations for the AR coefficients," *Internet Resour.*, vol. 2, pp. 68–73, 2003.

[7] C. Collomb, "Burg's Method, Algorithm and Recursion," *K. Elissa, Unpubl.*, 2010.

[8] G. Pfurtscheller, C. Neuper, A. Schlögl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *Rehabil. Eng. IEEE Trans.*, vol. 6, no. 3, pp. 316–325, 1998.

[9] M. Vourkas, S. Micheloyannis, and G. Papadourakis, "Use of ann and hjorth parameters in mental-task discrimination," in *Advances in medical signal and information processing, 2000. First international conference on (IEE conf. publ. no. 476)*, 2000, pp. 327–332.

[10] F. Cona, M. Zavaglia, L. Astolfi, F. Babiloni, and M. Ursino, "Changes in EEG power spectral density and cortical connectivity in healthy and tetraplegic patients during a motor imagery task," *Comput. Intell. Neurosci.*, vol. 2009, 2009.

[11] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, 1967.

[12] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 11, no. 7, pp. 674–693, 1989.

[13] N. Hazarika, J. Z. Chen, A. C. Tsoi, and A. Sergejew, "Classification of EEG signals using the wavelet transform," in *Digital Signal Processing Proceedings, 1997. DSP 97., 1997 13th International Conference on*, 1997, vol. 1, pp. 89–92.

[14] J. Rafiee, M. A. Rafiee, N. Prause, and M. P. Schoen, "Wavelet basis functions in biomedical signal processing," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 6190–6201, 2011.

[15] S. Theodoridis, A. Pikrakis, K. Koutroumbas, and D. Cavouras, *Introduction to Pattern Recognition: A Matlab Approach: A Matlab Approach*. Academic Press, 2010.

[16] A. Konar, *Computational intelligence: principles, techniques and applications*. Springer Science & Business Media, 2006.

[17] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, p. R1, 2007.

[18] T. Joachims, *Text categorization with support vector machines: Learning with many relevant features*. Springer, 1998.

[19] A. Saha, A. Konar, P. Das, B. Sen Bhattacharya, and A. K. Nagar, "Data-point and feature selection of motor imagery EEG signals for neural classification of cognitive tasks in car-driving," in *Neural Networks (IJCNN), 2015 International Joint Conference on*, 2015, pp. 1–8.

[20] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.

[21] T. M. Mitchell, "Machine learning," *Mach. Learn.*, 1997.

[22] L. M. Manevitz and M. Yousef, "One-class SVMs for document classification," *J. Mach. Learn. Res.*, vol. 2, pp. 139–154, 2002.

[23] I. Güler and E. D. Ü. Beyli, "Multiclass support vector machines for EEG-signals classification," *Inf. Technol. Biomed. IEEE Trans.*, vol. 11, no. 2, pp. 117–126, 2007.

[24] B. Blankertz, G. Curio, and K.-R. Muller, "Classifying single trial EEG: Towards brain computer interfacing," *Adv. Neural Inf. Process. Syst.*, vol. 1, pp. 157–164, 2002.

[25] M. Pal, "DESIGN OF AN EEG-DRIVEN ARTIFICIAL LIMB SYSTEM," 2014.

[26] D. Wang, D. Miao, and C. Xie, "Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection," *Expert Syst. Appl.*, vol. 38, no. 11, pp. 14314–14320, 2011.

[27] U. R. Acharya, F. Molinari, S. V. Sree, S. Chattopadhyay, K.-H. Ng, and J. S. Suri, "Automated diagnosis of epileptic EEG using entropies," *Biomed. Signal Process. Control*, vol. 7, no. 4, pp. 401–408, 2012.

[28] P. LeVan, E. Urrestarazu, and J. Gotman, "A system for automatic artifact removal in ictal scalp EEG based on independent component analysis and Bayesian classification," *Clin. Neurophysiol.*, vol. 117, no. 4, pp. 912–927, 2006.

[29] P. Langley and S. Sage, "Induction of selective Bayesian classifiers," in *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence*, 1994, pp. 399–406.

[30] R. Polikar, "Ensemble based systems in decision making," *Circuits Syst. Mag. IEEE*, vol. 6, no. 3, pp. 21–45, 2006.

[31] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants," *Mach. Learn.*, vol. 36, no. 1–2, pp. 105–139, 1999.

[32] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization," *Mach. Learn.*, vol. 40, no. 2, pp. 139–157, 2000.

[33] T. Windeatt, "Accuracy/diversity and ensemble MLP classifier design," *Neural Networks, IEEE Trans.*, vol. 17, no. 5, pp. 1194–1211, 2006.

[34] B. Yegnanarayana, *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009.

[35] L. Tarassenko and S. Roberts, "Supervised and unsupervised learning in radial basis function classifiers," in *Vision, Image and Signal Processing, IEE Proceedings-*, 1994, vol. 141, no. 4, pp. 210–216.

[36] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller, "Towards adaptive classification for BCIPart of the 3rd Neuro-IT and Neuroengineering Summer School Tutorial Series.," *J. Neural Eng.*, vol. 3, no. 1, p. R13, 2006.

[37] S. Datta, "COGNITIVE AND HAPTIC MODALITIES OF PERCEPTUAL ABILITY ANALYSIS FOR NEXT GENERATION HUMAN COMPUTER INTERFACES," 2014.

[38] D. J. C. MacKay, "A practical Bayesian framework for backpropagation networks," *Neural Comput.*, vol. 4, no. 3, pp. 448–472, 1992.

[39] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural networks*, vol. 6, no. 4, pp. 525–533, 1993.

[40] H.-C. Hsin, C.-C. Li, M. Sun, and R. J. Sclabassi, "An adaptive training algorithm for back-propagation neural networks," in *Systems, Man and Cybernetics, 1992., IEEE International Conference on*, 1992, pp. 1049–1052.

Chapter 3

Evolutionary Perspective for Optimal Selection of EEG Electrodes and Features

This chapter proposes a novel evolutionary approach to the optimal selection of electrodes as well as relevant EEG features for effective classification of cognitive tasks. The problem has been formulated in the framework of a single objective optimization problem with an aim to simultaneously satisfying three criteria. The first criterion deals with maximization of the correlation between the features of Electroencephalogram (EEG) sources before and after the selection of optimal electrodes. The second criterion is concerned with minimization of the mutual information between the features of the selected EEG electrodes. The last criterion aims at maximization of the ratio of the difference between the selected features of the EEG sources between and within any two cognitive tasks. A self-adaptive variant of Firefly Algorithm (FA) referred to as SAFA, is proposed to solve the above optimization problem by proficiently balancing the trade-off between the computational accuracy and the run-time complexity. The chapter is divided into five sections. Section 3.1 presents a brief introduction and describes the related works in this particular research genre. Section 3.2 provides a detailed description of the proposed framework. Section 3.3 recapitulates the traditional FA and presents the proposed self-adaptive FA (SAFA). The experimental results are reported in section 3.4. Section 3.5 concludes the paper.

3.1 INTRODUCTION

Brain computer interfacing (BCI) [1] is a multi-dimensional field of research, concerned with cognition, neurophysiology, psychology, sensors, machine learning, signal detection and processing, to name a few. Now a day, BCI [2] stands alone as the only modality of control and communication for patients suffering from diseases like amyotrophic lateral sclerosis, paralysis, cerebral palsy, and amputees [1]. Its contributions in medical fields range from prevention to neuronal rehabilitation for serious injuries. BCI addresses analyzing, conceptualization, monitoring, measuring, and evaluating the complex neuro-physiological behaviors detected and extracted from a set of electrodes over the scalp or from those implanted inside the brain.

These BCI interfaces bypass the natural pathways of neuro-muscular control and thus aim at serving an alternative means of communication/control in case of failure in neural/motor functioning. Several interfacing methodologies including invasive implants, semi invasive implants like electrocorticography (ECoG) [3] and non invasive modalities like electroencephalogram (EEG) [4], magnetoencephalogram (MEG) [5] and functional magnetic resonance imaging (fMRI) [6] have emerged in order to implement BCI successfully. EEG is the preferred technology for measuring brain activities for most BCI researchers because of its non-invasiveness, portability, easy availability, and high temporal resolution.

The basic BCI module consists of three steps, including i) pre-processing of the EEG signals dealing with artifact removal, identification of relevant electrodes and frequency bands of EEG signals, ii) feature extraction, and iii) classification, concerned with identification of different mental states. The classified results thus obtained, lead to the generation of the control signals required to drive an assistive device. The classification accuracy relies on the extent of detour the redundant information. This chapter addresses two crucial factors for effective classification of cognitive tasks using EEG based BCI systems, including,

- Optimal selection of electrodes [5] to facilitate faster processing of EEG signals for different cognitive tasks,
- Optimal selection of relevant EEG features to enhance the performance of a classifier.

The optimal selection of electrodes [5] is essentially influenced by the estimation of cortical sources. The EEG devices acquire raw cortical current signals, generated from different independent sources, through neuronal firing in outer cortex of the brain. These signals are then transformed to respective voltage signals by passing through different resistive devices. Finally, the voltage signals are recorded by placing electrodes at specified

scalp regions. Due to volume conduction, the signal acquired at the scalp electrodes is found to contain components of different sources. Moreover, for a particular cognitive task every electrode placed on the scalp cannot provide relevant information, in fact at times electrodes generate redundant information. Hence, optimal electrode selection is very important for signal analysis and relevant decision making in the following steps. Otherwise overlapping information will not only degrade performance metrics, but it involves processing of same signal components more than once, which negatively affects the time complexity as well.

Only optimal electrode selection [7] is not sufficient for a successful EEG based BCI implementation [8]. One of the significant concerns in BCI research is to deal with the high dimensionality of the features. Often it is observed that due to the presence of a large number of redundant features in the feature set, the accuracy of the classifier is greatly decreased. Researchers are now taking keen interest to select fewer discriminate features from the high dimensional EEG feature vectors for different cognitive tasks without sacrificing the classification accuracy. The chapter proposes a novel evolutionary approach to automatic selection of optimal set of EEG electrodes and EEG features (from the high dimensional feature space). The principle of evolutionary electrode and feature selection is outlined next.
In this chapter, the possible selection of EEG electrodes is realized by optimizing the scoring function, which deals with

- Maximizing the association between the estimated source signals corresponding to the original set and the reduced set of electrodes, and
- Minimizing the mutual information between two selected electrodes.

The first criterion reduces the loss in information of the cortical sources corresponding to a cognitive task after reducing the number of electrodes. The second criterion aims at identifying the relevant electrodes conveying unique information of specific cognitive tasks, thus discarding the redundant information.

The design philosophy adopted for the optimal selection of EEG features is to identify the set of features that are capable

- To uniquely represent a specific class of cognitive tasks and
- To effectively differentiate between any two classes.

It is realized by jointly serving the following two criteria. First, the selected *j*-th features of the data points in a given class should be close to each other. Contrarily, the difference between the means of the selected *j*-th feature of any two classes should be as high as possible.

A self-adaptive variant of the traditional firefly algorithm (FA) [9] is proposed here to select an optimal set of appropriate EEG electrodes and features by jointly optimizing the

above-mentioned objectives. FA is selected here partly heuristically and partly due to its established performance with respect to computational accuracy and run-time complexity[9].The self-adaptive variant of FA [10] assists the potential candidate solutions (of the optimization problem) to confine their search in their local neighborhood in the parameter space. On the other hand, the inferior members are equipped with global exploration capability.

The present work aims at improving the work proposed in [11] in three important aspects. First, the work proposed in [11] is primarily concerned with the optimal electrode selection [7] for a specific cognitive task. However, in this chapter, we aim at selecting the optimal electrodes for effective classification of different cognitive tasks [12]. This is a more realistic scenario of practical BCI applications. Second, Pearson correlation coefficient is utilized in [11] to capture the degree of relationship between the information conveyed by the original set of electrodes and the reduced number of selected electrodes. This correlation measure is found to be only sensitive to the linear relationship between two variables, incompetent to model non-linear or monotonous relationships. In other words, a zero value of Pearson coefficient does not always imply independence between two variables. Thus, it may fail to proficiently capture the dependence between two EEG sources. This is here circumvented by using an alternative measure of correlation, referred to as the distance correlation coefficient [13]. Third, the present work also attempts to select the optimal selection of EEG features along with the EEG electrodes. This improves the classification accuracy.

3.2 PROPOSED METHODOLOGY

This section provides a detailed overview of the proposed framework (Fig. 3.1). The proposed method involves a training data set $\boldsymbol{T}$, which is pictorially illustrated in Fig. 3.2, for a specific cognitive class K_c. The training dataset consists of L data points, each having information of N electrode (sink) signals. Each sink signal is represented as a F-dimensional feature vector. The chapter aims at the optimal selection of M ($\leq N$) electrodes and D ($\leq F$) features for effective classification of C cognitive tasks. It is to worth mentioning that K_c=[1, C]. Here $S_{i,k}^{t,c}$ and $S_{i,k}^{t,c}$ denote the k-th feature of the i-th electrode and the corresponding source for the t-th data point in the class K_c.

3.2.1 Independent Component Analysis as a Source Localization Tool

There are many sources inside human brain, which produce current signals due to neuronal firing during any cognitive task. The voltage signals recorded by the EEG electrodes (also

referred to as sink signals) placed on a human scalp are essentially mixtures of these source signals. In brain imaging, the problem of estimating the locations and distributions of the cortical sources, based on the voltage readings of the EEG electrodes, remains as an 'ill-posed' blind signal separation problem. Independent component analysis (ICA) [14] provides a solution to this problem [11] by transforming a multivariate signal (for instance, EEG signal) into a linear combination of independent non-Gaussian subcomponents (for example, non-Gaussian source signals).

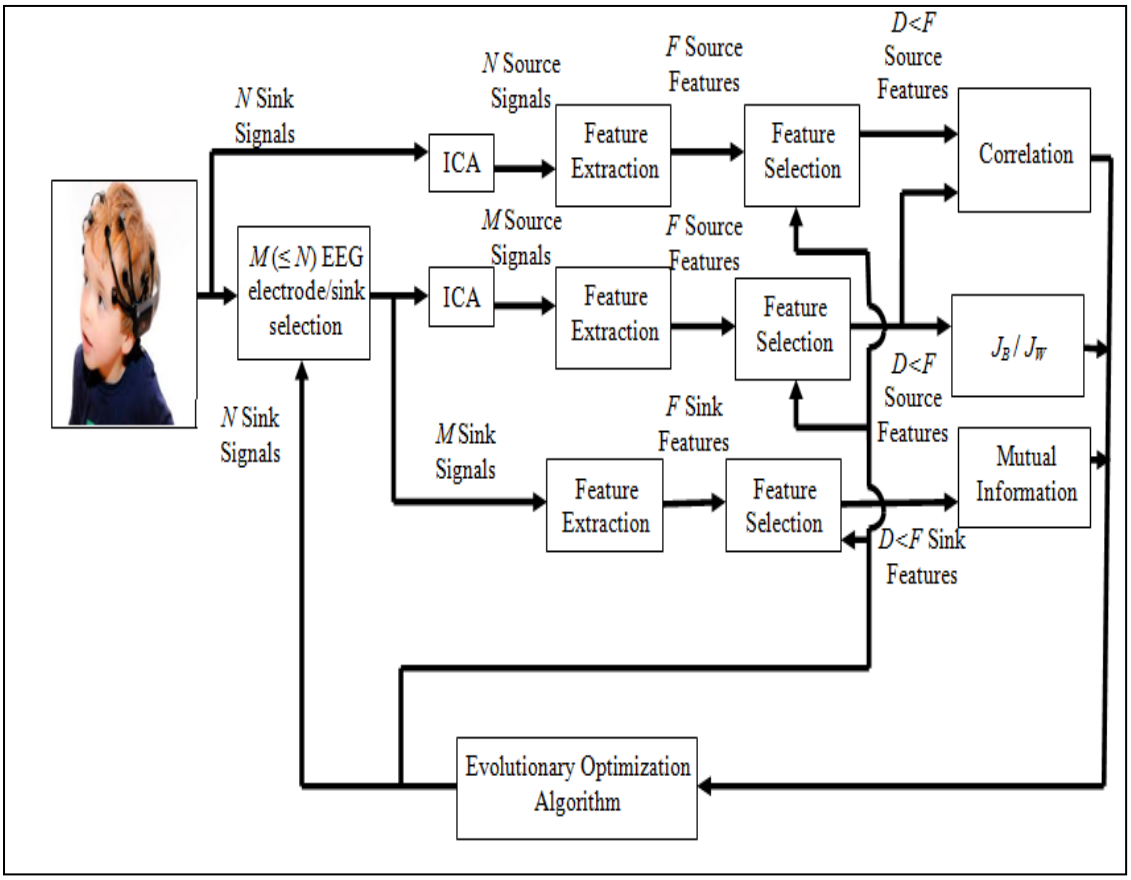


Fig.3.1 Overview of the Proposed Scheme

Let, $\mathbf{P} = [\vec{P}_1, \vec{P}_2, ..., \vec{P}_N]^T$ and $Q = [\vec{Q}_1, \vec{Q}_2, ..., \vec{Q}_N]^T$ be the N source and sink signals in time domain. ICA deals with representing each $\vec{Q}_i$ as a linear combination of $\vec{P}_j$ for j=[1, N], given by

$$\vec{Q}_i = w_{i,1}\vec{P}_1 + w_{i,2}\vec{P}_2 + ... + w_{i,N}\vec{P}_N = \sum_{j=1}^{N} w_{i,j}\vec{P}_j \quad \text{for } i=[1, N] \qquad (3.1)$$

or

$$\mathbf{Q = WP} \qquad (3.2)$$

with $w_{i,j}$ representing the degree of dependence of $\vec{Q}_i$ on $\vec{P}_j$ based on their distance and $\boldsymbol{W}$ denoting the mixing matrix with elements for i, j=[1, N]. ICA aims at determining the source

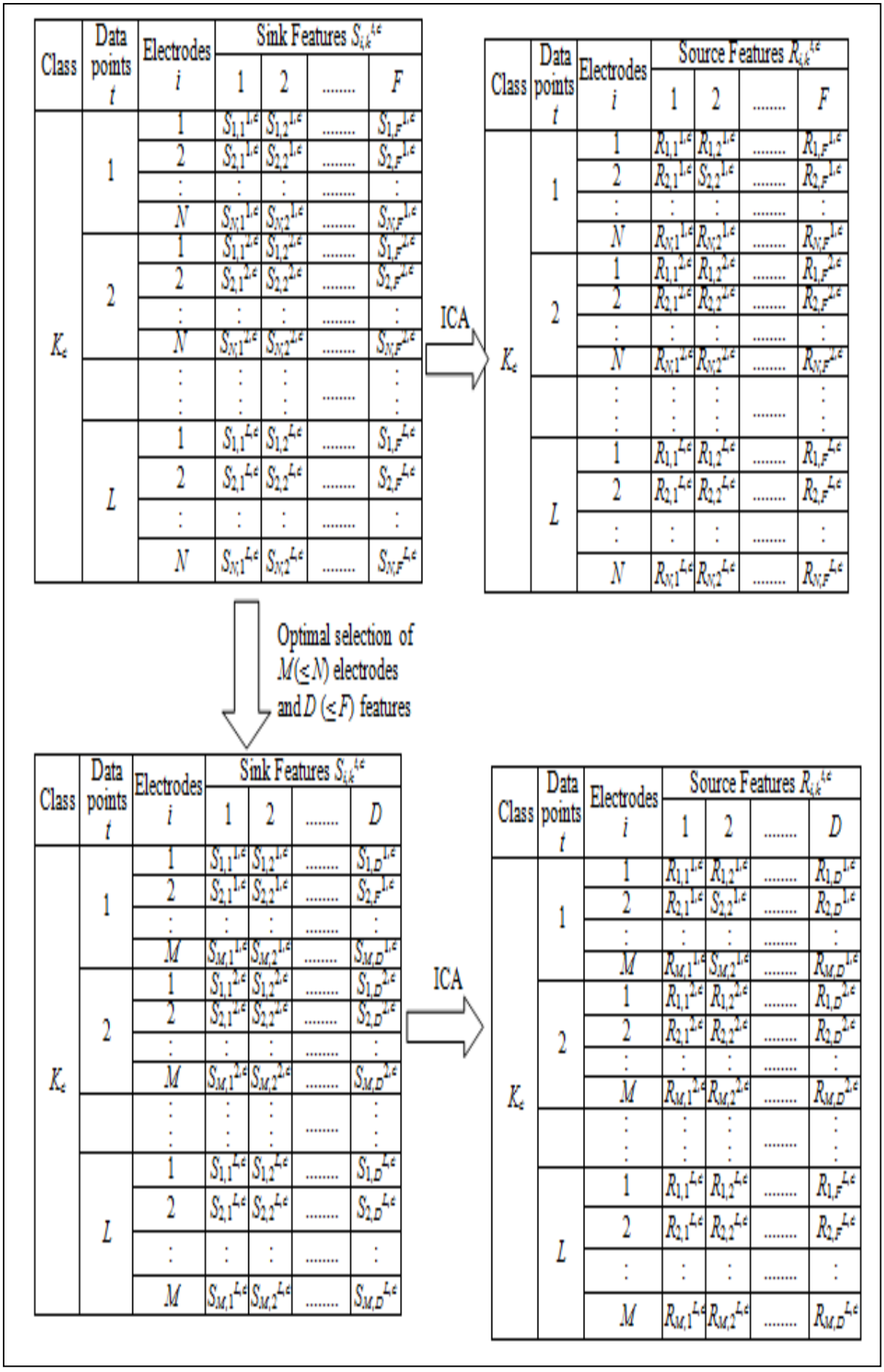


Fig. 3.2 Training Data Set for Cognitive Task K_c

signals by identifying the optimal mixing matrix, obtained by the minimization of the mutual information or the maximization of the non-Gaussianity of the estimated source signals. The members of former family use measures like K-L divergence and maximum entropy while the latter family utilizes kurtosis, negentropy and so on. In the present chapter, we employed a variant of ICA based on the Infomax algorithm [11] that uses the maximum entropy measure. Once the mixing matrix $\boldsymbol{W}$ is estimated, the source signals can be derived by multiplying the observed sink signals with the inverse of the mixing matrix, also known as demixing matrix $\boldsymbol{V}=\boldsymbol{W}^{-1}$, by

$$\mathbf{P} = \mathbf{W}^{-1}\mathbf{Q} = \mathbf{V}\mathbf{Q} \tag{3.3}$$

3.2.2 Feature Extraction

EEG signals of the sinks for two different cognitive tasks, as well as the sources, have been represented using five well-known features. The feature set consists of common spatial pattern feature (CSP) [7], two time domain features, including adaptive autoregressive (AAR) Parameters [11][15], and Hjorth parameters [11], a frequency domain feature, including power spectral density (PSD) [11] and a set of time-frequency correlated features, including discrete wavelet coefficients [11].

3.2.3 Optimum Features and Electrode Selection using Evolutionary Approach

The performance of any real world optimization algorithm [16], as in the present context, greatly relies on the judicial formulation of the objective function. In the present scenario, the optimal selection of M ($\leq N$) EEG electrodes and D ($\leq F$) salient EEG features for classification of two cognitive tasks is realized based on three significant performance characteristics. The first two characteristics deal with the optimal selection of electrodes, while the third one is concerned with the selection of optimal feature.

1. For a specific cognitive class K_c and t-th data point, the correlation between the D features of the source signals estimated from the optimally selected M electrodes and that of the original N electrodes should be maximized.

One of the major objectives in the present context is to preserve the relevant information of the recorded EEG signals for accurate classification of cognitive tasks, even after selecting the optimal set of M electrodes out of N. A higher degree of association between the sets of source features corresponding to the original N and the selected M ($\leq N$) sink signals indicate an efficient representation of the original number (N) of sink signals with the reduced number (M) of sinks signals. The degree of relationship here is captured by *distance correlation* [13]

between the feature space of *M* and *N* source signals. The distance correlation is an index in [0, 1] representing the statistical dependence between two sets of source features.

Distance correlation between any two sets of *D*-dimensional source features, $\vec{R}_i^{t,c}$ and $\vec{R}_j^{t,c}$, (for *t*-th data point of cognitive class K_c) is determined by,

$$\phi_{i,j}^{t,c} = \frac{dC_{i,j}}{\sqrt{dV_i \times dV_j}} \tag{3.4}$$

where

$$dC_{i,j} = \frac{1}{D}\sqrt{\sum_{k,l=1}^{D} A_{k,l}B_{k,l}} \tag{3.5}$$

and

$$\left.\begin{aligned} dV_i &= \frac{1}{D^2}\sum_{k,l=1}^{D} A_{k,l}^2 \\ dV_j &= \frac{1}{D^2}\sum_{k,l=1}^{D} B_{k,l}^2 \end{aligned}\right\} \tag{3.6}$$

with the components of the matrix **A** being determined by,

and

$$\left.\begin{aligned} a_{k,l} &= \left\| R_{i,k}^{t,c} - R_{i,l}^{t,c} \right\|, \quad \text{for } k,l = [1,F] \\ A_{k,l} &= a_{k,l} - \bar{a}_k - \bar{a}_l + \bar{a} \quad \text{for } k,l = [1,F] \\ \bar{a}_k &= \frac{1}{F}\sum_{l=1}^{F} a_{k,l}, \qquad \bar{a}_l = \frac{1}{F}\sum_{k=1}^{F} a_{k,l} \\ \bar{a} &= \frac{1}{F^2}\sum_{k,l=1}^{F} a_{k,l}. \end{aligned}\right\} \tag{3.7}$$

The components of matrix **B** are determined similarly as in (3.7), however, using the features of $\vec{R}_j^{t,c}$. Here ||.|| denotes the Euclidean norm. A value of $\phi_{i,j}$=0 is an implication of independence of $\vec{R}_i^{t,c}$ and $\vec{R}_j^{t,c}$. Distance coefficient is here adopted as a measure of correlation between source features, instead of Pearson correlation coefficient, to overcome its incapability to characterize non-linear or monotone dependency [13]. An accurate selection of optimal electrodes will increase the distance correlation, indicating that the source features corresponding to the original and reduced set of electrodes co-vary with each other.

2. For a specific cognitive class K_c and *t*-th data point, the mutual information between the *D* features of two selected electrodes should be minimized.

From the point of view of the present problem, in order to maximize the unique information content of two maximally distinct selected sinks, the redundant information is to be reduced. There are possibilities that for a particular feature, any two electrodes residing in

close proximity are providing similar information. Therefore, inclusion of any one between the two electrodes (instead of considering both) will serve the purpose with reduced computational complexity. The similarity in the information content of two sink feature sets is modeled here by their mutual information (MI).

Based on information theory, MI [17] is used to identify the amount of uncertainty about one sink feature vector $\vec{S}_i^{t,c}$ given knowledge of the other sink feature vector $\vec{S}_j^{t,c}$ (for cognitive class K_c and t-th data point). It equals zero if they are independent. Thus, more the mutual information between them, less the uncertainty in $\vec{S}_i^{t,c}$ given the knowledge of $\vec{S}_j^{t,c}$ or vice versa and hence, only one of them should be selected. Hence $MI_{i,j}^{t,c}$ is utilized here as a dependence measure of $\vec{S}_i^{t,c}$ on $\vec{S}_j^{t,c}$ and vice versa, given by

$$MI_{i,j}^{t,c} = H_i - H_{i|j} \tag{3.8}$$

$$H_i = -\sum_{k=1}^{D} p_i(k) \times \log_2 p_i(k) \tag{3.9}$$

$$H_{i|j} = -\sum_{k=1}^{D} p_{i,j}(k) \times \log_2 p_{i|j}(k) \tag{3.10}$$

where, the average information or entropy H_i is the uncertainty in $\vec{S}_i^{t,c}$ before observing $\vec{S}_j^{t,c}$ and the conditional entropy $H_{i|j}$ represents the uncertainty in $\vec{S}_i^{t,c}$ after observing $\vec{S}_j^{t,c}$. Here $p_i(k)$ is the probability of each feature in $\vec{S}_i^{t,c}$, $p_{i,j}(k)$ is the joint probability of $\vec{S}_i^{t,c}$ and $\vec{S}_j^{t,c}$ and $p_{i|j}(k)$ is the transition probability from $\vec{S}_i^{t,c}$ to $\vec{S}_j^{t,c}$.

3. For a specific cognitive task K_c, the similarity between the M source signals represented by the selected D features should be high. Contrarily, for better separability between two cognitive tasks, the difference between their respective source feature vectors should be maximized.

This objective is concerned with the optimal selection of D features out of F features of the source signals corresponding to the selected M ($\leq N$) sink signals. This aims at discarding the redundant information of the selected source signals for classifying two specific cognitive tasks. Let L be the number of data points (source EEG feature vectors) of any cognitive task.

This is accomplished here in two phases. First, it tries to minimize the difference between the selected D ($\leq F$) features of the source signals (corresponding to the selected $M \leq N$ sink signals), for all data points within a specific class K_c. The design philosophy is that if the k-th feature is selected as a unique representative feature of the class K_c, then it should

provide a high degree of similarity between the *k*-th feature of the *i*-th source for any two data points within the same class K_c. This is done for all classes c=[1, C]. This is realized here by minimizing

$$J_W = \sum_{c=1}^{C} \sum_{t=1}^{L-1} \sum_{u=t+1}^{L} \sum_{i=1}^{M} \sum_{k=1}^{D} \left(R_{i,k}^{t,c} - R_{i,k}^{u,c} \right)^2. \tag{3.11}$$

The second part aims at maximizing the ratio of the mean to the standard deviation between the selected D ($\leq F$) features of the source signals (corresponding to the selected $M \leq N$ sink signals), between any two classes. This ensures that the selected feature is capable enough to discriminate between any two classes. This is here realized by maximizing

$$J_B = \sum_{c=1}^{C} \sum_{\substack{d=1, \\ d \neq c}}^{C} \sum_{k=1}^{D} \sum_{i=1}^{M} \left(\bar{R}_{i,k}^{c} / \sigma_{i,k}^{c} - \bar{R}_{i,k}^{d} / \sigma_{i,k}^{d} \right) \tag{3.12}$$

where

$$\bar{R}_{i,k}^{c} = \frac{1}{L} \sum_{t=1}^{L} R_{i,k}^{t,c} \tag{3.13}$$

$$\sigma_{i,k}^{c} = \sqrt{\frac{(R_{i,k}^{t,c} - \bar{R}_{i,k}^{c})^2}{L}} \tag{3.14}$$

represent the mean and the standard deviation of the *k*-th feature of the *i*-th electrode over L data points in a specific class K_c for k=[1, D], i=[1, M] and K_c=[1, C].

With these three considerations, a composite objective function is formulated in (3.15), maximization of which yields the optimal sets of electrodes and features for classification of C cognitive tasks.

$$J = \sum_{c=1}^{C} \sum_{t=1}^{L} \sum_{i=1}^{N} \sum_{\substack{j=1, \\ j \neq i}}^{M} \phi_{i,j}^{t,c} - \lambda_1 \sum_{c=1}^{C} \sum_{t=1}^{L} \sum_{i=1}^{M} \sum_{\substack{j=1, \\ j \neq i}}^{M} M_{i,j}^{t,c} + \lambda_2 \frac{J_B}{J_W} \tag{3.15}$$

λ_1 and λ_2 are constants which are set in a manner so as to have all terms in the right hand side of (3.15) in the same order of magnitude. In our experiment, λ_1 and λ_2 are respectively set as 10 and 20.

Fig. 3.3 pictorially represents a four dimensional candidate solution $\vec{Y}$ for the present optimization problem with N=5 electrodes and F=4 features. Y_1 and Y_3 denote the values of M ($\leq N$) and D ($\leq F$). Y_2 and Y_4 are decimal values within [1, 2^N-1] and [1, 2^F-1] respectively. The binary decoding of Y_2 and Y_4 are used to identify M electrodes and D features. For

example, let $\vec{Z}_2$ (or $\vec{Z}_4$) be *N* (or *F*) dimensional binary string, obtained by decoding Y_2 (or Y_4). Intuitively, $Z_{2,j}$= 1 ($Z_{4,j}$= 1) indicates that the *j*-th electrode (or feature) selected or $j \in [1, N]$ (or $j \in [1, F]$). It is noteworthy that the number of ones in $\vec{Z}_2$ (or $\vec{Z}_4$) should be equal to *M* (or *D*).

3.2.4 Classification

Classification [18] is considered as an important step in EEG based research problems. A classifier primarily segregates unknown data samples into considered class labels after being trained with similar features. In context of BCI research, a classifier aims to distinguish between different brain activities after being optimally trained by following any 'learning' algorithm. While classifying EEG signals obtained from the electrodes placed on the scalp, every researcher strives to achieve high accuracy. There are numerous kinds of classifiers available which lets a user to choose the most suitable classifier according to the requirement of the problem. In the present chapter, we have used non-linear *support vector machine* (SVM) classifier with suitable kernel function for each of the tasks conducted. Three kernel functions, including

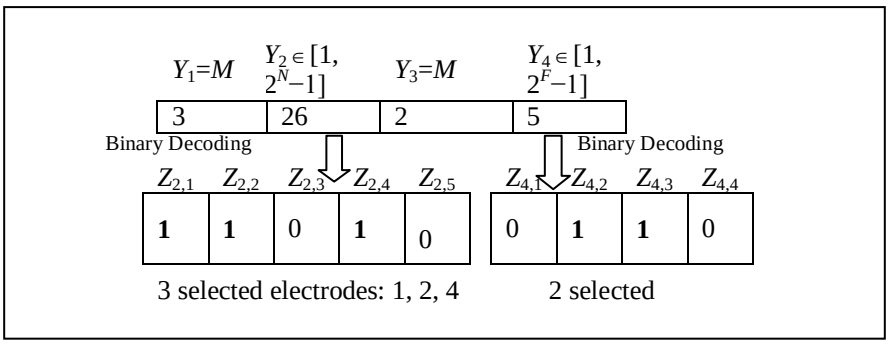


Fig. 3.3 Illustration of Encoding a Candidate Solution for *N*=5 and *F*=4

- Gaussian radial basis function:

$$k(X_1, X_2) = \exp(-\lambda \|X_1 - X_2\|^2) \text{ for } \lambda > 0$$

- Homogeneous polynomial function:

$k(X_1, X_2) = (X_1.X_2)l$ where l denotes the number of polynomials

- Hyperbolic tangent function:

$$k(X_1, X_2) = \tanh(kX_1.X_2 + c) \in k < 0 \text{ and } c > 0 .$$

To determine the efficacy of the proposed methodology, the SVM classifier is trained with the selected features of the selected electrodes (and corresponding sources). The testing data set is also prepared with the selected feature-based information contents of the selected electrodes and the sources.

3.3 SELF-ADAPTIVE FIREFLY ALGORITHM

The optimization problem of optimal selection of electrode positions and EEG features for classification of cognitive tasks here has solved using a self-adaptive variant of the traditional firefly algorithm (FA) [19]. In this section, first an overview of the traditional FA is provided. Next, we propose the self-adaptive variant of FA, referred to as SAFA [20], which adaptively tunes its control parameters to balance the trade-off between the computational accuracy and the run-time complexity effectively.

3.3.1 Firefly Algorithm

In firefly algorithm (FA) [9], a possible solution of an optimization problem is encoded by the position of a firefly in the parameter space and its light intensity signifies the fitness of the respective solution. An overview of FA is given next.

1. Initialization: Initially, a population **P**(t) of NP, D-dimensional firefly positions, $\vec{Y}_i(t) = \{y_{i,1}(t), y_{i,2}(t), \ldots, y_{i,D}(t)\}$ for i= [1, NP] is uniformly randomized in the search range $[\vec{Y}^{\min}, \vec{Y}^{\max}]$ where $\vec{Y}^{\min} = \{y_1^{\min}, y_2^{\min}, \ldots, y_D^{\min}\}$ and $\vec{Y}^{\max} = \{y_1^{\max}, y_2^{\max}, \ldots, y_D^{\max}\}$ at the current generation t = 0, given by

$$y_{i,d}(0) = y_d^{\min} + \text{rand}(0,1) \times (y_d^{\max} - y_d^{\min}) \tag{3.16}$$

for j=[1, D] and i=[1, NP] where rand(0, 1) is a uniformly distributed random number in [0, 1]. The objective function value $f(\vec{Y}_i(0))$ of $\vec{Y}_i(0)$ is evaluated for i= [1, NP].

2. Attraction to Brighter Fireflies: Now the firefly $\vec{Y}_i(t)$ is attracted towards the positions of the brighter fireflies $\vec{Y}_j(t)$ for i, j= [1, NP] but $i \neq j$ such that $f(\vec{Y}_j(t)) > f(\vec{Y}_i(t))$ (for maximization problem). Apparently, the attractiveness $\beta_{i,j}$ of $\vec{Y}_i(t)$ towards $\vec{Y}_j(t)$ decreases exponentially with the distance between them, denoted by $d_{i,j}$ as given in (3.17).

$$\beta_{i,j} = \beta_o \exp(-\gamma \times d_{i,j}^m), \quad m \geq 1 \tag{3.17}$$

where β_0 represents the maximum attractiveness felt by $\vec{Y}_i(t)$ at its own position (i.e., at $d_{i,j} = d_{i,i} = 0$) and γ denotes the light absorption coefficient, which controls the rate of change of $\beta_{i,j}$ with $d_{i,j}$. Intuitively, γ governs the convergence speed of FA [9]. A setting of γ=0 is concerned with a constant attractiveness of β_0 for all fireflies, while γ approaching infinity implies complete random search [9]. The possible range of γ is found to be [0.01, 10] in the existing literature. In (3.17), m is a pre-defined positive non-linear modulation index. The distance between $\vec{Y}_i(t)$ and $\vec{Y}_j(t)$ is computed using the Euclidean norm as follows.

$$d_{i,j} = \| \vec{Y}_i(t) - \vec{Y}_j(t) \| \tag{3.18}$$

This step is repeated for $i, j = [1, NP]$.

3. Movement of Fireflies: The firefly at position $\vec{Y}_i(t)$ flies towards a more attractive location $\vec{Y}_j(t)$ (in the parameter space) of a brighter firefly (i.e., $f(\vec{Y}_j(t)) > f(\vec{Y}_i(t))$) for $j = [1, NP]$ but $i \neq j$ following

$$\begin{aligned} \vec{Y}_i^{next}(t) &= \vec{Y}_i^{cur}(t) + \beta_{i,j} \times (\vec{Y}_j(t) - \vec{Y}_i(t)) + \alpha \times (\vec{r} - 0.5) \\ \vec{Y}_i^{cur}(t) &\leftarrow \vec{Y}_i^{next}(t) \end{aligned} \tag{3.19}$$

where $\vec{Y}_i^{cur}(t)$ is initialized with $\vec{Y}_i(t)$ before its movement and $\vec{r}$ denotes a D-dimensional random position vector with its d-th component uniformly distributed in [0, 1] for $d = [1, D]$. The movement of the i-th firefly, governed by (3.19), is carried on for $j = [1, NP]$, but $i \neq j$ such that $f(\vec{Y}_j(t)) > f(\vec{Y}_i(t))$. This step is repeated for $i = [1, NP]$. The first term in (3.19) represents the firefly's position after its last movement. The second term in (3.19) denotes the positional change of $\vec{Y}_i(t)$ due to the attraction towards $\vec{Y}_j(t)$. Apparently, this term has no contribution towards controlling the movement of the brightest firefly and hence, it may be stuck at the local optima in the parameter space. This problem is overcome by inducing a random movement of the fireflies with a step-size of $\alpha \in (0, 1)$. The upgraded position of the i-th firefly, after being controlled by the brighter ones, is represented by $\vec{Y}_i(t+1)$ for $i = [1, NP]$.

4. Convergence: After each evolution, the steps 2 and 3 are repeated until one of the terminating conditions is satisfied. The conditions include restricting the number of maximum generations, preserving error limits, or the both, whichever occurs earlier.

3.3.2 *Self-Adaptive Firefly Algorithm* (*SAFA*)

The population members of the traditional FA are equipped with the exploitation capability capable to escape the local optima due to their random movements with step-size α in (3.19). Intuitively, the step-size (α) profile governs the convergence of fireflies towards global optimum. However, in traditional FA, a constant value of α is used for the random movement, irrespective of the position of the fireflies in the parameter space. A setting of large value of α may result in the deviation of a quality firefly from the global optimum, while a small value of α may take a relatively long time to effectively orient a poor firefly towards the global optimum. The exploitation capability of the traditional FA, being a decisive factor of its performance, here has been farther improved by self-adapting the step-size parameter α within a range [α^{min}, 1) based on the relative position of a firefly with respect to the current best firefly position. This is realized by setting

$$\alpha_{i,d} = \alpha^{\min} + (1-\alpha^{\min}) \times \text{rand}(0,1) \times \frac{|\, y_d^{best}(t) - y_{i,d}(t) \,|}{y_d^{\max} - y_d^{\min}} \tag{3.20}$$

for d=[1, D]. The step-size is now treated as a D-dimensional vector as symbolized by $\vec{\alpha}_i$ ={$\alpha_{i,1}(t)$, $\alpha_{i,2}(t)$, …, $\alpha_{i,D}(t)$} with its d-th component $\alpha_{i,d} \in [\alpha^{\min}, 1)$ for d=[1, D]. Here $\vec{Y}^{best}(t)$ ={$y_1^{\text{best}}(t)$, $y_2^{\text{best}}(t)$, …, $y_D^{\text{best}}(t)$} the best position of the firefly in the t-th generation and |.| represents the absolute value. The dynamic in (3.20) guarantees that a firefly at $\vec{Y}_i(t)$, close to $\vec{Y}^{best}(t)$, should exploit the local neighborhood with a small step-size $\alpha_{i,d} \approx \alpha^{\min}$ to prevent the exclusion of the global optimum. Contrarily, an inferior firefly, far away from $\vec{Y}^{best}(t)$, should take part in the global search (with step size $\alpha_{i,d}$ approaching unity for d=[1, D]) to explore the potential zones in the parameter space.

Moreover, a simple strategy is proposed to update the values of γ in each generation, based on the knowledge of its potential values that were able to generate better firefly positions in the last generation. At every generation, the light absorption coefficient $\gamma_i(t)$ of each individual firefly $\vec{Y}_i(t)$ is independently generated as

$$\gamma_i(t) = Rayleigh(\sqrt{2/\pi}\,\bar{\gamma}(t)) \tag{3.21}$$

where $Rayleigh(\sqrt{2/\pi}\,\bar{\gamma}(t))$ is a random number sampled from a Rayleigh distribution with mean $\bar{\gamma}(t)$ (scale parameter $\sqrt{2/\pi}\,\bar{\gamma}(t)$). The value of $\gamma_i(t)$ is reproduced if it is beyond its allowable range [0.01, 10].

Let $\gamma_s(t)$ be the set of the successful light absorption coefficients of all fireflies of the current generation t producing better positions for the next generation t+1.

$$\gamma_s(t) = \{\gamma_i(t) \text{ for } i = [1, NP] : f(\vec{Y}_i(t+1)) > f(\vec{Y}_i(t))\} \tag{3.22}$$

Here, $\bar{\gamma}(0)$ is initialized to be 0.5. After every generation, it is updated as

$$\bar{\gamma}(t+1) = w \times \bar{\gamma}(t) + (1-w) \times \mu_p(\gamma_s(t)) \tag{3.23}$$

where $\mu_{pow}(.)$ denotes the power mean [21], given by

$$\mu_p(\gamma_s(t)) = \sum_{\gamma \in \gamma_s(t)} [\gamma^n / |\gamma_s(t)|]^{1/n} \tag{3.24}$$

with $|S|$ denotes the cardinality of set S. The weight factor w in (3.23) is randomly selected from [0.7, 1]. The weighted sum of $\bar{\gamma}(t)$ and $\mu_p(\gamma_s(t))$ helps in the effective tuning of $\bar{\gamma}(t+1)$ based on the successful values of the light absorption coefficients in the past and present generations respectively. The positive constant n in (3.24) is taken as 1.5 after wide variety of experiments to avoid premature convergence at local optima. The design philosophy adopted here relies on the principle of selecting large diversified values of γ from the Rayleigh distribution (having longer tails than the normal distribution) when the population is far away from the global optimum.

3.4 EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

This section describes each step of the proposed framework with the obtained readings arranged in a tabulated manner. Apart from that, the performance of the proposed system has been analyzed with respect to certain standard methods in terms of different performance indices.

3.4.1 EEG Signal Acquisition

Fifteen different binary classification experiments are carried out for this particular framework, in order to validate the obtained results for any kind of cognitive task without loss of generality. In each case, ten healthy subjects aged between 22 and 30 years have participated (five male and five female) in the experiments. The experiments are undertaken in multiple sessions, each of 2 minutes duration and every subject is asked to perform each

session 10 times. It is thus evident from Fig. 3.2 that the number of data points L in each cognitive task equals to (5+5)×10=100. All the signals are recorded using a stand-alone EEG machine manufactured by Nihon Kohden (200 Hz sampling frequency) comprising N=19 electrodes, placed using the standard international 10-20 electrode placement method [22], as shown in Fig.3.4. A_1 and A_2 are considered as the reference electrodes.

TABLE 3.1 ALGORITHM FOR SAFA INDUCED FEATURE SELECTION

Procedure SAFA_Induced_Electrode_Feature_Selection

Input: L data points, each comprising the EEG signals of N electrodes for each cognitive task class K_c for c=[1, C], F EEG features, non-linear modulation index m.
Output: Optimally selected M ($\leq N$) electrodes and D ($\leq F$) features.

Begin

1. Initialize a population **P**(t) of NP, 4-dimensional firefly position vectors $\vec{Y}_i(t)$ at generation t=0 using (3.16) and Fig. 3.3 for i= [1, NP].
2. Set $\bar{\gamma}(0) \leftarrow 0.5$.
3. Decode $\vec{Y}_i(0)$ using Fig. 3.3 and evaluate $f(\vec{Y}_i(0))$ using (3.15) for i= [1, NP].
4. Set $\vec{Y}^{best}(0) \leftarrow \arg\left(\max_{i=1}^{NP} f(\vec{Y}_i(0)) \right)$.
5. **While** termination condition is not reached **do**
   Begin
   I. Set $\gamma_s(t) \leftarrow NULL$.
   II. **For** i=1 to NP **do**
      Begin
      (i) Select $\gamma_i(t)$ using (3.21).
      (ii) **For** j=1 to NP, $j \neq i$ **do**
         Begin
         If $f(\vec{Y}_j(t)) > f(\vec{Y}_i(t))$ **Then do**
            Determine $\vec{Y}_i^{next}(t)$ using (3.19) and (3.20).
         End If.
         End For.
      (iii) Set $\vec{Y}_i(t+1) \leftarrow \vec{Y}_i^{next}(t)$.
      (iv) Decode $\vec{Y}_i(t+1)$ using Fig. 3.3 and evaluate $f(\vec{Y}_i(t+1))$ using (3.15) for i= [1, NP].
      (v) **If** $f(\vec{Y}_i(t+1)) > f(\vec{Y}_i(t))$ **Then**
         Set $\gamma_s(t) \leftarrow \gamma_s(t) \cup \gamma_i(t)$.
         End If.
      End For.
   III. Set $\vec{Y}^{best}(t) \leftarrow \arg\left(\max_{i=1}^{NP} f(\vec{Y}_i(t)) \right)$.
   IV. Set $t \leftarrow t+1$.
   V. Determine $\bar{\gamma}(t+1)$ from (3.23).
   End While.
6. Decode $\vec{Y}^{best}(t)$ using Fig. 3.3
7. Return optimally selected M ($\leq N$) electrodes and D ($\leq F$) EEG features.

End.

TABLE 3.2: EXPERIMENTS CONDUCTED FOR PROPOSED SCHEME

Index	Description
Experiment 1	Left hand and right hand motor execution
Experiment 2	Left hand and right hand motor imagery
Experiment 3	Left leg and right leg motor execution
Experiment 4	Left leg and right leg motor imagery
Experiment 5	Tongue and finger motor execution
Experiment 6	Happiness and sadness emotion recognition
Experiment 7	Happiness and anger emotion recognition
Experiment 8	Happiness and fear emotion recognition
Experiment 9	Happiness and disgust emotion recognition
Experiment 10	Disgust and fear emotion recognition
Experiment 11	Disgust and anger emotion recognition
Experiment 12	Disgust and sadness emotion recognition
Experiment 13	Anger and fear emotion recognition
Experiment 14	Anger and sadness emotion recognition
Experiment 15	Sadness and fear emotion recognition

Table 3.2 presents an overview of the cognitive tasks undertaken during different experiments for the concerned problem. From Table-3.2, it can be seen that the experiments are not limited to only motor execution or motor intention based tasks [23], but the list also includes several experiments of segregation of five basic emotions, including happiness, sadness, anger, fear and disgust. To conduct experiments related to emotion recognition, certain video clips are shown to the users with sufficient amount of relaxation interval in between two clips corresponding to two different emotions. The clips of small duration are chosen such that it takes minimal time for the user to understand the clip and to experience such emotions freely.

While performing the experiments the subjects are asked to sit on a chair with arms at rest position and eyes stable. After a certain interval, a cross appears on the screen along with a beep sound; thereafter the subjects are instructed to perform the required tasks according to commands appearing on screen. The structure of a stimulus used for left and right hand motor execution task has been shown in Fig. 3.5 as an example.

3.4.2 Preprocessing

The acquired EEG signals are passed through certain pre-processing steps in order to remove the artifacts generated due to eye blinking or spurious pick-ups from the power supply. The responses to the experiments undertaken are better captured by the rhythmic brain activity in the frequency band of 4-40Hz. Hence, to filter the signals within the specified frequency band, a 6-th order elliptical filter is used with 1dB pass-band ripple and 50dB stop-band ripple.

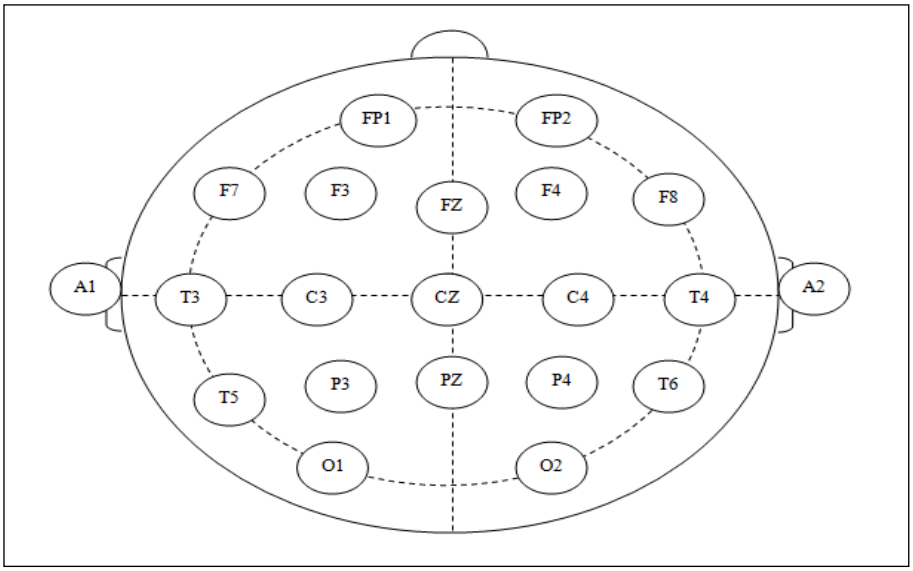


Fig. 3.4 Electrode positions according to 10-20 electrode placement system

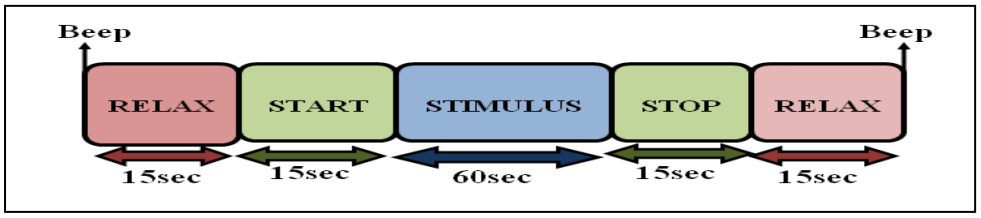


Fig. 3.5 Queue for stimulus presentation for left and right hand motor execution

3.4.3 Feature Extraction

In this step, five well known features for EEG based research have been considered including temporal features like Hjorth parameters [3], AAR parameters [3], frequency domain features like PSD [3], time-frequency correlated features like discrete wavelet coefficient [3] and CSP features [2]. In this case, AAR parameters of 6-th order have been calculated using Kalman filter as the estimator with an update coefficient of 0.0085. PSD features have been extracted

using Welch method with 50% overlapped signal segments using a Hamming window. For wavelet coefficients, Daubechies 4-th order mother wavelet has been used.

3.4.4 Optimal Feature and Electrode Selection using Optimization Algorithm

A potential solution of this optimization problem of selection of the optimal set of EEG electrodes [24] and EEG features is encoded as a four dimensional firefly position (Fig. 3.3). Using a set of such firefly positions, the proposed SAFA is executed. In each generation of SAFA, three steps are followed to determine the objective function value of a firefly position. First, each firefly position is decoded using Fig. 3.3 to obtain M ($\leq N$) and D ($\leq F$). The source signals, corresponding to the selected M ($\leq N$) electrodes, are estimated using ICA. The selected D ($\leq F$) features are then extracted from the selected M ($\leq N$) sources and sinks. The fitness of the solution is then determined using (3.15). Maximization of (3.15) using FASA ultimately provides the optimal set of electrodes and EEG features for the effective representation of C cognitive tasks i) without dropping out any useful information and ii) without any redundant information.

3.4.5 Experimental Results

Table 3.3 reports the mean (standard deviation within parenthesis) of the following measures as obtained by the proposed SAFA-based realization over 50 independent runs, each with 4×10^4 maximum number of function evaluations. The performance metrics include

- The average of the distance correlation coefficients between the selected D features of the selected M and the original N set of sources,
- The average of the mutual information between the selected D features of the selected M and the original N set of sinks, and
- The average value of J_B/J_W due to the selected M sources, each with D features over 50 runs and for all possible combinations of fifteen cognitive tasks. It is evident from Table-3.3 that for emotion-based experiments, frontal and prefrontal electrodes have been chosen as the optimal ones apart from a few other channels. It justifies the effectiveness of the proposed scheme as it is known from elementary neuro-physiological knowledge that frontal and prefrontal brain area are majorly responsible for emotion based task executions.

A comparative analysis of the proposed SAFA with other existing algorithms, including self-adaptive artificial bee colony (SAABC) [25], traditional ABC [26] and traditional differential evolution (DE) [27], are undertaken in Table 3.4 with respect to the mean (standard

deviation within parenthesis) objective function values of the optimal solutions (using (3.15)) over 50 independent runs. All the algorithms commence from the same initial population of size 50. The maximum number of function evaluations for each run of an algorithm is set equal to 50×10^4. Their control parameters are set in a manner to have their individual best performance in the present context after a wide variety of experiments. The parameters of SAFA including α^{min}, β_0, and m are respectively set as 0.2, 1 and 1.5. The *limit* cycle for SAABC and ABC is set to 50. A crossover rate of 0.9 is used in case of DE/current-to-best/1.

The statistical significance level of the difference of the 50 samples of the optimal objective function of any two competitive algorithms is verified by the Wilcoxon rank sum test [28] with a significance level α=0.05 [29]. The *p*-values obtained through the rank sum test between the best algorithm and each of the remaining algorithms over all combinations of cognitive tasks is reported in third brackets in Table3.4. Here NA stands for *not applicable* covering the cases of comparing the best algorithm with itself. The null hypothesis of this statistical test is concerned with the equivalent performance of all the competitor algorithms. If the *p*-value, corresponding to the relative performance analysis of the *i*-th and *j*-th algorithms, is less than α, then the respective null hypothesis is rejected. The reported results in Table 3.4 clearly indicate the superiority of the proposed SAFA to its contenders in a statistically significant fashion over most of the combinations of cognitive tasks. SAABC, which yields the best objective function values for three cases (including experiments 7, 11 and 14) attain the second best rank. However, for experiment 11, the statistical test indicates an insignificant dominance of SAABC over SAFA. It is noteworthy that DE based realization of the problem outperforms the proposed SAFA for task 9, however, insignificantly.

It is also remarkable from Tables 3.3 and 3.4 that the performance of each algorithm remains better for selection of optimal electrodes and EEG features for different classes of motor intensions or motor imagery rather than emotion recognition. An obvious reason may be that the emotional stimuli produce the brain rhythms essentially in α and θ bands, but here the signals are band-pass filtered as a whole in the 4-40 Hz band. It may have resulted in a degraded performance of an algorithm as compared to other cognitive tasks.

The comparative analysis of the competitor algorithms is also undertaken with respect to the classification accuracy of three different SVM classifiers [30], considered in this chapter. This is accomplished by first creating a testing dataset for each task combinations in Table-3.2. The testing dataset is created by following the same principle as in section 3.4.1. Then for an algorithm, say *i*, we obtain optimal set of *M* electrodes and *D* features. Then these

TABLE 3.3.A: PERFORMANCE OF SAFA BASED SELECTION OF OPTIMAL EEG ELECTRODES AND FEATURES FOR COGNITIVE TASK CLASSIFICATION FOR EXPERIMENTS 1 TO 14

Experiment Index	Number of Selected Electrodes	Optimal Electrode Positions	Number of Selected Features	Correlation	J_B/J_W	Mutual Information
1	5	C3, C4, Cz, P4, P3	545	320.51 (0.047)	11.09 (0.068)	12.66 (0.043)
2	4	C3, C4, Cz, Pz	412	288.75 (0.013)	9.43 (0.027)	10.06 (0.031)
3	6	C3, C4, Cz, P3, P4, Pz	755	373.89 (0.057)	18.23 (0.024)	20.65 (0.075)
4	3	P3, P4, Pz	323	180.76 (0.062)	8.39 (0.061)	8.85 (0.028)
5	5	C3, C4, P3, P4, Pz	563	331.87 (0.023)	16.51 (0.053)	16.78 (0.011)
6	4	FP1, FP2, Cz, F4	472	270.57 (0.039)	12.64 (0.038)	11.57 (0.027)
7	4	FP1, C3, Fz, F3	397	225.64 (0.032)	8.57 (0.014)	9.68 (0.017)
8	5	FP1, FP2, Fz, F3, C3	513	321.65 (0.076)	11.67 (0.032)	10.89 (0.072)
9	6	FP1, F3, F4, O1, C3, Cz	781	365.45 (0.083)	17.85 (0.011)	18.95 (0.069)
10	4	T3, FP1, F3, F4	456	265.88 (0.028)	8.58 (0.045)	9.98 (0.083)
11	5	O_2, FP2, Fz, F3, F4	529	318.77 (0.048)	11.96 (0.087)	13.46 (0.071)
12	3	FP1, FP2, Fz	293	200.45 (0.022)	7.43 (0.069)	8.93 (0.059)
13	5	T3, FP1, F3, F4, FP2	527	322.68 (0.059)	10.08 (0.052)	12.89 (0.081)
14	6	FP1, Pz, Fz, T3, F4, T4	746	354.05 (0.017)	18.79 (0.047)	19.58 (0.079)

TABLE 3.3.B: PERFORMANCE OF SAFA BASED SELECTION OF OPTIMAL EEG ELECTRODES AND FEATURES FOR COGNITIVE TASK CLASSIFICATION FOR EXPERIMENT 15

Experiment Index	Number of Selected Electrodes	Optimal Electrode Positions	Number of Selected Features	Correlation	JB /JW	Mutual Information
15	4	T3, FP1, FP2, F4	468	278.92 (0.086)	11.23 (0.033)	11.87 (0.066)

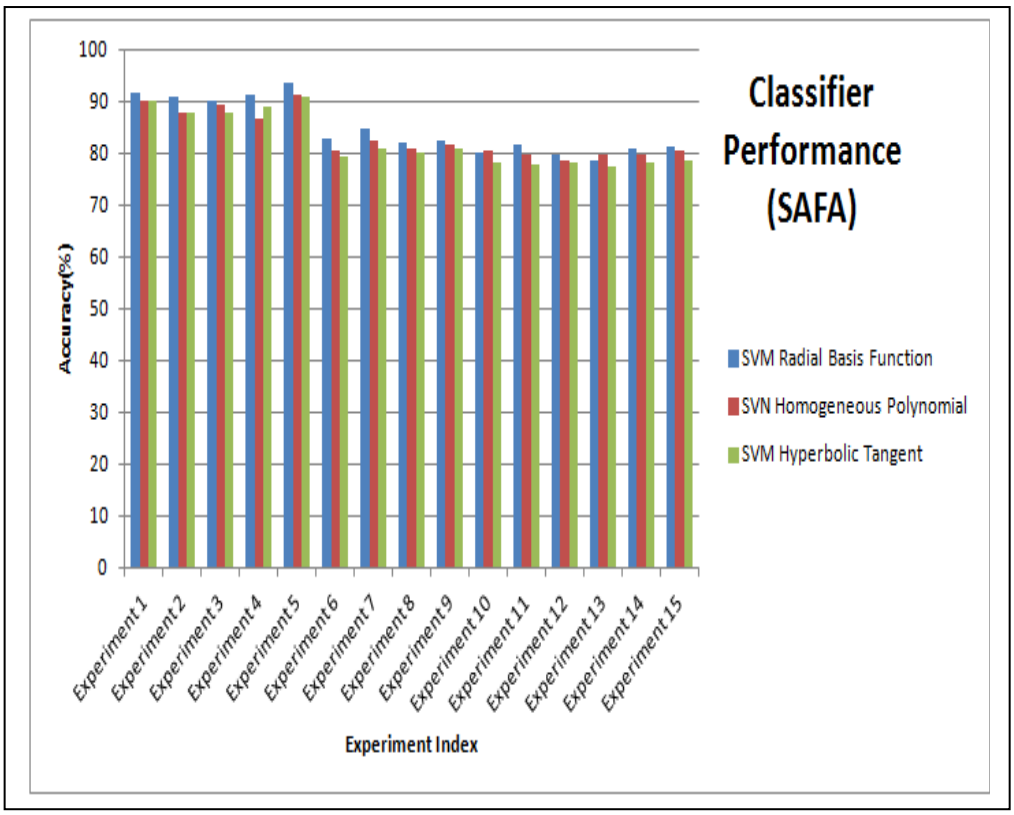
TABLE 3.4.A: COMPARISON OF THE PROPOSED ALGORITHM WITH OTHER STANDARD EVOLUTIONARY ALGORITHMS BASED ON OBJECTIVE FUNCTION VALUES FOR EXPERIMENTS 1 TO 7

Experiment Index	SAFA	SAABC	ABC	DE
1	0.2562 (0.056) [NA]	0.2068 (0.049) [0.0034]	0.0167 (0.011) [0.0053]	0.1583 (0.027) [0.0016]
2	0.2963 (0.095) [NA]	0.2477 (0.064) [0.0048]	0.0543 (0.031) [0.0074]	0.1948 (0.072) [0.0031]
3	0.3549 (0.025) [NA]	0.3249 (0.112) [0.0073]	0.0851 (0.105) [0.0084]	0.2057 (0.087) [0.0027]
4	**0.3688 (0.145) [NA]**	0.2911 (0.194) [0.0046]	0.1049 (0.162) [0.0033]	0.2648 (0.121) [0.0051]
5	**0.3344 (0.271) [NA]**	0.3105 (0.127) [0.0053]	0.1078 (0.186) [0.0082]	0.2420 (0.191) [0.0154]
6	**0.3231 (0.084) [NA]**	0.2865 (0.073) [0.0049]	0.0989 (0.085) [0.0064]	0.1997 (0.066) [0.0129]
7	0.3825 (0.094) [0.0044]	**0.3941 (0.105) [NA]**	0.2732 (0.147) [0.0018]	0.3475 (0.194) [0.0003]

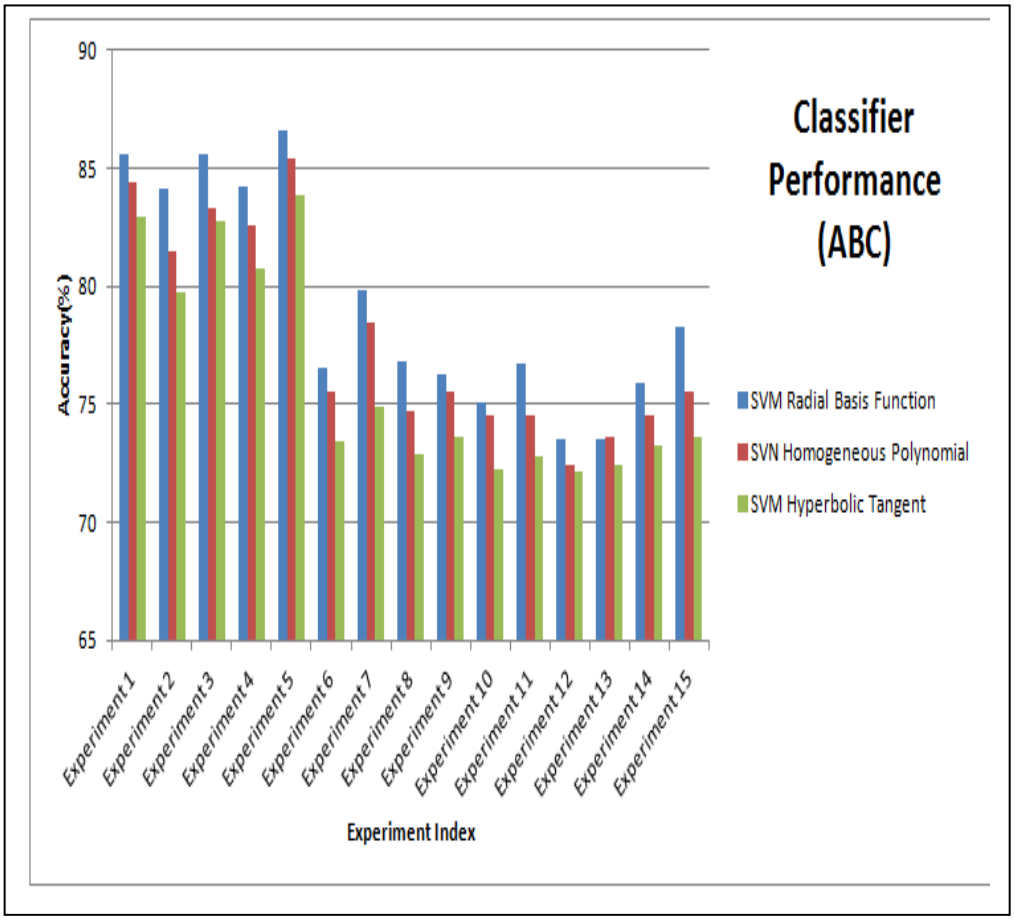
TABLE 3.4.B: COMPARISON OF THE PROPOSED ALGORITHM WITH OTHER STANDARD EVOLUTIONARY ALGORITHMS BASED ON OBJECTIVE FUNCTION VALUES FOR EXPERIMENTS 8 TO 15

Experiment Index	SAFA	SAABC	ABC	DE
8	**0.3376 (0.034) [NA]**	0.2951 (0.072) [0.0059]	0.0611 (0.128) [0.0011]	0.2559 (0.136) [0.0046]
9	0.2931 (0.241) [0.0564]	0.2883 (0.262) [0.0077]	0.2691 (0.177) [0.0003]	**0.3259 (0.167) [NA]**
10	**0.3596 (0.184) [NA]**	0.3115 (0.179) [0.0165]	0.1572 (0.086) [0.0234]	0.2754 (0.109) [0.0196]
11	0.3104 (0.097) [0.0602]	**0.3155 (0.116) [0.0007]**	0.2158 (0.123) [0.0097]	0.2795 (0.152) [0.0197]
12	**0.3963 (0.096) [NA]**	0.3386 (0.188) [0.0217]	0.1652 (0.192) [0.0194]	0.2964 (0.271) [0.089]
13	**0.3713 (0.290) [NA]**	0.3494 (0.306) [0.0204]	0.1922 (0.327) [0.0028]	0.2278 (0.315) [0.0011]
14	0.3353 (0.224) [0.0189]	**0.3686 (0.219) [0.0173]**	0.2263 (0.254) [0.0184]	0.3281 (0.278) [0.008]
15	**0.3533 (0.172) [NA]**	0.3467 (0.245) [0.0309]	0.0476 (0.239) [0.0061]	0.2343 (0.241) [0.0149]

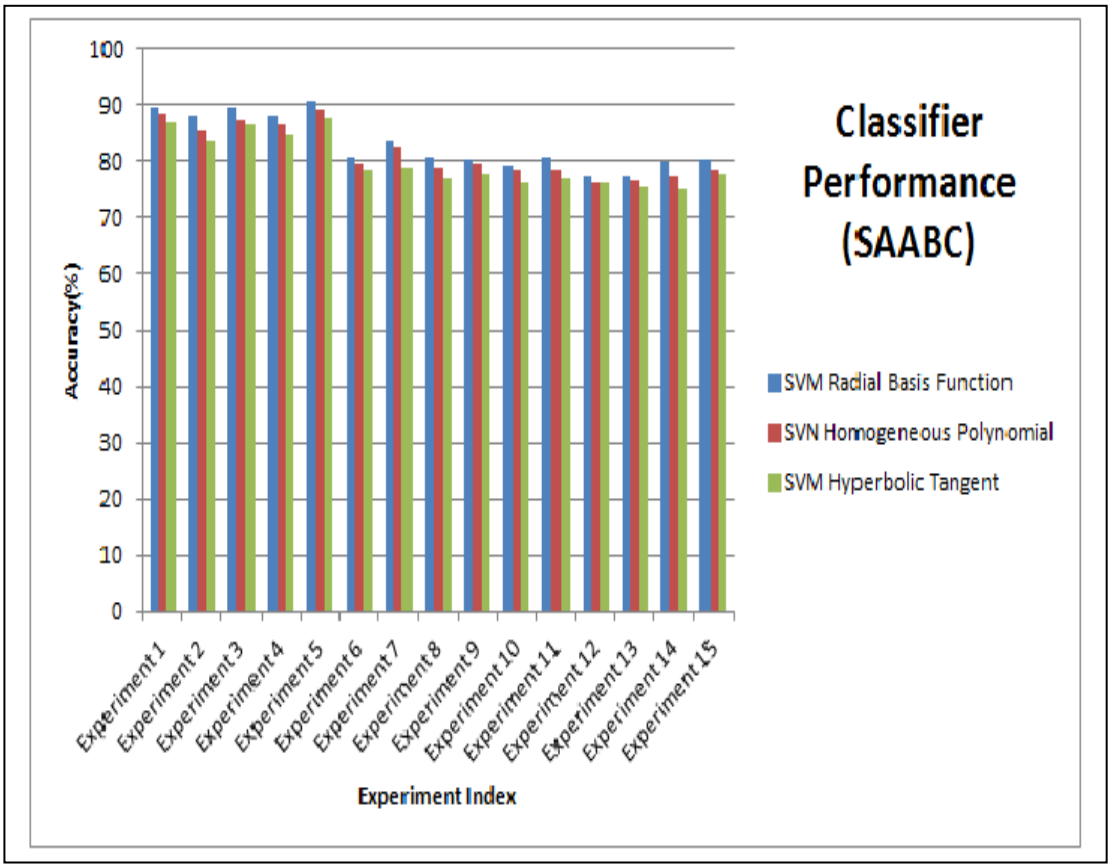
D-dimensional *M* source and sink feature vectors of the training dataset are used to train the SVM individually. After completing the training cycle, the *D*-dimensional *M* source and sink feature vectors are extracted from the testing set and are used to verify the classification accuracy of the trained SVM. This is repeated for all four competitors with i=[1, 4] for all three variants of nonlinear SVM. Fig. 3.6 plots the classification accuracy obtained by individual SVM classifier due to selection of optimal set of EEG electrodes and features by each contender algorithm. The plot clearly reveals that the proposed SAFA here too outperforms its contenders, however, marginally for SAABC [25].


Classifier Performance (SAFA)
Accuracy(%)
100
90
80
70
60
50
40
30
20
10
0
Experiment 1
Experiment 2
Experiment 3
Experiment 4
Experiment 5
Experiment 6
Experiment 7
Experiment 8
Experiment 9
Experiment 10
Experiment 11
Experiment 12
Experiment 13
Experiment 14
Experiment 15
Experiment Index
SVM Radial Basis Function
SVN Homogeneous Polynomial
SVM Hyperbolic Tangent

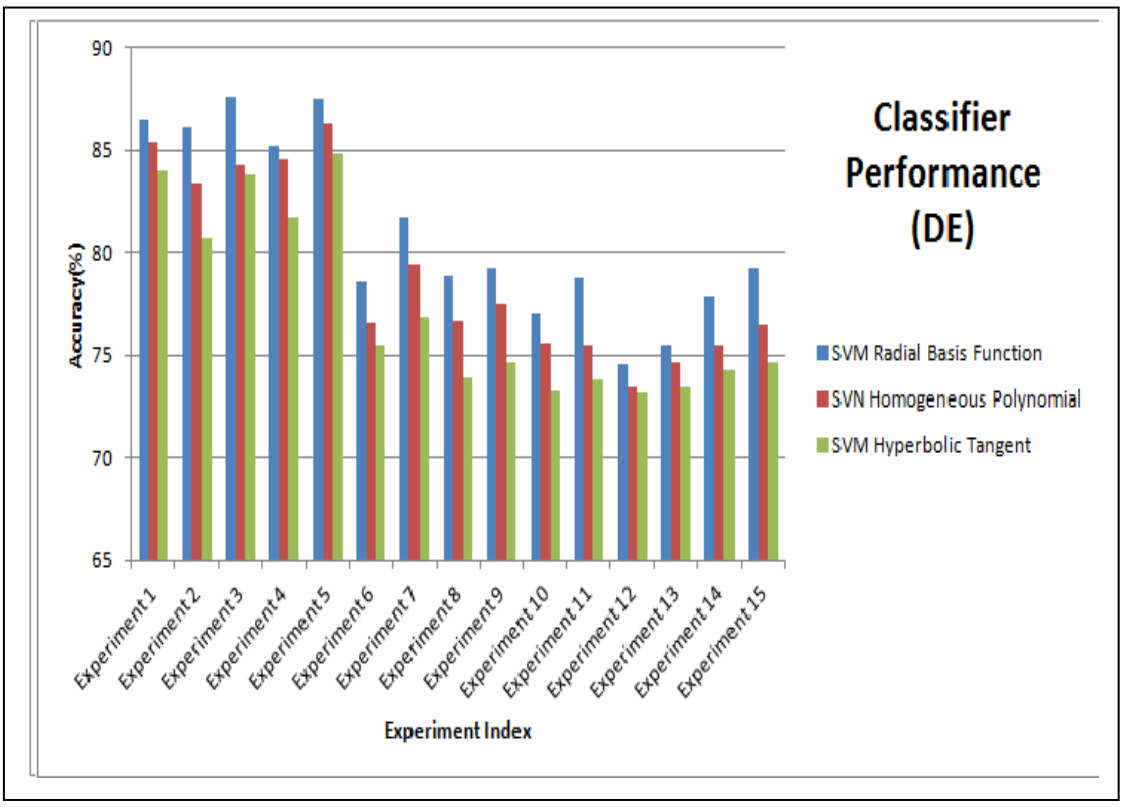
(a)


Classifier Performance (ABC)
Accuracy(%)
90
85
80
75
70
65
Experiment 1
Experiment 2
Experiment 3
Experiment 4
Experiment 5
Experiment 6
Experiment 7
Experiment 8
Experiment 9
Experiment 10
Experiment 11
Experiment 12
Experiment 13
Experiment 14
Experiment 15
Experiment Index
SVM Radial Basis Function
SVN Homogeneous Polynomial
SVM Hyperbolic Tangent

(b)



(c)



(d)

Fig. 3.6 Classification Accuracy of SVM Classifiers Due to Training with EEG Electrodes and Features Selected by (a) SAFA), (b) ABC, (c) SAABC and (d) DE

3.5 CONCLUSION

The chapter proposes a novel evolutionary optimization approach to the simultaneous selection of significant electrodes and relevant EEG features for classification of cognitive tasks in EEG based BCI paradigm. The present work can get extreme appreciation, because till date separate techniques of electrode and feature selections are adopted by the researchers, but a combined methodology of simultaneous dealing with these two important aspects of EEG signal processing has not been used in large scale. It is noteworthy that the efficiency of classification of cognitive tasks is enhanced by a great extent using the proposed framework. The main novelty of the work lies in the formulation of the problem in an optimization framework and in solving the problem using the proposed SAFA with an aim to satisfy three criteria. 1) To minimize the loss of information of the cortical sources for a cognitive task by discarding a specific set of redundant electrodes, the correlation between the EEG sources before and after the electrode selection should be as high as possible (in the feature space). 2) To ensure that only relevant EEG electrodes to be used for characterizing the mental states of cognitive tasks, the mutual information between the selected electrodes should be as low as possible indicating their independence (in the feature space). 3) The optimal set of features is selected by identifying the features that well differentiate between EEG sources of two cognitive tasks, while representing a specific cognitive task uniquely. To analyze the relative performance of SAFA with other existing evolutionary algorithms, three variants of non-linear SVM [4] classifier (including Gaussian radial basis function, homogeneous polynomial function and hyperbolic tangent function) are individually trained with the selected feature sets of the selected EEG electrodes and the corresponding EEG sources, as obtained by individual competitor algorithms. Experiments undertaken with SAFA reveal the statistically significant superiority of the proposed method over other existing evolutionary algorithms, with respect to classification accuracy of the SVM classifier, irrespective of the non-linearity. Despite the superiority of the above discussed methodology there exist ample scopes of improvement that can be carried out in future in order to reach closer to the desired objective. In the present scenario, the data obtained from the undertaken experiments have been exploited to compute merely five different types of features although the resulting features belong to very high dimension, hence the proposed work can be extended further in case of larger types of features and finally the classification accuracy obtained collectively after considering all the feature set is likely to provide relatively unbiased and fair conclusion. Moreover, although the list of experiments tabulated in Table 3.2 includes a wide variety of experiments, but it has majorly emphasized on emotion based experiments, so it is wise to

include other experiments that empowers any motor execution or motor imagery movements apart from the few basic ones mentioned here.

REFERENCES

[1] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. M??ller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[2] B. Graimann, B. Allison, and G. Pfurtscheller, "Brain – Computer Interfaces : A Gentle Introduction," pp. 1–28, 2010.

[3] E. C. Leuthardt, K. J. Miller, G. Schalk, R. P. N. Rao, and J. G. Ojemann, "Electrocorticography-based brain computer interface-the Seattle experience," *Neural Syst. Rehabil. Eng. IEEE Trans.*, vol. 14, no. 2, pp. 194–198, 2006.

[4] E. Niedermeyer and F. H. L. da Silva, *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[5] O. Hauk, D. G. Wakeman, and R. Henson, "Comparison of noise-normalized minimum norm estimates for MEG analysis using multiple resolution metrics," *Neuroimage*, vol. 54, no. 3, pp. 1966–1974, 2011.

[6] L. I. Kuncheva, J. J. Rodríguez, C. O. Plumpton, D. E. J. Linden, and S. J. Johnston, "Random subspace ensembles for fMRI classification," *Med. Imaging, IEEE Trans.*, vol. 29, no. 2, pp. 531–542, 2010.

[7] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2006, pp. 5392–5395.

[8] G. Vecchiato, L. Astolfi, F. De Vico Fallani, J. Toppi, F. Aloise, F. Bez, D. Wei, W. Kong, J. Dai, and F. Cincotti, "On the use of EEG or MEG brain imaging tools in neuromarketing research," *Comput. Intell. Neurosci.*, vol. 2011, p. 3, 2011.

[9] X.-S. Yang, "Firefly algorithms for multimodal optimization," in *Stochastic algorithms: foundations and applications*, Springer, 2009, pp. 169–178.

[10] I. Fister, X.-S. Yang, J. Brest, and I. Fister Jr, "Memetic self-adaptive firefly algorithm," *Swarm Intell. bio-inspired Comput. theory Appl.*, pp. 73–102, 2013.

[11] S. Datta, P. Rakshit, A. Konar, and A. K. Nagar, "Selecting the optimal EEG electrode positions for a cognitive task using an artificial bee colony with adaptive scale factor optimization algorithm," in *Evolutionary Computation (CEC), 2014 IEEE Congress on*, 2014, pp. 2748–2755.

[12] A. K. Mohamed, T. Marwala, and L. R. John, "Single-trial EEG discrimination between wrist and finger movement imagery and execution in a sensorimotor BCI," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 6289–6293.

[13] G. J. Székely, M. L. Rizzo, and N. K. Bakirov, "Measuring and testing dependence by correlation of distances," *Ann. Stat.*, vol. 35, no. 6, pp. 2769–2794, 2007.

[14] L. Zhukov, D. Weinstein, and C. Johnson, "Independent component analysis for EEG source localization," *Eng. Med. Biol. Mag. IEEE*, vol. 19, no. 3, pp. 87–96, 2000.

[15] B. Xu, A. Song, and J. Wu, "Algorithm of imagined left-right hand movement classification based on wavelet transform and AR parameter model," in *Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on*, 2007, pp. 539–542.

[16] S. Bhattacharyya, P. Rakshit, A. Konar, D. N. Tibarewala, and R. Janarthanan, "Feature selection of motor imagery EEG signals using firefly temporal difference Q-Learning and support vector machine," in *Swarm, Evolutionary, and Memetic Computing*, Springer, 2013, pp. 534–545.

[17] C. Guerrero-Mosquera, M. Verleysen, and A. N. Vazquez, "EEG feature selection using mutual information and support vector machine: A comparative analysis," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 2010, pp. 4946–4949.

[18] A. Khasnobish, S. Bhattacharyya, A. Konar, and D. N. Tibarewala, "K-Nearest neighbor classification of left-right limb movement using EEG data," in *International conference on Biomedical Engineering and assistive technologies, NIT Jalandhar*, 2010.

[19] C. Blum and X. Li, "Swarm Intelligence in Optimization," in *Swarm Intelligence: Introduction and Applications*, C. Blum and D. Merkley, Eds. Berlin: Springer Verlag, 2008, pp. 43–86.

[20] T. Niknam, R. Azizipanah-Abarghooee, and A. Roosta, "Reserve constrained dynamic economic dispatch: a new fast self-adaptive modified firefly algorithm," *Syst. Journal, IEEE*, vol. 6, no. 4, pp. 635–646, 2012.

[21] T.-P. Lin, "The power mean and the logarithmic mean," *Am. Math. Mon.*, vol. 81, no. 8, pp. 879–883, 1974.

[22] R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10–20 system electrode placement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 66, no. 4, pp. 376–382, 1987.

[23] C. Qiang, P. Hu, and F. Huanqing, "Experiment study of the relation between motion complexity and event-related desynchronization/synchronization," in *Neural Interface and Control, 2005. Proceedings. 2005 First International Conference on*, 2005, pp. 14–16.

[24] A. Saha, A. Konar, P. Rakshit, A. L. Ralescu, and A. K. Nagar, "Olfaction recognition by EEG analysis using differential evolution induced Hopfield neural net," in *Neural Networks (IJCNN), The 2013 International Joint Conference on*, 2013, pp. 1–8.

[25] W. Gu, M. Yin, and C. Wang, "Self adaptive artificial bee colony for global numerical optimization," *IERI Procedia*, vol. 1, pp. 59–65, 2012.

[26] M. Dorigo and G. . Caro, "The ant colony optimization meta-heuristic," in *New Ideas in Optimization*, D. Corne, M. Dorigo, and F. Glover, Eds. London,UK: McGraw Hill, 1999, pp. 11–32.

[27] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[28] D. J. Sheskin, *Handbook of parametric and nonparametric statistical procedures*. crc Press, 2003.

[29] J. H. Zar, *Biostatistical analysis*. Pearson Education India, 1999.

[30] A. Saha, A. Konar, P. Das, B. Sen Bhattacharya, and A. K. Nagar, "Data-point and feature selection of motor imagery EEG signals for neural classification of cognitive tasks in car-driving," in *Neural Networks (IJCNN), 2015 International Joint Conference on*, 2015, pp. 1–8.

Chapter 4

CSP & Its Regularization

This chapter primarily aims at describing the fundamental aspects of a well known feature extraction algorithm formally termed as Common Spatial Pattern (CSP).Firstly, it establishes the requirement of feature extraction in a rehabilitative BCI system. Then after briefly recapitulating the reasons behind the introduction of spatial filter, it describes the CSP methodology in details incorporating the future research directions regarding the improvement of CSP. As every entity has got its pros and cons, CSP too comes with its own set of limitations. To overcome these issues, several regularization techniques have been presented by different research enthusiasts throughout the world, thus researchers proposed a bunch of novel algorithms that enhances the existing framework in terms of most of the performance metrics, which are broadly classified to one category that is formally termed as Regularized CSP(RCSP).Section 4.1 presents a brief introduction and section 4.2 summarizes the existing works that are closely relevant to this chapter. Section 4.3 describes CSP algorithm from mathematical point of view and section 4.4 presents CSP methodology from optimization aspect. Section 4.5 describes the methods adopted for regularizing CSP and the existing algorithms and finally section 4.6 introduces the novel penalty terminologies as an extension of existing RCSP algorithms.

4.1 INTRODUCTION

A *Brain Computer Interfacing (BCI*) system [1] primarily attempts to translate a user's intentions into control commands without using the brain's normal output pathways of peripheral nerves and muscles. These days BCI is considered as one of the most developing research disciplines due to its growth, implementation of cutting edge and novel methodologies and scopes of improvement which bridges the theoretical gap between large numbers of research fields including medicines, psychology, signal processing, machine learning and so on. Since BCI interfaces mainly deal with brain rhythmic signals only, it can provide help to people suffering from several disabilities where the patients cannot instruct an assistive device with the help of muscular actions, but the cognitive components of the patients are not damaged and thus he/she can obtain the required service simply by using a BCI interface that is capable of decoding his/her thinking and executing the specific action. There are different modalities that are available through which brain signals can be acquired, including both invasive as well as non invasive implants. In this chapter we are considering one of the most common and user friendly non invasive modality termed as Electroencephalogram (EEG), which is reliable, easy to acquire and portable also.

EEG signals [2] are acquired by placing external electrodes on scalp following a standard electrode placement rules, but the EEG signals acquired from human scalp are highly contaminated by noise, moreover the oscillatory components of EEG signals are often found to be extremely non stationary because of the presence of sharp waves, spikes, electrical discharges etc. In fact, at times the amount of noise is so high that actual signal attributes are likely to be buried away in the noise. Hence, it is important to extract the useful components from the concerned signals for further analysis, because of these issues feature extraction is considered to be one of the key steps in an EEG based BCI system. Technically, a feature is a distinctive or characteristic measurement, a transform or a recognizable structural component that is extracted from a segment of a pattern. Usually, extracted features are meant to minimize the loss of the fundamental components that are embedded in a signal or a pattern, and to simplify the need of exploiting multiple information sources to describe larger amount of data with precision. In other words, the feature extraction phase primarily takes care of three important aspects,

- Reducing system implementation complexity
- Reducing cost of information processing
- Reducing the requirement of further information compression.

Literature show several feature extracting mechanisms have been adopted by the research fraternity in the past yielding moderately good experimental results including time

domain features [3] (AAR parameters, Hjorth parameters), frequency domain features (PSD, FFT) and time frequency correlated features (Wavelet coefficients). In case of basic BCI design, the number of electrodes considered and the number of subjects participating in the experiment is assumed to be relatively small, hence apparently the task looks relatively simple, but in case of larger data the complexity increases. Although from preliminary neuro physiological knowledge [4] we are aware of the particular brain regions that are likely to be responsible for a particular cognitive task, but there are high possibilities that even the surrounding brain regions may have a contribution in that particular task, however small it is. For example, for a motor execution task instead of C_3 and C_4 the optimal electrodes for a particular subject may be FC_3 and FC_4, or CP_3 and CP_4 etc. So, to obtain complete information for further analysis, it is recommended to use larger number of electrodes that covers maximum of the scalp regions so that none of the relevant information is missed out. Besides having a large number of electrodes, the optimal frequency band for different subjects may vary which may lead to a large amount of data and more importantly increasing the number of channels may incorporate redundancy also. In order to deal with such issues, the concept of spatial filtering has been introduced.

Spatial filtering basically derives a reduced number of new channels which are formed by linear combination of the original ones, following the equation (4.1),

$$x' = \sum_i w_i x_i = wX \tag{4.1}$$

Here w signifies the weight vector and X is the data obtained from the original electrodes placed upon human scalp. In this context, some of the basic spatial filters can be described in terms of the electrodes shown in Fig 4.1 [5].

The Fig. 4.1 depicts a set of 75 electrodes placed on different regions of human scalp following a standard electrode placement scheme including the reference electrodes. For a left hand and right hand motor imagery task, the electrodes that are considered for an experiment has been coloured red. In this chapter, only Bipolar and Laplacian filters has been described amongst the basic spatial filters and further the results obtained from them has been illustrated. Using the electrodes shown in Fig. 4.1 the Bipolar and Laplacian filter can be implemented as (4.2) and (4.3) respectively [5],

$$C_3^{'} = FC_3 - CP_3 \tag{4.2}$$

$$C_3^{'} = 4*C_3 - FC_3 - C_5 - C_1 - CP_3 \tag{4.3}$$

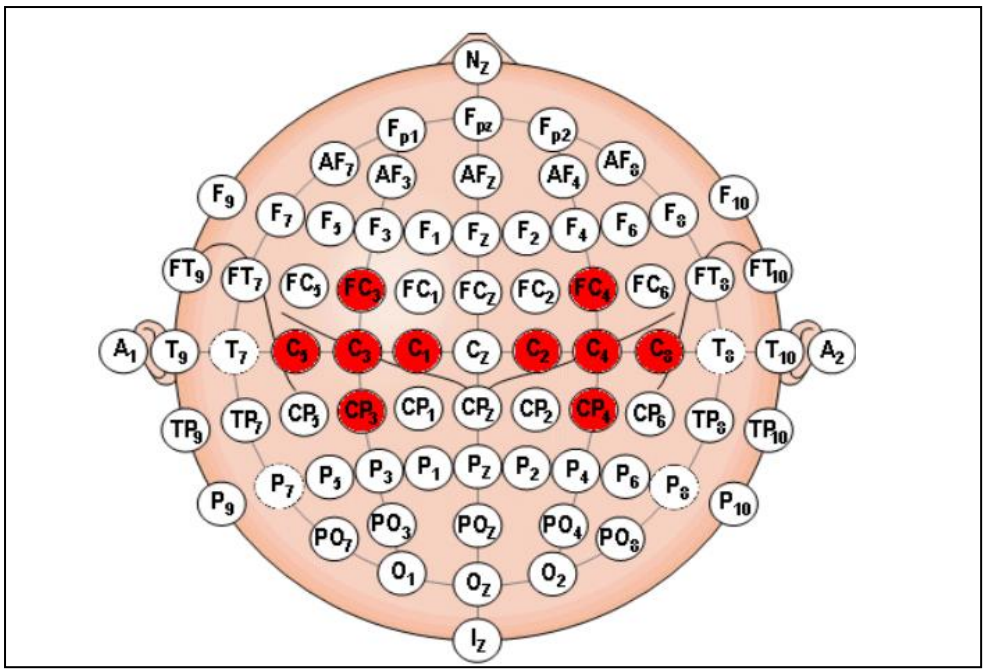


Fig 4.1 Electrodes Placed on Human Scalp Using Standard System

These spatial filters mainly emphasizes on localized activity, they help to extract the relevant information corresponding to any cognitive task. For example, due to volume conduction the information the signal originating from a specific source within the brain often gets scattered over more number of electrodes placed on scalp, in these cases spatial filtering can be employed to identify actual sources in brain localization problems.

Lotte *et al.* in his lecture [6] at BBCI Winter School of Neurotechnology has presented the performance of the above two basic spatial filters on BCI competition IV, dataset IIa using LDA classifier. This particular dataset comprises of EEG recordings from 9 different subjects performing left hand and right hand motor imagery tasks. After extracting the band power features the accuracy attained with Laplacian C_3-C_4 and Bipolar C_3-C_4 has been tabulated in Fig.4.2.

It is clearly revealed from Fig 4.2 that the accuracy obtained with the spatial filters clearly surpasses that obtained from the original C_3-C_4 electrodes, for all the subjects. The average classification accuracies for C_3-C_4, Laplacian C_3-C_4 and Bipolar C_3-C_4 are computed as 60.7%, 68% and 70.5% respectively. However, in this case, Bipolar C_3-C_4 has outperformed its other contender nearly in each case, but the results may not be the same for all cognitive task based experiments. Moreover, barring the performance attained for few electrodes the average accuracy obtained for all the three methods have not been satisfactory, as in most of the cases the accuracy obtained has been under the minimum threshold value of 70% which will not allow these filters to be implemented in any real time online systems. Hence, it is required to move on to more advanced variants of spatial filtering and thus

Common Spatial Pattern has been introduced as a member of the supervised spatial filtering family to deal with the limitations of its predecessors.

In this chapter, we are focussing upon CSP as a feature extractor algorithm. From the pattern recognition point of view, an EEG based BCI system comprises of the following phases as shown in Fig. 4.3 [7].

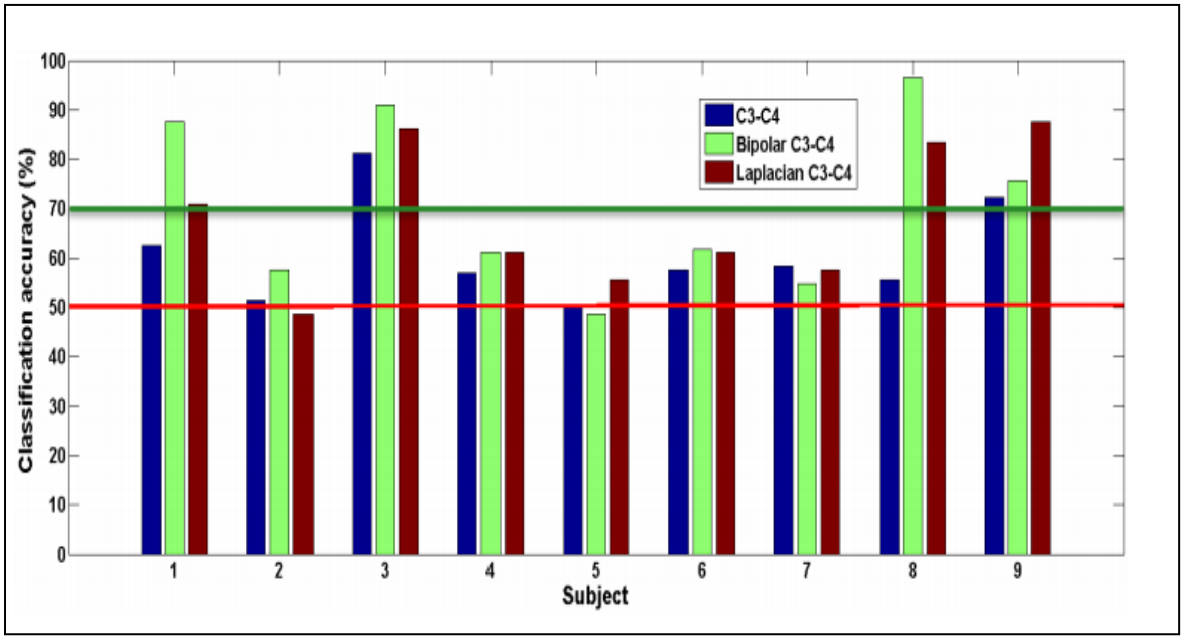


Fig. 4.2 Classification Accuracy Obtained with LDA Classifier for Basic Spatial Filters

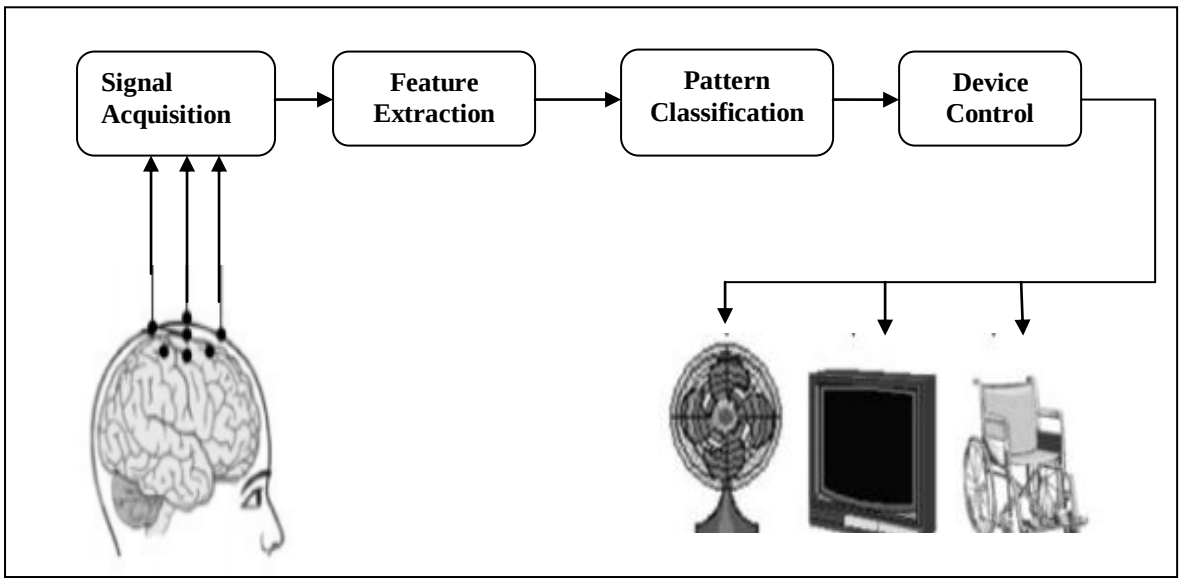


Fig. 4.3 BCI Overview

CSP is one of the most efficient feature extraction methods that enable a researcher to transform the acquired signals in a linear subspace such that the projected signals are

maximally discriminate. CSP algorithms provides the best results in deciphering motor imagery signals such as, left hand and right hand movements [8]. For binary classification problems, CSP can be defined as a linear spatial filter that seeks a projection direction such that the variance of the spatially filtered signals of one class is maximized and those of the other class is minimized. The reason behind the usage of the variance of the signals lies in the fact that variance is actually nothing but the band power of the signals, hence dealing with variance is supposed to yield better results. Basically, CSP algorithm attempts to derive optimal spatial filters that can best discriminate two different populations of EEG signals, in this approach simultaneous diagonalization of the covariance matrices has been used and in recent days, it is employed successfully for classifying movement related trials or EEG motor imagery.

4.2 RELATED WORKS

Due to the poor spatial resolution of EEG signals [9], CSP is considered to be a preferred alternative for enhancing the system performance and information transfer rate as well. Fukunaga *et al.* first introduced the concept of CSP [10] as an application of *Karhunen Loeve Transform* for classification and it was employed by Koles *et al.* later to detect the abnormal components of EEG scalp potential [11]. Ramoser *et al.* in [12] proposed a novel way estimating spatial filters using CSP algorithm and also stated the spatial filters for multichannel EEG analysis are capable of extracting discriminate information from two classes of EEG data and the recognition accuracies obtained after conducting experiments signify the proposed scheme as a promising one for EEG based BCI design. Lemm *et al.* in [13] attempted to alleviate the adverse impacts of non stationarity and artifacts in EEG based BCI system by embedding a finite impulse response filter in the CSP methodology, termed as Common Spectral Spatial Pattern (CSSP) and EEG recordings from the experiments of imagined limb movements have shown noticeable improvement in terms of Information Transfer Rate. Although CSSP provided certain improvement over the classical CSP algorithm, but first order FIR filters lacks sufficient flexibility. To overcome this drawback Dornhege *et al.* presented a novel technique Common Sparse Spectral Spatial Pattern (CSSSP) for simultaneous optimization of spatial and spectral filters in the periphery of CSP analysis [14]. To avoid the overfitting problem, the authors incorporated a regularization term to sparsify the solution that leads to an increase of the computational complexity. Since computationally expensive systems are not feasible to be implemented in online real time mechanisms, Tamioka *et al.* [15],[16] proposed an alternate version of CSP that exploits the temporal features by optimizing spectral filters in the frequency domain by considering the square root of the Rayleigh coefficient as the objective function. Motivated by these

techniques, later Wu *et al*. [17] later developed a novel algorithm termed as Iterative Spatio-Spectral Pattern Learning (ISSPL) that utilizes the concepts of statistical learning theory to optimize spatio-spectral filters.

After the introduction of CSP, gradually the needs of generalization of the algorithm is noticed, as to test the efficacy of an algorithm it is important to check how far it is performing on the testing data from the training data. Hill *et al.* in [18], compared the classifiability of the three brain signal acquisition methods, namely EEG, magnetoencephalography (MEG) and electrocorticography (ECoG) and after conducting a comparative study on the performance of spatial filters in each of the three sensor types, the authors inferred that spatial filtering enhances performance by a great deal in EEG, improves a little in ECoG and practically has no positive impact in MEG. It is also found due to overfitting problem CSP appears to be redundant to detect ERD in MEG and ECoG signals. Reuderink *et. al.* described the generalization of CSP over time, number of trials and subjects, but he could not conclude clearly regarding the actual reason of the overfitting problem observed with CSP [19]. The complexity of the EEG signal creates hindrance in the process of exploring the reasons for overfitting of CSP. The complexity of CSP can also be described from different aspects, firstly the nonstationarity of CSP does not entirely come from a single trial, so it consists of the between trial nonstationarities also. Further, still now there is no clear idea of the actual mechanism of brain that yields such signals for a particular cognitive task, hence the entire procedure is treated as an output from a black box, with no exact formulation of brain system. Moreover, the classification accuracy attained from various classifiers also varies amongst different subjects which lead to more confusion. In order to deal with these issues, the concept of regularization of CSP [20]-[22] has been introduced and from the past two decades the researchers are delving further to generate new ideas that can enhance the performance of spatial filtering in modern day BCI paradigms.

Lotte *et al*. in [23] has presented a unified framework based on the theoretical concepts of designing Regularized CSP algorithm, along with revisiting the already established RCSP algorithms, the researchers also formulated certain new algorithms.

4.3 CSP ALGORITHM FOR BINARY CLASSIFICATION PROBLEMS

CSP is a decomposition method that derives the spatial filters that generates maximally discriminate variance of the EEG signals in two different conditions. Basically, CSP maps the samples in a new space based on the covariance matrix of the signals [8]. Let us consider a task of discriminating between left hand and right hand motor imagery signals using CSP as a feature extractor. This can be seen as a simplified exemplary solution of the optimization criterion of the CSP algorithm: maximizing variance for the class of right-hand trials and at

the same time minimizing variance or left-hand trials. To illustrate mathematically, let X_k denotes the band pass filtered EEG recordings of *k*-th trials having a dimension of $C \times T$, where *C* and *T* denotes the total number of electrodes and the number of time samples per electrode in each trial respectively. Now, $Y_k \in (1, 2)$ denotes the class label of the *k*-th trial. The covariance matrix corresponding to both the classes can be calculated as,

$$\sum_{1(k:Y_k=1)} = (X_k X_k^T) \quad \text{and} \quad \sum_{2(k:Y_k=2)} = (X_k X_k^T) \tag{4.4}$$

The aim of CSP is to derive the coefficients of the weight matrix *W* and the elements (lying in the closed interval [0, 1]) of diagonal matrix *D*, using simultaneous diagonalization such that,

$$W \sum_1 W^T = I \quad \text{and} \quad W \sum_2 W^T = I - D \tag{4.5}$$

The derivations are carried out by first employing whitening transformation of ($\sum_1 + \sum_2$) and calculation of *P* such that,

$$P(\sum_1 + \sum_2) P^T = I \tag{4.6}$$

Now, using CSP analysis two more calculations are carried out to define another two matrices,

$$S_1 = P \sum_1 P^T \quad \text{and} \quad S_2 = P \sum_2 P^T \tag{4.7}$$

Moreover, a orthogonal matrix *R* and diagonal matrix *D* is calculated in terms of S_1 using the rules of spectral theory such that,

$$S_1^T = RDR^T \quad \text{and} \quad S_2^T = R(I - D)R^T \tag{4.8}$$

given, ($S_1 + S_2 = I$). It is needless to say, that the decomposition is possible due to the positive definiteness of the ($\sum_1 + \sum_2$). Here, it is quite remarkable to notice that the projection given by the *p*-th row of matrix *R* offers a relative variance equal to d_p (*p*-th element of diagonal matrix *D)* for the trials corresponding to class 1 and a relative variance of (1- d_p*)* for trials corresponding to class 2. It is evident that if d_p is close to 1 then it is the case of maximization of variance of the filtered signals corresponding to class 1 trials and obviously, (1- d_p) close to zero also signifies the same case. Finally, the weight matrix *W* can be obtained as,

$$W = R^T P \tag{4.9}$$

Using the weight matrix *W*, the EEG recordings of X_k is transformed in a subspace using,

$$Z_k = WX_k \tag{4.10}$$

The rows of *W* are basically spatial filters. In order to obtain the feature set that can be fed to the classifier for further proceedings in terms of signal analysis, using the normalized variance, the features are extracted as,

$$f_p = \log[\frac{\text{var}(Z_p)}{\sum_{i=1}^{2m} Z_i}] \tag{4.11}$$

In this case, *p* varies from *1* to *2m*, which basically corresponds to first *m* and last *m* rows of Z_k according to the largest Eigen values of each condition.

4.4 CSP ALGORITHM AS AN OPTIMIZATION PROBLEM

The CSP algorithm is highly successful in detecting ERD/ERS effects in motor imagery based BCI systems. The working principle of basic CSP algorithm can also be described from the pattern recognition aspect, incorporating the concept of optimization of an objective functions that is best suited for a particular problem. From this perspective, CSP algorithm attempts to obtain optimal spatial filters that maximizes the variance of EEG band pass filtered signals for one class while minimizes the same for the other class. Mathematically, CSP seeks to find the filters *w* that extremize the following objective function [23],

$$J(w) = \frac{w^T X_1^T X_1 w}{w^T X_2^T X_2 w} = \frac{w^T C_1 w}{w^T C_2 w} \tag{4.12}$$

Here, *T* denotes the transpose operation, X_i is the data matrix of *i*-th class, and C_i is the centred spatial covariance matrix from class *i*. To be precise extremizing (4.12), basically indicates extremizing $w^T C_1 w$ subjected to constraint $w^T C_2 w = 1$. Employing Lagrange's multiplier method, the constrained optimization problem reduces to extremizing the following function,

$$L(\lambda, w) = w^T C_1 w - \lambda(w^T C_2 w - 1) \tag{4.13}$$

To compute the filters *w* that extremizes the above function, it is required to equate the derivative of the function with respect to *w* to zero. Mathematically,

$$\frac{\partial L}{\partial w} = 2w^T C_1 - 2\lambda w^T C_2 = 0 \tag{4.14}$$

$$\Rightarrow C_1 w = \lambda C_2 w \tag{4.15}$$

$$\Rightarrow C_2^{-1} C_1 w = \lambda w \tag{4.16}$$

This is a mere Eigen value decomposition problem that is the spatial filters are the Eigen vectors of $M = C_2^{-1} C_1$, corresponding to largest and lowest Eigen values to generate the maximally discriminate outcome.

4.5 REGULARIZED CSP METHODOLOGY (RCSP)

Since recorded EEG signals are likely to be contaminated with noise and artifacts, to deal with the sensitivity of CSP to overfitting and noise, it is always recommended to regularize it by adding prior information [24]-[27]. Now, the regularization of CSP algorithms can be primarily categorized into two levels,

- Regularization at the covariance matrix estimation level
- Regularization at the objective function level.

4.5.1 Regularizing Covariance matrix estimates

The CSP algorithm depends majorly on the covariance matrix estimates of both the classes, so if such estimates suffer from the issues of noise or small training sets, then there is high probability that those matrices to be poor or non reliable estimates of the mental states involved and thus leading to poor spatial filters. In such cases, it is always wise to improve the estimates by adding prior knowledge to it in the form of regularization. The regularized covariance matrix estimated can be calculated as,

$$\tilde{C}_c = (1-\gamma)\hat{C}_c + \gamma I \tag{4.17}$$

where,

$$\hat{C}_c = (1-\beta)s_c C_c + \beta G_c \tag{4.18}$$

Here, $\tilde{C}_c$ denotes the regularized spatial covariance matrix estimate for class *c*, C_c is the initial spatial covariance matrix, s_c is a scaling parameter (a scalar), γ and β are two user defined regularization parameters in the closed interval [0, 1] and G_c signifies the 'generic' covariance matrix. The generic matrix is developed exploiting the information collected from

the EEG signals that have been recorded previously and it provides a rough estimate regarding how the covariance matrix corresponding to a specific cognitive task should be. It is seen clearly in (4.17) and (4.18) that using the regularization parameters γ and β, (4.18) reduces the original covariance matrix close the generic matrix and further, (4.17) shrinks the intermediate covariance matrix estimate to the identity matrix, to nullify the effect of an estimation bias that could have been generated as a result of a small training set and also to obtain a more stable estimate. To employ RCSP using this method, one has to replace the covariance matrix with its regularized estimates having one or two regularizing parameters.

4.5.2 Regularizing the CSP objective function

As discussed in section 4.4, the CSP algorithm primarily aims to find spatial filters that can extremize the objective function (4.12). In case of RCSP, a penalty function is added to the denominator containing another regularization parameter α, to outlaw solutions that do not satisfy a predetermined constraint. Formally, the objective function can be expressed as,

$$J_P(w) = \frac{w^T C_1 w}{w^T C_2 w + \alpha P(w)} \tag{4.19}$$

Generally, in our studies, we have calculated the spatial filters considering quadratic penalty functions of the form, $P(w) = w^T K w$ where the matrix K represents the aspect of the constraints that needs to be satisfied. So (4.19), can be written as,

$$J_P(w) = \frac{w^T C_1 w}{w^T (C_2 + \alpha K) w} \tag{4.20}$$

Following a similar approach, the Lagrangian equation in this case becomes,

$$L_P(\lambda, w) = w^T C_1 w - \lambda (w^T (C_2 + \alpha K) w - 1) \tag{4.21}$$

Thus the filters w that maximize (20) are the Eigen vectors corresponding to the largest Eigen values of, $M_1 = (C_2 + \alpha K)^{-1} C_1$. However, it is remarkable to note that unlike (4.12), to derive the optimize filters (4.20) is not being extremized. The reason can be explained as, minimizing (4.20) is essentially equivalent to maximizing the denominator, which contains a penalty term, will eventually maximize the penalty also and that is extremely undesirable, since the concept of penalty is introduced to deemphasize the filters not satisfying the constraint. In order to calculate the filters maximizing C_2 and minimizing C_1, the objective function is slightly modified as the positions of the variances corresponding to both the

classes are swapped and the filters are the Eigen vectors corresponding to the largest Eigen values of, $M_2 = (C_1 + \alpha K)^{-1} C_2$.

4.6 Improvising the existing RCSP Algorithms

Among the existing RCSP algorithms, till date Weighted Tikhonov Regularised CSP (WTRCSP) has been found to be the most efficient one amongst all the variants of RCSP algorithms that are presently available for EEG signal processing. However, in this case, the CSP algorithm is regularized at the objective function level where the penalty is assigned according to the usefulness of a specific electrode. To illustrate further, if an EEG electrode is likely to provide relevant contribution then those filters are given small penalty, on the other hand negligibly contributing electrodes are penalized greatly to improve the quality of the solution. According to WTRCSP, the penalty assigned to an electrode is calculated as the inverse of the average absolute value of the normalized weight of the corresponding electrode in CSP filters obtained from data acquired by other subjects, using (4.22),

$$w_G(i) = \left[\frac{1}{2 \times N \times |\Omega|} \sum_{i \in \Omega} \sum_{j=1}^{2N} \left| \frac{w_j^i}{\left\| w_j^i \right\|} \right| \right]^{-1} \tag{4.22}$$

where Ω is the total number of subjects participating in the experiment and w_j^i refers to the *j*-th spatial filter (out of Eigen vectors corresponding to *N* largest and *N* lowest Eigen values) corresponding to the *i*-th subject. Finally the penalty term is calculated as,

$$P(w) = w^T D_w w \text{ where } D_w = diag(w_G) \tag{4.23}$$

In (4.22), electrode usefulness is calculated after providing equal priority to each subject considered for the experiment, but it is important to note that all subjects do not provide equal performance, moreover number of trials acquired from different subjects, subject specificness, inter trial variance etc. all of these parameters have immense impact in this scenario and so neglecting these factors can generate degraded results.

- Firstly, (4.22) is modified to incorporate more constraints such that each subject is weighted using the number of trials he has performed. It is needless to say, a subject having more number of trials is likely to provide more promising results than another one having relatively lesser number of trials. In this way, the subjects having more number of trials are given more priority while deemphasizing the other subjects with lesser number of trials, as shown in (4.24),

$$w_G(i) = [\frac{1}{2 \times N \times |\Omega|} \sum_{i \in \Omega} \frac{N_c^i}{N_{t,c}} \sum_{j=1}^{2N} \left| \frac{w_j^i}{\left\| w_j^i \right\|} \right|]^{-1} \tag{4.24}$$

Here, N_c^i and $N_{t,c}$ denote the total number of trials corresponding to subject *i* and the total number of trials considering all the subjects.

- Secondly, the weights are defined as the Kullback Leibler divergence (K L divergence) between subject's data and can be expressed as,

$$w_G(i) = [\frac{1}{2 \times N \times |\Omega|} \sum_{\substack{i \in \Omega, m \in \Omega \\ (i \neq m)}} \frac{1}{Z \times KL(m,i)} \sum_{j=1}^{2N} \left| \frac{w_j^i}{\left\| w_j^i \right\|} \right|]^{-1} \tag{4.25}$$

where $Z = \sum_{m \in \Omega} KL(m,i)$ and the K L divergence can be mathematically expressed as,

$$KL(m,i) = \frac{1}{2}[\log(\frac{\det(C_c^i)}{\det(C_c^m)}) + tr(C_c^{i-1} C_c^m) - N_e] \tag{4.26}$$

where C_c^m and C_c^i denote the covariance matrices corresponding to *m*-th and *i*-th subject respectively and N_e denotes the number of considered electrodes, while *det* and *tr* respectively denote trace and determinant.

- Finally, the subject weights are defined according to inter trial variance, that is the subjects having high intertribal variance are given less priority as shown in,

$$w_G(i) = [\frac{1}{2 \times N \times |\Omega|} \sum_{i \in \Omega} s_i \sum_{j=1}^{2N} \left| \frac{w_j^i}{\left\| w_j^i \right\|} \right|]^{-1} \tag{4.27}$$

where

$$s = \sum_{m=1}^{C} \frac{1}{N_t} \sum_{n=1}^{N_t} e^{-\frac{1}{2}[\sum_{p=1}^{N_s} x_p^n - m_p]^2} \tag{4.28}$$

Here, *C* is the number of classes, N_t and N_s denote the number of trials and number of samples respectively and m denotes the mean vector obtained over all the trials.

4.7 CONCLUSION

This chapter basically recapitulates the background for the emergence of CSP along with showing innovative avenues for improvising the existing technologies in this domain. After revisiting the backdrop of Regularised CSP (RCSP), it also provides interesting insights about different aspects of cognitive penalty computation and finally three novel penalty terms have been derived as an extension of the well known Weighted Tikhonov Regularised CSP (WTRCSP) and the application of the same has been illustrated in Chapter 5.

REFERENCES

[1] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. M??ller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[2] G. Pfurtscheller and C. Neuper, "Motor Imagery and Direct Brain–Computer Communication," vol. 89, no. 7, pp. 1123–1134, 2001.

[3] S. Datta, P. Rakshit, A. Konar, and A. K. Nagar, "Selecting the optimal EEG electrode positions for a cognitive task using an artificial bee colony with adaptive scale factor optimization algorithm," in *Evolutionary Computation (CEC), 2014 IEEE Congress on*, 2014, pp. 2748–2755.

[4] S. Datta, A. Khasnobish, A. Konar, and D. N. Tibarewala, "Cognitive Activity Classification from EEG Signals with an Interval Type-2 Fuzzy System," in *Advancements of Medical Electronics*, Springer, 2015, pp. 235–247.

[5] D. J. McFarland, L. M. McCane, S. V David, and J. R. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 3, pp. 386–394, 1997.

[6] F. Lotte, "Oscillatory EEG-based BCI design : signal processing and more," 2011.

[7] S. Sun and J. Zhou, "A Review of Adaptive Feature Extraction and Classification Methods for EEG-Based Brain-Computer Interfaces."

[8] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2006, pp. 5392–5395.

[9] P. L. Nunez, R. Srinivasan, A. F. Westdorp, R. S. Wijesinghe, D. M. Tucker, R. B. Silberstein, and P. J. Cadusch, "EEG coherency: I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 5, pp. 499–515, 1997.

[10] K. Fukunaga, *Introduction to statistical pattern recognition*. Academic press, 2013.

[11] Z. J. Koles, "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 79, no. 6, pp. 440–447, 1991.

[12] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *Rehabil. Eng. IEEE Trans.*, vol. 8, no. 4, pp. 441–446, 2000.

[13] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, "Spatio-spectral filters for improving the classification of single trial EEG," *Biomed. Eng. IEEE Trans.*, vol. 52, no. 9, pp. 1541–1548, 2005.

[14] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Muller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *Biomed. Eng. IEEE Trans.*, vol. 53, no. 11, pp. 2274–2281, 2006.

[15] R. Tomioka, G. Dornhege, G. Nolte, K. Aihara, and K.-R. Müller, "Optimizing spectral filters for single trial EEG classification," in *Pattern Recognition*, Springer, 2006, pp. 414–423.

[16] R. Tomioka, G. Dornhege, G. Nolte, B. Blankertz, K. Aihara, and K.-R. Müller, "Spectrally weighted common spatial pattern algorithm for single trial EEG classification," *Dept. Math. Eng., Univ. Tokyo, Tokyo, Japan, Tech. Rep*, vol. 40, 2006.

[17] W. Wu, X. Gao, B. Hong, and S. Gao, "Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL)," *Biomed. Eng. IEEE Trans.*, vol. 55, no. 6, pp. 1733–1743, 2008.

[18] N. J. Hill, T. N. Lal, M. Schröder, T. Hinterberger, G. Widman, and C. E. Elger, "Classifying eventrelated desynchronization in EEG, ECoG and MEG signals. DAGM." Berlin, Heidelberg (Germany): Springer-Verlag, 2006.

[19] B. Reuderink and M. Poel, "Robustness of the common spatial patterns algorithm in the BCI-pipeline," 2008.

[20] G. Townsend, B. Graimann, and G. Pfurtscheller, "A Comparison of Common Spatial Patterns With Complex Band Power Features in a Four-Class BCI Experiment," vol. 53, no. 4, pp. 642–651, 2006.

[21] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 683–686, 2009.

[22] K. P. Thomas, C. Guan, C. T. Lau, A. P. Vinod, K. K. Ang, and A. Event-related, "A New Discriminative Common Spatial Pattern Method for Motor Imagery Brain – Computer Interfaces," vol. 56, no. 11, pp. 2730–2733, 2009.

[23] F. Lotte, C. Guan, F. Lotte, C. Guan, and S. Member, "Regularizing Common Spatial Patterns to Improve BCI Designs : Unified Theory and New Algorithms Regularizing Common Spatial Patterns to Improve BCI Designs : Unified Theory and New Algorithms," 2011.

[24] F. Lotte and C. Guan, "Spatially regularized common spatial patterns for EEG classification," *Proc. - Int. Conf. Pattern Recognit.*, pp. 3712–3715, 2010.

[25] H. Ghaheri and A. Ahmadyfard, "Temporal windowing in CSP method for multi-class Motor Imagery Classification," pp. 1602–1607, 2012.

[26] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain-computer interfacing.," *J. Neural Eng.*, vol. 9, no. 2, p. 026013, 2012.

[27] A. Ashok, A. K. Bharathan, V. R. Soujya, and P. Nandakumar, "Tikhonov regularized spectrally weighted common spatial patterns," *2013 Int. Conf. Control Commun. Comput. ICCC 2013*, no. Iccc, pp. 315–318, 2013.

Chapter 5

Detecting Motor Imagery EEG Signals Using CSP

This chapter provides an insight to the role of CSP algorithm as a feature extration procedure in the context of EEG motor imagery signal detection. It addresses two distinct perspectives of motor imagery signal discrimination and each one of them have different highlighted sections in the work that has been carried out. Section 5.2 introduces a novel scheme of motor imagery classification using a Regularised variant of CSP using an ensemble of k-NN classifiers, and the main focus of the work lies in the regularization of CSP objective function by adding subject specific trial information into it. Section 5.3 proposes another novel scheme of deciphering motor imagery EEG signals, this time the focus has been shifted to the classification phase. The major novelty lies within an ADE induced sparse network following the principles of traditional artificial neural network.

5.1 INTRODUCTION

Brain Computer Interfacing (BCI) has been emerged as one of the most interesting research disciplines during the past few decades because of its outstanding contributions towards rehabilitative applications that provide aid to the patients suffering from diseases like Amyotrophic Lateral Sclerosis (ALS), Locked-in syndrome or Brain stem stroke [1]. As more and more research enthusiasts are investing their physical and intellectual resources in this domain, the researchers are enlightened about the limitations of the concerned technology and they are coming up with novel ideas to obtain the desired output either by solving those drawbacks or by circumventing the limitations using certain intelligent strategies. BCI systems enable us to instruct a machine with brain activation patterns thus not requiring the peripheral nervous system for conveying any message. Electroencephalogram (EEG) is a non-invasive modality that is comprised of multivariate data corresponding to neuronal electrical potentials at sampled time intervals measured from scalp electrodes [2]. For EEG driven BCI systems, motor imagery (MI) is found to be one of the most popular paradigms to stimulate a subject's brain potentials. It is accompanied by neuronal power enhancement or attenuation of signals generated during performing imagined body part movements, such phenomena are formally termed as Event Related Synchronization (ERS) and Event Related Desynchronization (ERD) respectively [3]. To deal with the drawbacks of EEG paradigm, the researchers have introduced the concept of spatial filtering with an aim to project the acquired raw brain signals in a subspace such that it brings out the most discriminative components of those signals.

Common Spatial Pattern (CSP) [4] is an efficient as well as discriminative spatial filter that yields the most discriminative features corresponding to a binary classification problem, such that variance of the filtered signals of one class is maximized and those of the other class is minimized. This feature extraction algorithm was primarily formulated in the context of EEG/MEG analysis considering two classes only and it generated extremely precise classification accuracies when applied to BCI systems utilizing MI paradigms (for example left hand and right hand motor imagery). Despite its efficiency and adaptability, CSP lacks robustness while dealing with noisy data and outliers, which in turn degrades the system performance. To deal with these issues, researchers have designed a novel framework of adding prior information to the existing CSP methodology in terms of regularization parameters to ensure improved performance against outliers. This new variant of CSP, termed as Regularized CSP (RCSP) has outperformed the traditional one by a large extent in most of the cases.

5.2 DISCRIMINATING MOTOR IMAGERY EEG SIGNALS USING RCSP ALGORITHM

Common Spatial Pattern is considered as one of the most efficient feature extraction methods for rehabilitative BCI paradigms. In the past many researchers have devoted their financial and intellectual resources in the said domain with an aim to regularize the concept of CSP by adding prior information in order to reach closer to the desired objective. In this section, a novel penalty term has been introduced as an extension of one of the most popular variants of RCSP algorithms, formally termed as Weighted Tikhonov Regularized CSP. The proposed strategy has been implemented for deciphering four class Motor Imagery signals recorded from different users while performing four imagined movements of right hand and left hand. An ensemble classifier comprising of k-NN layers has been used for the classification purpose and finally, the efficiency of the proposed framework has been tested against the traditionally used standard classifiers and in each case the designed algorithm outperformed the others.

Literature shows Ramoser *et al.* first introduced CSP as an optimal spatial filter that can generate maximally different features that leads to reliable discrimination of single trial left and right hand MI signals [5]. As mentioned earlier, CSP was initially formulated for binary classification problems; later its multiclass extensions have been proposed and found to generate good results. Kang *et al.* in his attempt to consider the changes of subject specific spatial distributions of ERD/ERS in time domain in particular frequency bands, designed an optimal spatial filter by decomposing the raw EEG signals in a space-time-frequency feature space [6]. Wentrup *et al.* proposed a novel multiclass extension of traditional CSP by incorporating mutual information based Information Theoretic Feature Extraction (ITFE) methodology after inferring that CSP using simultaneous diagonalization procedure is equivalent to Independent Component Analysis (ICA) [7]. Moreover, after introduction of RCSP, in [8]-[10], although the researchers presented different possible aspects of the same, but they were unable to express the entire concept in a unified presentation instead of expressing different regularized formulations. Lotte *et al*. in [11] formulated a unified framework after accumulating almost all the existing regularization concepts and introduced a few improvised frameworks and presented a detailed analysis of all the methodologies by applying those on standard datasets acquired from BCI competition.

The present section mainly addresses two major issues. Firstly, CSP is known to be a user specific spatial filter, which does not take into account any other subject's data while performing the same cognitive task except the concerned subject. Due to this reason, often the performance of CSP is degraded drastically in case of users having lesser number of trials. Further, for a particular cognitive task not all scalp regions are supposed to provide relevant signals, hence it is wise to rank the EEG electrodes in order of their usefulness corresponding

to a particular cognitive task. In the present scenario, a penalty function, considering the above two aspects has been derived which on one side penalizes the channels which are supposed to contribute less corresponding to a specific cognitive task and on the other hand it judiciously focuses on the subjects having more number of trials while deemphasizing the other subjects.

5.2.1 Proposed Methodology

Literature shows already many researchers have devoted their extraordinary efforts in order to derive different penalty terms that covers various aspects of the drawbacks faced while using an EEG based BCI system, but none of them could satisfactorily merge the methodologies so that one unified framework can take care of all the limitations with efficiency. These issues motivated us to delve into the matter and come up with a solution that can serve the purpose.

5.2.1.1 Novel Penalty Term

As mentioned in [11], out of the penalty terms that has been applied so far, CSP with Weighted Tikhonov Regularization (WTRCSP) is found to generate the best performance in terms of several performance indices like accuracy, robustness etc.

As stated earlier, for a particular cognitive task not all scalp regions will be contributing equally, but again choosing only a smaller subset of electrodes may lead to loss of relevant information which can have a negative impact on the system performance. Instead, it is an intelligent way to rank the electrodes according to their usefulness for a concerned cognitive task and then introduce the concept of penalty by deemphasizing the channels having lesser contribution and focusing more on the channels that are likely to provide relevant data.

Similarly, not all subjects provide optimal data which can degrade system performance especially for the subjects having lesser number of trial recordings. In this paper, we propose a novel penalty, formally termed as Composite WTRCSP which computes the usefulness coefficient of every channel considering each and every subject's data weighted by a term depending on their individual number of trials. Earlier, similar work has been introduced in regularization of CSP at covariance matrix estimate level, but combining these two genre of regularization to cover all the aspects is not an easy work, in fact that leads to an increased system complexity which is not acceptable.

In the proposed framework, a quadratic penalty term $P(w)$ is considered, where K is a diagonal matrix such that $K=Diag(U_G)$ and $Diag$ denotes the diagonalize operation. $U_G(i)$ signifies the level of penalty assigned to the channel i. Using preliminary neurophysiologic knowledge it is often possible to assign penalty values to the concerned electrodes, but theoretic explanations may not always fit well to all the subjects participating in the experiment and due to this inexact information the system performance can be deteriorated.

Hence, in this work, the penalty values have been obtained using other subjects' data, which is nothing but the inverse of the normalized average absolute weight of the electrode obtained from the composite weighted spatial filters corresponding to other subjects. The mathematical formulation can be expressed as,

$$U_G = \left(\frac{1}{2\times f\times|\Omega|}\sum_{i\in\Omega}\frac{N_i}{N_{total}}\sum_{j=1}^{2f}\left|\frac{w_j^i}{\left\|w_j^i\right\|}\right|\right)^{-1} \tag{5.1}$$

Since the more will be the usefulness of an electrode, the less is supposed to be the assigned penalty; to carry out this purpose inverse operation is chosen. Ω is the set of all the subjects participating in the experiment, and w_j^i denotes the weight of j^{th} spatial filter of i^{th} subject obtained using CSP. N_i and N_{total} respectively denote the total number of trials of i^{th} subject and the total number of trials conducted for all the subjects. The usefulness coefficients are calculated in a way such that the information obtained from the subjects having more number of trials are given more priority in comparison to the subjects having relatively lesser number of trials, thus this modified Composite WTRCSP is likely to reflect the channel usefulness more accurately than previous one.

5.2.1.2 Ensemble of k-NN Classifiers

In supervised learning problem Ensemble classifier is an important part as it trains a number of individual classifiers in co-operative manner. It is seen that depending upon the distribution of the training data, all the training data are not learned well by any single classifier and accuracy on test dataset becomes poor [12]. A better approach is to use number of classifiers as a group where each individual classifier is called base learners. Base learners are taught separately and decision obtained from base learners are combined to generate a single decision. Base learners may be heterogeneous or homogenous, to perform a classification task with coordination [13].

A classification problem can be stated as mapping between set *X* and set *Y* where $X=\{(x_1,y_1),(x_2,y_2).....(x_n,y_n)\}$ is set of *n* training data and each instance belongs to some unique domain. Y is label data in integer form such that $Y=\{1,2,....K\}$ Here, the objective is to find the function which maps each element of set *X* to set *Y*.

k-NN can be regarded as such function for mapping, where the method consists of storing *k*-patterns which are nearest to the test pattern under consideration and then it calculates the Euclidian distance of the test pattern from all other patterns. For any two vector x_1 and x_2 in the same feature space, Euclidean distance D is calculated as [14],

$$D^2_{1,2} = (x_1 - x_2)(x_1 - x_2)' \tag{5.2}$$

The most probable classes get the nomination from previously selected k-nearest neighbours and the class which gets the most nomination is selected as the class of that particular test pattern [15]. *K* is very important parameter of the k-NN classifier as the local density of data is controlled by it and of course it is much smaller than the training sample size. The distances from the test sample to other samples are stored in ascending order such that $D1_{(x)} \leq D2_{(x)} \leq ....Dn_{(x)}$ where $D1_{(x)}$ is the distance from the nearest neighbour and so on. The density estimate of k-NN is given as follows [16];

$$\hat{U}(x) = \frac{k}{2NDk_{(x)}} \tag{5.3}$$

The ensemble classifier performs the two prime tasks, first it infers the classes from individual classifiers and second it generates a combination rule which ultimately decides the class based on the results of individual classifier.

Out of many combination rules, this section uses random subspace method for constructing ensembles [17]. Here the method generates random subspaces from original feature space to train the base learners. Suppose $X = \{ x_1, x_2, ...x_n \}$ be the *n* input feature. To generate a random space ensemble with *D* number of classifiers, *D* number of samples each size of *W* is drawn at random, this is done without replacement and over uniform distribution *X*. This method requires two parameters for constructing ensemble, first the number of ensemble size and other one is cardinality of subspace. The final decision is taken based on majority vote. It needs to be mentioned here that subspace method is more useful than any other rules which modifies the input training data, as k-NN is stable with the modification of input data but sensitive to any input perturbation [16].

5.2.2 Experimental Setup

This section provides a detailed description of various experiments that were undertaken to detect the efficacy of the proposed framework. EEG data was acquired from 8 healthy subjects having undergone no such major diseases in the recent past. As EEG signal is subject specific, so equal number of participants were chosen from both the gender (male & female). Age group of all the subjects were between 22-30. Objective and methods of the experiment were made clear to them prior to the experiment. A consent form was also signed by them

stating their willingness to participate in the study. All other ethical and safety issues were maintained according to the Helsinki Declaration of 1972, revised in 2000 [18].

The main objective of the section is to introduce a novel approach on RCSP to decode the motor imagination signal, therefore motor cortex along with parietal and temporal region of brain has been considered for acquiring EEG signals. 6 electrodes viz. *C3, Cz, C4, P3, Pz, P4* were placed over the scalp. EEG amplifier has the sampling frequency of 200 Hz and made by NIHON-KOHDEN and all the electrodes were made of AgCl.

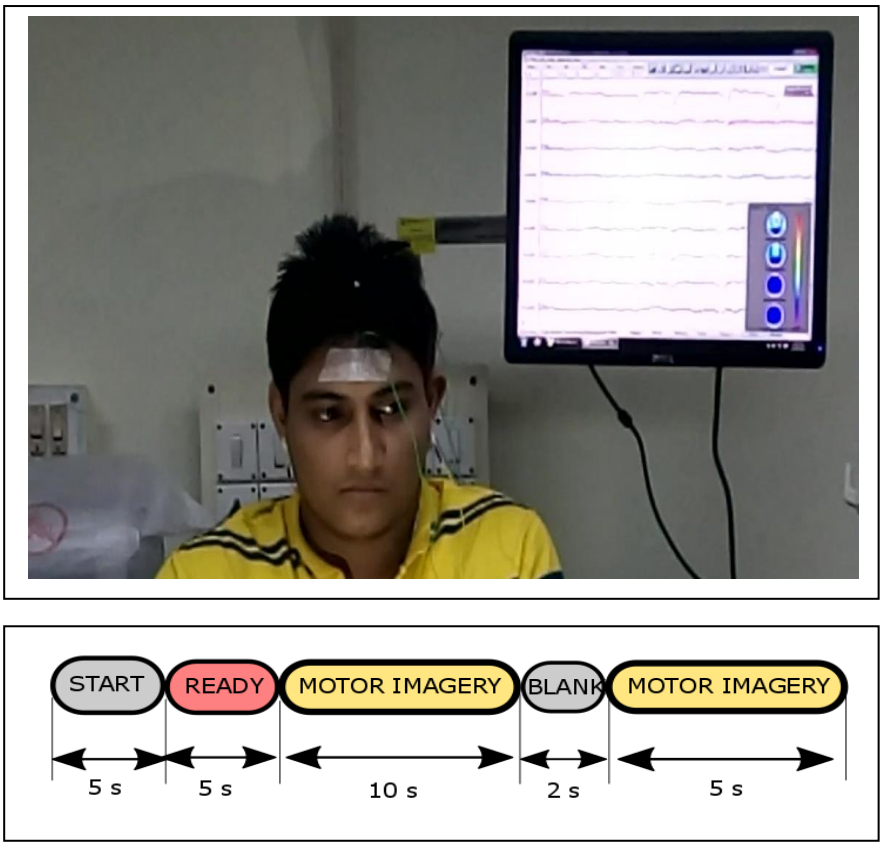


Fig.5.1 Experimental Setup and Visual Stimuli

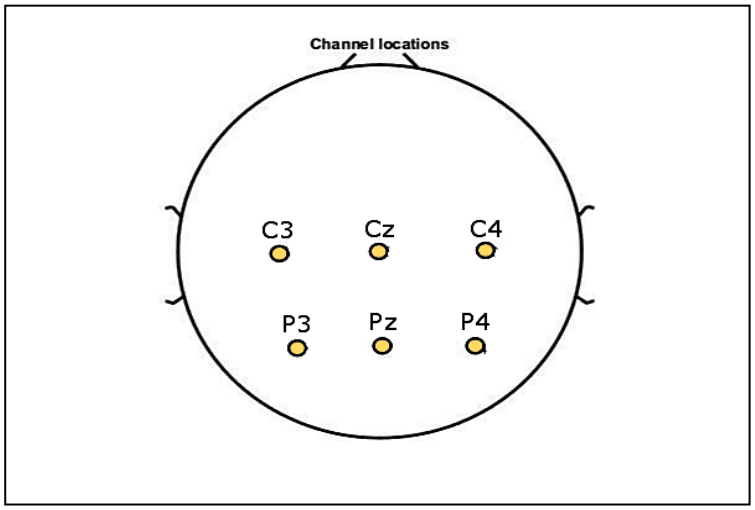


Fig. 5.2 Electrode Locations

5.2.2.1 Design of Visual Stimuli

To maintain the synchronisation of response timing, visual stimuli were shown to all subjects and they had to perform the imagination task accordingly. This section employs four class motor imagery given as, right arm stretch about shoulder, left arm stretch about shoulder, left arm fold about elbow, right arm fold about elbow. Each of the instruction appears on the screen randomly but with equal duration of 5 seconds. Between each such command a blank screen appears for 2 seconds duration. At beginning a 'start' screen appears for 5 seconds followed by a mark screen of 5 seconds to make subjects concentrate on study. Each subject performed 5 such sessions in each trial, but since the proposed methodology of RCSP concerns mainly about penalizing subjects having lesser number of trials, so in order to check the adaptability of the proposed framework, the subjects were asked to perform unequal number of trials unlike the traditional ways of equal trials. Fig. 5.1 and Fig. 5.2 depict the stimuli timing diagram and electrode placement of the experiment respectively.

5.2.2.2 Preprocessing of EEG Signals

Raw EEG data is contaminated with various artefacts due to eye blinking and head movement. Power spectral density of raw data reveals that signal power is dominant in the frequency range 0.1-32 Hz, an elliptical bandpass filter of order 10 has been employed to extract the desired frequency range.

TABLE 5.1 AVERAGE CLASSIFICATION ACCURACY(%) OF THE EMPLOYED ENSEMBLE CLASSIFIER IN COMPARISON WITH OTHER STANDARD CLASSIFIERS

Class / Classifier	Class1 (Left Hand Shoulder Stretch)	Class2 (Right Hand Shoulder Stretch)	Class3 (Left Hand Elbow Stretch)	Class4 (Right Hand Elbow Stretch)
Proposed Ensemble Classifier	88.57	86.72	82.35	81.48
k-NN Classifier	82.65	83.79	80.86	79.53
MLP SVM Classifier	74.39	72.22	70.68	71.21
Naïve Bayes Classifier	65.47	66.07	62.49	61.13

- ***CAR Filtering***

Common average referencing has been done to spatially filter the EEG signal and to get rid of local interference caused by neighbour electrodes. The mean of all the channels are

subtracted from each individual channel to eliminate the influence of far field sources. If raw signal of 6 electrodes is designated as $a_i(t)$ where $i = 1,2,...6$, then CAR filter can be defined by following formula,

$$a_i(t) - \frac{1}{6}\sum_{i=1}^{6} a_i(t) \tag{5.4}$$

5.2.2.3 Feature Extraction

In the present work, after introduction of the novel penalty term based Composite WTRCSP, only the largest two spatial filters out of six spatial filters have been used for filtering purpose. After obtaining the spatially filtered signals, final feature set is computed as the logarithm of the variance of the filtered signals.

5.2.2.4 Classification

Table 5.1 provides a tabulated description of the mean classification accuracies obtained class wise after employing the described Ensemble classifier and studying the results mentioned in Table 5.1, it is quite clear that the Ensemble classifier outperforms the other standard classifiers in terms of precision and thus validates our preference for the same over the other classifiers.

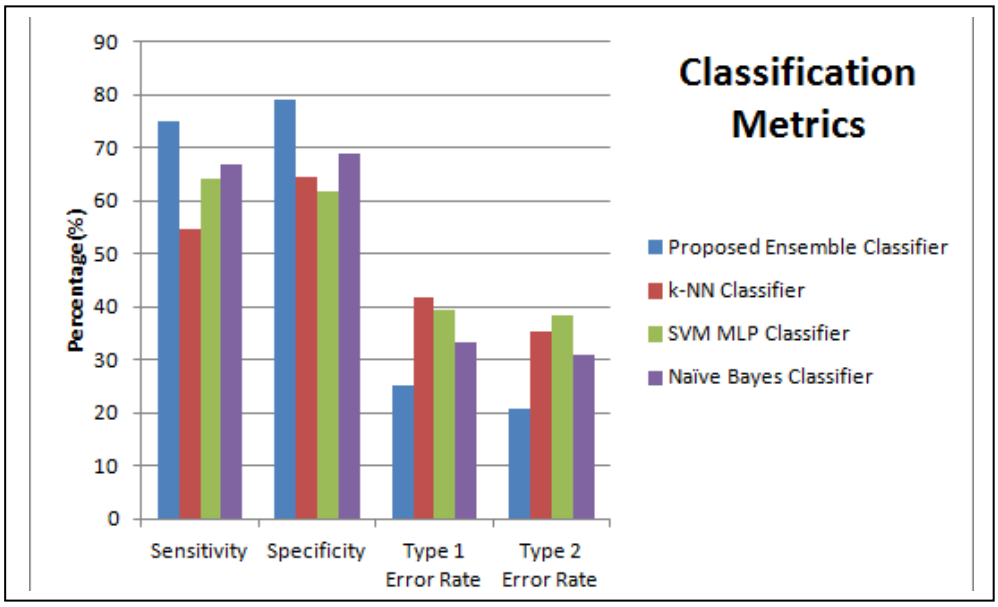


Fig. 5.3 Classification Metrics

Fig.5.3. presents a detailed description regarding the performance of the Ensemble Classifier in terms of Sensitivity, Specificity, Type 1 Error Rate and Type 2 Error Rate in comparison with three other standard classifiers. In the present work, Sensitivity, Specificity, Type 1 Error Rate and Type 2 Error Rate are defined as,

$$Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative} \tag{5.5}$$

$$Specificity = \frac{TrueNegative}{TrueNegative + FalsePositive} \tag{5.6}$$

$$Type1Error = \frac{FalsePositive}{FalsePositive + TrueNegative} \tag{5.7}$$

$$Type2Error = \frac{FalseNegative}{FalseNegative + TruePositive} \tag{5.8}$$

It is evident from Fig.5.3 also that the chosen classifier surpasses its competitors in terms of the specified parameters by a large margin.

5.2.2.5 Statistical McNemar's Test

In this section, McNemar's test [19] has been employed for detection of the performance of two classification algorithms for correctly classifying data samples. Suppose, using a common input phylogenetic sequence, the outputs generated by the algorithm A and algorithm B are f_A and f_B respectively. Let n_{01} be the number of data samples misclassified by algorithm A but not by algorithm B and n_{10} be the number of data samples misclassified by algorithm B but not by algorithm A. Then, a statistic termed as *Z* score is defined by using the following equation,

$$Z = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}} \tag{5.9}$$

Here *Z* has been evaluated in order to denote a comparator statistic of misclassification between the proposed Ensemble classifier (Algorithm A) and any of the above mentioned competitor algorithms (Algorithm B). In Table 5.2 the null hypothesis has been rejected, if *Z*> 3.84, where 3.84 is the critical value of the chi square distribution for 1 degree of freedom at probability of 0.05. It can be clearly seen from the results that the proposed classification algorithm outperforms the other competitors except k-NN classifier.

TABLE 5.2 STATISTICAL ANALYSIS OF CLASSIFIERS USING MCNEMAR'S TEST

Reference Algorithm: Proposed Ensemble Classifier				
Classifier algorithm used for comparison using desired features d=50	**Parameters used for McNemar's test**		***Z***	**Comments on Acceptance/Rejection of Hypothesis**
	n_{01}	n_{10}		
MLP SVM	87	120	4.946	Rejected
Naïve Bayes	90	121	4.265	Rejected
k-NN	97	116	1.521	Accepted

This chapter introduces a novel penalty term in the objective function of RCSP based BCI systems such that the concerned system emphasizes on the subjects having performed larger number of trials while computing the usefulness coefficients of different electrodes and thus penalizing the subjects having relatively lesser number of trials. The proposed framework has been implemented in a system considering the reduced numbers of chosen electrodes placed on certain specific scalp regions that are likely to provide relevant signals while performing a chosen cognitive task (MI) according to prior neurophysiological knowledge, but to detect the efficiency of the said system it is equally important to employ the proposed methodology in a system considering a larger number of electrodes placed on the entire scalp following any standard electrode placement system. Moreover, the newly derived penalty term takes into account subject-specificness, while there are several other important aspects that should be addresses while dealing with such a system. Further, the mean classification accuracy of (88%) can be considered only to be fairly good in this research discipline, hence further efforts should be devoted to enhance the accuracy precision as far as possible.

5.3 DECIPHERING MOTOR IMAGERY EEG DATA USING PROXIMITY BASED ADE INDUCED SPARSE NETWORK

This chapter proposes a novel approach for classification of four class motor imagery data using a sparse neural network classifier. EEG signals were recorded using 19 different electrodes placed on scalp excluding the reference electrodes, from 9 different subjects corresponding to imagined movements of left hand, right hand, foot and tongue. The proposed framework presents a novel Adaptive Differential Evolution based weight adaptation scheme and implemented the said scheme in a classical Back Propagation neural network (BPNN). The features were extracted using Common Spatial Pattern (CSP) algorithm, and the Principal Component Analysis (PCA) was employed as a dimension reduction tool. Further, to validate the performance of the proposed methodology, the concerned classifier performance has been compared with other standard classifiers. From the

experimental results, it is concluded that CSP filtered signals along with the designed classifier outperforms the other traditionally used procedures. Apart from improving the classification accuracy by the newly designed classifier, the proposed framework also checks the impact of physiologically generated artifacts (eye movement) upon the system performance by conducting the concerned procedures without removing those artifacts and infers that such artifacts have negligible impact in case of motor imagery based cognitive tasks.

Due to cost efficiency, high time resolution, less environmental limits and most importantly its wide application in decoding mental states using pattern classifiers, electroencephalogram (EEG) is always preferred in practical BCI applications compared to other non-invasive modalities [20] like fMRI, fNIRS etc. Motor Imagery is one of the most popular paradigms used in EEG based BCI system. Motor imagery based BCI is basically formulated using a pattern recognition approach [21], where a subject issues a command by imagining a movement of a particular limb, which in turn causes a change in the rhythmic activities in specific locations of brains corresponding to the concerned limb. After extracting relevant features from raw EEG signals it is easy to detect the subject's mental state with the help of a classifier that has been trained using similar features with known class labels.

To check the accuracy of the entire BCI system, electrodes are placed over the entire scalp area as skipping any particular area may lead to the loss of information. While recording EEG signal readings from the subjects corresponding to any cognitive task, the resulting signals are often contaminated by artifacts and noise. A common source of such noise is artifacts generated due to eye blinking. To avoid dealing with such noisy data, different filters [22] are used for raw EEG data preprocessing to lessen the impact of noise. But, there is a specific kind of artifact generated due to eye movement (blinking, flutter) that has very minimal impact on system performance.

Moreover, in EEG based BCI research, feature extraction is a key step, but irrespective of the feature type selected, the resulting data dimension is always found to be pretty high. Hence, various data dimension reduction techniques are introduced which enable analysts to avoid working with higher dimensional data thus reducing system complexity. The final step is related to classification which assigns a particular class label to an unknown data with the help of an already trained classifier.

Because of the growing interest in this particular domain, a large number of research groups have put their efforts in this area. Literature show that Pfurtscheller *et al*. first used EEG classification on Motor Imagery based Event Related Desynchronization (ERD) for a BCI application [2]. Later, Ramoser *et al.* used the common spatial pattern algorithm for feature extraction, in a single trial motor imagery based BCI [5]. Popescu *et al*. proposed a new

EEG cap consisting of dry EEG electrodes avoiding the bottle-neck of time consuming gel application and also considering much less number of electrodes than a standard EEG setup [23]. Lal *et al*. proposed a novel channel selection method based on recursive elimination of least contributing electrodes according to the classification accuracy with respect to a classifier like Support Vector Machine (SVM) [24]. Wang *et al.* used common spatial pattern algorithm (CSP) to obtain the spatial patterns corresponding to hand and foot motor imagery and further channels were selected based on the CSP spatial filter co-efficient [25].

Much work has also been carried out with an aim to enhance the efficiency of the classifiers in various real time applications. Heerman *et al*. employed Back Propagation Neural Network (BPNN) for multispectral image data classification purpose by applying it to simulated as well as real satellite imagery [26]. Paola *et al.* presented a detailed comparison between Back Propagation Neural Network (BPNN) and maximum likelihood classifiers in terms of working principle to find out how these two algorithms perform differently on the same image [27]. Qin et al. developed a self adaptive Differential Evolution algorithm, where prior assumption of the learning rule and the parameters were no longer required, they were adapted through learning experience only [28].

Although this chapter also attempts to classify motor imagery signals [29], it majorly emphasizes on two aspects of EEG motor imagery classification. Firstly, the artifacts generated due to eye movements, blink or eyelid flutter has minimal impact on the classification of motor imagery signals, since the noise generated for the above mentioned physiological processes has maximum impacts on electrodes placed close to eye (prefrontal electrodes). With the elementary neurophysiological knowledge, it can be said easily that in case of motor imagery classification, motor cortex electrode signals generate more relevant information than the other channels, so conducting experiments without artifact removal does not have a major impact on system performance. Secondly, the paper proposes a novel weight adaptation approach for designing a neural network classifier which can recognize the class of an unknown motor imagery signal after being trained by similar data acquired from various subjects. The highlighted part of the approach lies in rejection of the links having minimum weight without sacrificing the accuracy. A variant of Differential Evolution has been developed and employed in order to collect the optimized weights. Differential Evolution has been preferred over other alternatives because of its rapid rate of convergence and high performance precision. The proposed framework may find tremendous appreciation in the field of cognitive task based BCI systems, because accurate classification of motor imagery signals still remains a challenging one. Apart from that, the paper employs common spatial pattern (CSP) feature extraction algorithm and Principal Component Analysis (PCA) for dimension reduction purpose.

5.3.1 Preliminaries

This section provides an elementary idea about the working principle of the Differential Evolution (DE) algorithm and Artificial Neural Network (BPNN).

5.3.1.1 Differential Evolution

Differential Evolution is undoubtedly belongs to the category of popular evolutionary algorithms because of its ease of use, low complexity, higher speed of convergence, less requirement of parameters and most importantly high precision [30]. The Differential Evolution Algorithm used here is can be explained as follows,

1) **Initialization:** Initialize a population of *NP D* dimensional members ($\overrightarrow{X_{i,G}}$) as trial solutions to an optimization problem. The i^{th} data point of the population is expressed as,

$$\overrightarrow{X}_{i,G} = [x_{1,i,G}, x_{2,i,G}, ..., x_{D,i,G}] \tag{5.10}$$

where *G* is the generation number. The vectors are chosen by maintaining a uniform distribution of the entire range of $\overrightarrow{X}$ so that the DE algorithm can easily search for the global optima. So the k^{th} data point of the i^{th} vector can be defined as,

$$x_{k,i,0} = x_{k,\min} + r_{i,k} * (x_{k,\max} - x_{k,\min}) \tag{5.11}$$

where *G=0* denotes the initial population, $r_{i,k}$ denotes a uniformly distributed parameter in [0,1], and

$$x_{k,\min} \in \overrightarrow{X}_{\min} = \{x_{1,\min}, x_{2,\min}, ..., x_{D,\min}\} \tag{5.12}$$

$$x_{k,\max} \in \overrightarrow{X}_{\max} = \{x_{1,\max}, x_{2,\max}, ..., x_{D,\max}\} \tag{5.13}$$

2) **Mutation:** For each candidate solution, 3 helping agents called target vectors are generated to the present generation from the parent vector. Now, a mutant vector termed as donor vector is generated using differential mutation operation and finally an offspring is formed by recombination of donor and target vectors, known as trial vector.

$$\overrightarrow{X_{i,G}}' = \overrightarrow{X_{m^i,G}} + F(\overrightarrow{X_{n^i,G}} - \overrightarrow{X_{o^i,G}}) \tag{5.14}$$

where $\overrightarrow{X_m}$, $\overrightarrow{X_n}$ and $\overrightarrow{X_o}$ are helping agents, such that the indices *m*, *n* and *o* are chosen randomly [1,*NP*] so that they are mutually different and also different from base index *i*.

The donor vectors are chosen from the current population only. The value of *F* is assumed in (0, 1).

3) **Recombination:** To boost up the population after mutation, the donor vector interchanges its components with the trial vector depending upon the value of a randomly generated number. Basically, For each pair of $\overrightarrow{X_{i,G}}$ and $\overrightarrow{X_{i,G}}'$ we randomly generate a no in (0, 1) and if it is greater than Cr we consider $\overrightarrow{X_i}'$, otherwise we take $\overrightarrow{X_i}$ to form the target vector $\overrightarrow{X_{i,G}}'' = [x''_{1,i,G}, x''_{2,i,G}, ..., x''_{D,i,G}]$.

4) **Selection:** From each pair of $\overrightarrow{X_i}$ and $\overrightarrow{X_{i,G}}''$, allow only one member to enter the population of the next generation using the following policy, if $f(\overrightarrow{X_i}) < f(\overrightarrow{X_{i,G}}'')$ then $\overrightarrow{X_{i,G+1}} = \overrightarrow{X_{i,G}}$ else $\overrightarrow{X_{i,G+1}} = \overrightarrow{X_{i,G}}''$ depending upon fitness function value. Thus DE allows the fittest chromosome to generate offspring to the next generation, hence using this procedure the fitness value of the population either improves or remains same but never worsens.

5.3.1.2 Artificial Neural Network

The applicability of artificial neural network has been explored by the research professionals by employing the architecture in various domains including computer technology, satellite communication, biology, psychology etc. The purpose and the methods of application varied but the central idea remained similar in each of these cases. The development of artificial neural networks is primarily based upon parallel processing technique. The initial steps include selection of appropriate neural network architecture depending on the problem parameters and then tuning the said network to generate desired output. Every neural network follows certain 'learning' paradigms for training purpose using numerous examples.

Back Propagation Neural Network (BPNN) is one of the most popular neural networks used these days for classification purpose. The mostly used network topology consists of multiple layers with connections established between neurons of adjacent layers only such that flow of information remains in one direction only, such type of networks are termed as feed forward network. The layer through which data is fed into the network is termed as input layer, the layer from which processed information is retrieved is termed as output layer and the intermediate layers are called hidden layers.

Fig. 5.4 presents a detailed overview of a three layer Back Propagation Neural Network (BPNN), where x_{ij} is the connection between the i^{th} node of the input layer and j^{th} node of the hidden layer and y_{ij} is the connection between the i^{th} node of the hidden layer and j^{th} node of the output layer. The main element of a network is the processing node, which serves basic

two purposes. Firstly, it calculates the weighted sum of the input activations and then passes the summation through an arbitrary activation function to generate the output response using the equation,

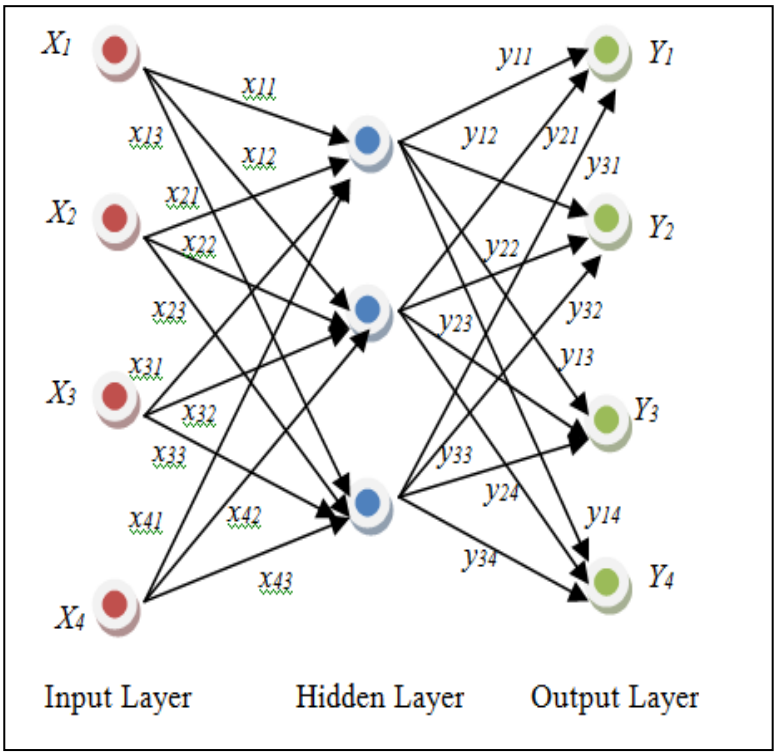


Fig. 5.4 Architecture of Back Propagation Neural Network (BPNN)

$$net_{pi} = \sum_{j} w_{ij} a_{pj} + bias_i \tag{5.15}$$

where, w_{ij} signifies the weight between current i^{th} node and previous j^{th} node, whereas a_{pj} denotes the input activation from previous node for pattern p and $bias_i$ is the bias term corresponding to the current node. Further, net_{pi} is passed through a nonlinear function to produce the output activation response a_{pi},

$$a_{pi} = \frac{1}{1+e^{-net_{pi}}} \tag{5.16}$$

A Back Propagation Neural Network is trained using numerous samples after configuring input and output pattern according to the problem requirements. The easiest approach is to assign one input node to each input channel and one output node to each of the desired class labels. As an input pattern is fed to the network the connections are modified in

order to produce the output activation response close to the desired output as much as possible. The patterns are repeatedly fed to the networks until it completely learns it. The main aim of the Back Propagation Learning algorithm is to minimize the sum of the error E between the output activity response and the desired response.

$$E = \sum_{p} \sum_{i} (t_{pi} - O_{pi})^2 \tag{5.17}$$

where, p and i denotes the patterns and output nodes respectively. The weights are adapted in a way such that above mentioned error is minimized and the network learns the pattern, so that it can recognize a similar pattern by assigning the concerned class label to it, in the testing phase.

5.3.2 Proximity Based ADE

In this section, Proximity based Adaptive Differential Evolution (PADE) algorithm is proposed. The algorithm is based on a few observations.

5.3.2.1. Mutant Vectors based on Success Rate

S. Basu Roy *et al*. [31] has used controlled mutation to search for better donor vectors. The main reasons for using more than one type of mutation vectors' choice technique within a single algorithm are that

- DE/rand/1 technique has more exploration capability since it uses random vectors from the pool of population vectors for perturbation. This nature makes this technique a useful tool to pull out algorithm from local minima with a limitation of less convergence ability.
- DE/best/1 or DE/target-to-best/1 techniques use best vectors of the generation. Thus, the algorithm is guided towards the direction of the best solution found so far. It is highly probable that the new offspring vector may be guided towards local minima with high convergence rate.

The effort towards finding a better optimal solution by creating a stand-off between exploration and exploitation is the main reason behind proposing this novel mutant vectors' selection technique. We have defined a mutation control parameter (*M*) used in the next generation depending on the success rate (*SR*) of current generation. Success rate is defined by the ratio of number of successfully generated offspring vectors which are carried to the next generation to total population. This success rate is used to find *M* according to the logarithmic rule, expressed as,

$$M = \log_2(1+SR) \tag{5.18}$$

This strategy has a strong potential for better donor vector selection than the traditional family of DE algorithm.

5.3.2.2. Proximity based Scale Factor Adaptation

We have seen the traditional use of fixed scale factor [32] or use of adaptive scale factor [31], [33] generated randomly using some statistical distribution. Without depending upon the random generation of a scale factor from any statistical distribution, we propose an approach for selecting a scale factor for each target vector *i* depending upon the success rate and the distance of the candidate vector from the best population vector, which is mathematically formulated as,

$$F_i(n) = SR.F_i(n-1) + (1-SR).(1-e^{-\|X_i - X_{best}\|}), \tag{5.19}$$

where, $F_i(n)$ is the scale factor the current generation. So in best case scenario when *SR* is 1, the old scale factors are carried over to the next generation without alternation. In the worst case, scale factor for each target vector is regenerated according to the distance formula. These changes are incorporated on the main DE frame.

5.3.3 System Overview

This section describes the basic steps of a four class motor imagery classification system.

5.3.3.1 Feature Extraction

This step mainly deals with extracting features from preprocessed EEG signals. There is no standardized feature set for EEG signal analysis, depending on the requirement of the problem, researchers usually choose a type of or a combination of features from a pool of time domain, frequency domain or time frequency correlated features. All the above mentioned feature types carry relevant information, but since classifying motor imagery data is still considered to be a challenging problem, hence it is necessary to deal with efficient data set [16]. So, we have used Common Spatial Pattern (CSP) algorithm for feature extraction purpose. CSP is a spatial filtering technique that represents the EEG signals as a linear combination over the number of channels taken into consideration. Moreover, unlike other feature sets CSP is directly applicable to multichannel data while on the other hand, to extract other feature types like wavelet coefficients from multiple electrode EEG signals, the extraction technique needs to be applied individually on each channel to serve the purpose. CSP basically aims to map the data samples in a linear subspace such that the variance is

maximized in one condition for one class and minimized for the other class, to yield more discriminate filtered signals which can be easily classified by the classifier.

5.3.3.2 Dimension Reduction

Dimension reduction is also a key step in the proposed framework, since feature extraction generates a huge dimensional feature set (here after termed as 'data points') of the order of several thousands approximately; hence it becomes very difficult to use that large dimensional data for classification purpose. So, several data point reduction techniques have been introduced in order to deal with such problems. In this chapter, we have used Principal Component Analysis (PCA) [34], to extract ideal data points for each subject corresponding to each class. The steps of PCA in this context have been briefly outlined as follows,

1) Let $A_k = \{\overrightarrow{X_1^k}, \overrightarrow{X_2^k}, ..., \overrightarrow{X_m^k}\}$ denotes the set of *m* extracted features from k^{th} subject for each class, where $\vec{X}_i^k = \{x_{i,1}^k, x_{i,2}^k, ..., x_{i,d}^k\}$ denotes a *d*- dimensional vector and $k \in [1, K]$ where *K* is the maximum number of subjects for each class. Firstly, for each data point $\overrightarrow{X}_i^k$, a mean subtracted vector $\overrightarrow{X'}_i^k$ is constructed, where $\bar{x}_i^k$ is the mean of all data points of $\vec{X}_i^k$.

$$\overrightarrow{X'}_i^k = \{x_{i,1}^k - \bar{x}_i^k, x_{i,2}^k - \bar{x}_i^k, ..., x_{i,d}^k - \bar{x}_i^k\} \tag{5.20}$$

2) A matrix D_k of $m \times d$ dimension is constructed containing the newly formed mean subtracted vectors. Next, the data covariance matrix is formed as

$$C_k = \frac{D_k . D_k^T}{d-1} \tag{5.21}$$

Finally, the first principal component $\overrightarrow{PC_k}$ is obtained by Eigen value decomposition, which is nothing but the Eigen vector corresponding to the largest Eigen value.

$$D_k = [\overrightarrow{X'}_1^k, \overrightarrow{X'}_2^k, ..., \overrightarrow{X'}_m^k] \tag{5.22}$$

3) Each member of the matrix D_k is projected along the first principal component to obtain class representative principle component $\overrightarrow{\phi_k}$ using the equation,

$$\overrightarrow{\phi_k} = \overrightarrow{PC_k}^T \times D_k \tag{5.23}$$

Thus, we obtained a reduced dimensional data points for each subject. If there are *K* subjects and *R* classes, then the entire procedure needs to be repeated $K \times R$ times.

5.3.3.3 Classification

In order to classify motor imagery signals, we have used neural network with sparse architecture. In sparse neural network, some of the neurons of each layer are not connected to the nodes or neurons of the next layer. The training phase of the network has been divided into two sub phases – weight adaptation and weight tuning.

- *Weight Adaptation.* The weights of ANN have been adapted using a proximity based adaptive differential evolution (PADE) algorithm. The dimension of the population vectors is same as the number of total weights present in the network. The vectors are changed iteratively such that the error defined in equation (5.17) is minimized. After adaptation, *15%* connections having low absolute weight values are ignored to get the sparse architecture assuming that they have low contributions towards classification.
- *Weight Tuning.* After the sparse architecture is achieved, the weights of the existing connection are gone through fine tuning by basic back propagation (BP) algorithm to increase the accuracy using the same error equation (5.17). The necessity of this step is that it put more emphasis on the existing connections assuming that they have significant contribution in the classification of MI data.

In this approach though the network has gone through the training phase twice, the less complex NN architecture is found. The pseudo code for weight adaptation using proposed PADE algorithm along with the training procedure is outlined in Table 5.3.

5.3.4 Experimental Results and Statistical Analysis

This section aims to provide detailed description of the different steps of the cognitive task based experiments that have been undertaken to check the performance of the proposed framework and validate the performance of the designed neural network classifier with other standard classifiers.

5.3.4.1 Experimental Setup and EEG Signal Acquisition

To record EEG signals, the subjects were asked to sit on a chair comfortably with hands kept in relaxed position. At the beginning of the experiment a beep sound is heard followed by a relax phase of 5 seconds, after which an arrow directed in (left, right, up or down) appears on screen to instruct the subjects to perform the specific motor imagination. The arrow stays on screen for 10 seconds followed by another relaxation phase of 5 seconds and a beep sound to indicate the end of trial.

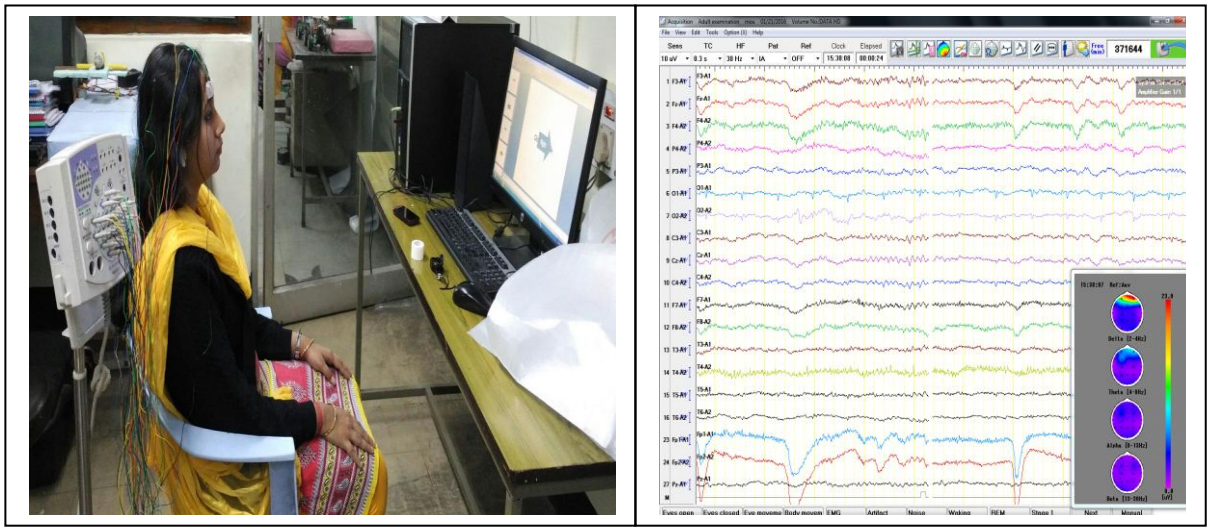


Fig.5.5 Experimental Setup

TABLE 5.3 PSEUDO ALGORITHM FOR SPARSE NETWORK WEIGHT ADAPTATION

1. Initialization:
a. Initialize NP number of weight vectors whose dimensions are equal to the total number of weight present in the neural network.
b. Initialize F = 0.8 and CR = 0.7 and Mutation factor M = 0.5 and
calculate the fitness value of equation (5.10) of each individual population vector.
c. Initialize success rate SR = 0.
2. Mutation:
If iteration>1
Choose scale factor for each target vector according to equation (5.19).
End if
if rand(0,1] > M
Choose DE/rand/1/bin
Else
Choose DE/target_to_best/1/bin for donor vector selection
End if
3. Crossover:
Recombine donor vector with the target candidate vector to generate offspring vector.
4. Selection :
Find the best vector between candidate and offspring.
If the offspring is better according to the minimum objective function value, increase the success rate by (1/NP).

5. Repeat step 1.c to 4 for all the NP vectors.

6. Choose control parameter M using equation (5.18).

7. Repeat step 1.c to 5 for a predefined number of iteration.

8. Find out the best weight vector after all the iteration completes.

9. 15% weights between each two consecutive layer with low values are ignored and connections are terminated to get sparse network.

10. Fine tune the weights of the existing connections with BP algorithm to get better accuracy

A pictorial presentation of a trial is shown in Fig. 5.6. To predict the extent of impact of artifacts generated due to eye movement, the subjects were not restricted to keep their eye movement stable while performing the imagined movements for the concerned cognitive tasks.

EEG signals are acquired using a stand alone EEG machine (manufactured by Nihon Kohden) with 19 Ag/ AgCl (*FP1, FP2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2*) electrodes placed on different scalp locations according to standard 10-20 electrode placement rule. Here A_1 and A_2 have been considered as the reference electrodes and the sampling frequency is 200Hz. Fig. 5.5. describes the entire experimental setup briefly. Fig. 5.7 presents 19 unique component scalp maps of different brain locations obtained by performing ICA, where the red colour indicates highest activation whereas the blue colour denotes lowest activation.

5.3.4.2 Frequency Band Selection

From the experimental results and elementary neurophysiological knowledge we have seen, the maximum change in the brain signals occurs in the range of 4-30 Hz. Hence in this case, to extract the signals of the above mentioned frequency band, a bandpass elliptical filter of 4^{th} order has been used.

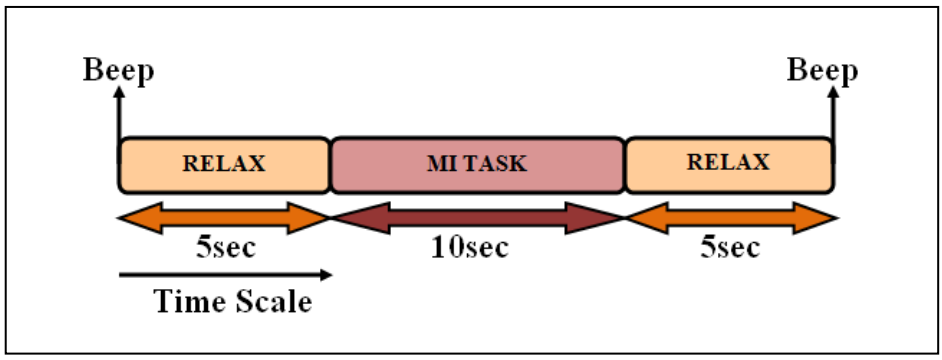


Fig. 5.6 Time Division of a Particular Trial of a Class

5.3.4.3. Experimental Results

Fig. 5.8 presents the classification accuracy of the newly designed proximity based ADE induced sparce neural network classifier and validated it with other standard classifier in terms of average classification accuracy. From the figure, it is easily noticed that subject 2 has outperformed all the subjects and the proposed classifier provides better performance even in the presence of artifacts generated due to eye movement infering that such artifacts have minimal impact on motor imagery based classification tasks. In case of all the subjects, RBF

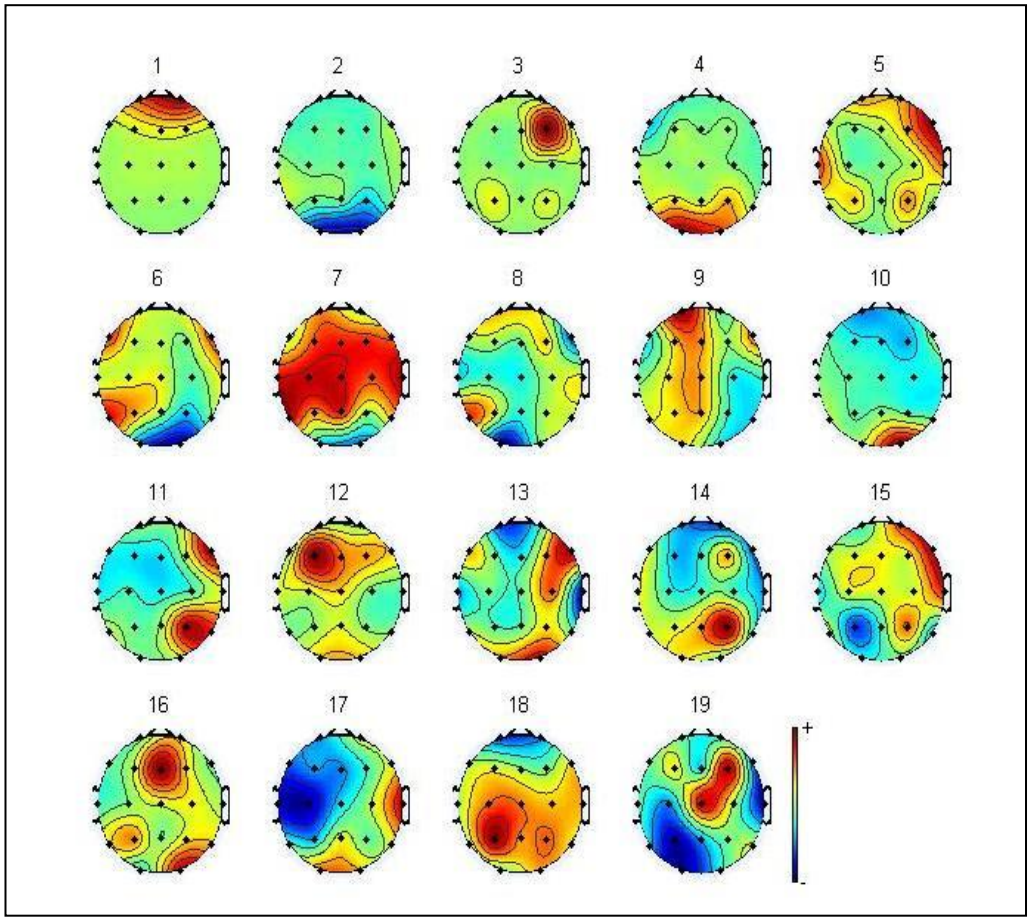


Fig. 5.7 Component wise Scalp Maps for All the Electrodes Number wise Arranged as, F3, Fz, F4, P4, P3, O1, O2, C3, Cz, C4, F7, F8, T3, T4, T5, T6, Fp1, Fp2, Pz

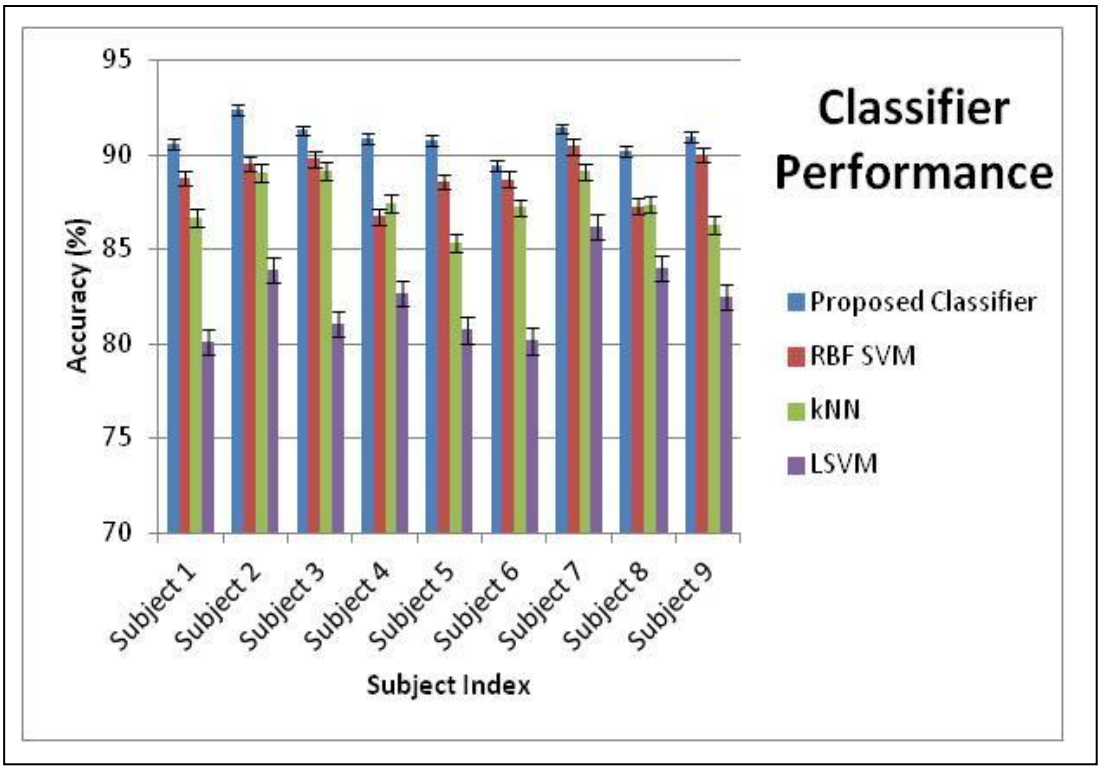


Fig. 5.8 Subjectwise Comparison of the Performance of the Proposed Classifier with Other Variants of Neural Network Classifiers in Terms of Mean Classification Accuracy

SVM has been found to be the closest competitor of the proposed classifier, and Linear SVM generated poor performance for all the subjects for this problem.

Along with classification accuracy, it is equally important to keep an eye on the misclassification rate of the proposed classifier. Hence we have included the confusion matrix in Table 5.4 corresponding to the four cognitive tasks which tabulates the readings obtained using the newly designed classifier. The large entries on the diagonal positions indicate the satisfactory performance index in all the four classes, besides the misclassification rate is also smaller in the proposed strategy.

In Table 5.5, the average classification accuracies obtained for four classes using different set of EEG features have been tabulated. In this case, we have validated CSP with other standard temporal, frequency and time frequency correlated features. From the experimental results, it is readily understood that amongst the four classes left hand (class 2) has been detected most accurately. Among all set of different features, Common Spatial Pattern (CSP) provided the best performance among the five well known features.

TABLE 5.4 CONFUSION MATRIX OF THE COGNITIVE TASKS USING PROPOSED FRAMEWORK

		Predicted Class			
		Left Hand	**Right Hand**	**Tongue**	**Foot**
Actual Class	**Left Hand**	92.278	3.125	1.562	3.037
	Right Hand	4.983	91.184	1.273	2.56
	Tongue	4.237	5.175	89.578	1.01
	Foot	4.123	5.247	0.149	90.481

In Table 5.6 the null hypothesis has been rejected, if $Z > 3.84$, where 3.84 is the critical value of the chi square distribution for 1 degree of freedom at probability of 0.05. It can be clearly seen from the results that the proposed classification algorithm outperforms the other competitors except RBF SVM.

TABLE 5.5 AVERAGE CLASSIFICATION ACCURACY OBTAINED WITH DIFFERENT EEG FEATURES FOR THE FOUR CLASSES

EEG Features	Cognitive tasks undertaken presented as classes			
	Class 1 (Right Hand)	Class 2 (Left Hand)	Class 3 (Tongue)	Class 4 (Foot)
AAR	72.56	74.78	64.57	68.92
Hjorth	78.48	79.21	70.12	69.25
PSD	80.56	82.36	75.61	76.44
DWT	82.39	84.85	78.37	80.29
CSP	89.67	92.61	83.45	86.78

5.4 CONCLUSION

This section basically infers three important findings, 1) using Common Spatial Pattern (CSP) as the feature extraction tool generates better performance than any other feature extraction technique, 2) the usage of the novel proximity based ADE induced neural network classifier has enhanced the classification accuracy by quite a great extent, 3) the artifacts generated due to eye movements while EEG data recording have negligible effect on the overall system performance, because without removing such artifacts we have acquired quite a high accuracy (92%) for each of the classes. Hence, the proposed framework should definitely be recommended for application of real time systems where high precision is required.

TABLE 5.6 STATISTICAL ANALYSIS OF CLASSIFIERS USING MCNEMAR'S TEST

Reference Algorithm: Proximity based ADE induced sparse neural network classifier				
Classifier algorithm used for comparison using desired features d=50	**Parameters used for McNemar's test**		***Z***	**Comments on Acceptance/Rejection of Hypothesis**
	n_{01}	n_{10}		
LSVM	87	120	4.946	Rejected
kNN	90	121	4.265	Rejected
RBF SVM	97	116	1.521	Accepted

REFERENCES

[1] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. M??ller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[2] G. Pfurtscheller and C. Neuper, "Motor Imagery and Direct Brain–Computer Communication," vol. 89, no. 7, pp. 1123–1134, 2001.

[3] F. Lotte and M. Congedo, "A review of classification algorithms for EEG-based brain – computer interfaces," *J Neural Eng*, vol. 4, pp. R1–R13, 2007.

[4] M. Arvaneh, S. Member, C. Guan, K. K. Ang, and C. Quek, "Optimizing the Channel Selection and Classification Accuracy in Introduction Experiments Results And Conclusion," vol. 58, no. 6, pp. 1865–1873, 2012.

[5] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *Rehabil. Eng. IEEE Trans.*, vol. 8, no. 4, pp. 441–446, 2000.

[6] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 683–686, 2009.

[7] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1991–2000, 2008.

[8] F. Lotte and C. Guan, "Spatially regularized common spatial patterns for EEG classification," *Proc. - Int. Conf. Pattern Recognit.*, pp. 3712–3715, 2010.

[9] A. Ashok, A. K. Bharathan, V. R. Soujya, and P. Nandakumar, "Tikhonov regularized spectrally weighted common spatial patterns," *2013 Int. Conf. Control Commun. Comput. ICCC 2013*, no. Iccc, pp. 315–318, 2013.

[10] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain-computer interfacing.," *J. Neural Eng.*, vol. 9, no. 2, p. 026013, 2012.

[11] F. Lotte, C. Guan, F. Lotte, C. Guan, and S. Member, "Regularizing Common Spatial Patterns to Improve BCI Designs : Unified Theory and New Algorithms Regularizing Common Spatial Patterns to Improve BCI Designs : Unified Theory and New Algorithms," 2011.

[12] N. García-Pedrajas and D. Ortiz-Boyer, "Boosting k-nearest neighbor classifier by means of input space projection," *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10570–10582, 2009.

[13] B. Zhang, "Reliable classification of vehicle types based on cascade classifier ensembles," *Intell. Transp. Syst. IEEE Trans.*, vol. 14, no. 1, pp. 322–332, 2013.

[14] V. Athitsos and S. Sclaroff, "Boosting nearest neighbor classifiers for multiclass recognition," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, 2005, p. 45.

[15] M. Derlatka and M. Bogdan, "Ensemble kNN classifiers for human gait recognition based on ground reaction forces," in *Human System Interactions (HSI), 2015 8th International Conference on*, 2015, pp. 88–93.

[16] S. Bhattacharyya, A. Konar, D. N. Tibarewala, A. Khasnobish, and R. Janarthanan, "Performance analysis of ensemble methods for multi-class classification of motor imagery EEG signal," in *Control, Instrumentation, Energy and Communication (CIEC), 2014 International Conference on*, 2014, pp. 712–716.

[17] L. I. Kuncheva, J. J. Rodríguez, C. O. Plumpton, D. E. J. Linden, and S. J. Johnston, "Random subspace ensembles for fMRI classification," *Med. Imaging, IEEE Trans.*, vol. 29, no. 2, pp. 531–542, 2010.

[18] J. Mellin-Olsen, S. Staender, D. K. Whitaker, and A. F. Smith, "The Helsinki declaration on patient safety in anaesthesiology," *Eur. J. Anaesthesiol.*, vol. 27, no. 7, pp. 592–597, 2010.

[19] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Comput.*, vol. 10, no. 7, pp. 1895–1923, 1998.

[20] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *Biomed. Eng. IEEE Trans.*, vol. 51, no. 6, pp. 993–1002, 2004.

[21] G. Townsend, B. Graimann, and G. Pfurtscheller, "Continuous EEG classification during motor imagery-simulation of an asynchronous BCI," *Neural Syst. Rehabil. Eng. IEEE Trans.*, vol. 12, no. 2, pp. 258–265, 2004.

[22] R. Vigário, J. Särelä, V. Jousmiki, M. Hämäläinen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings," *Biomed. Eng. IEEE Trans.*, vol. 47, no. 5, pp. 589–593, 2000.

[23] F. Popescu, S. Fazli, Y. Badower, B. Blankertz, and K.-R. Müller, "Single trial classification of motor imagination using 6 dry EEG electrodes," *PLoS One*, vol. 2, no. 7, p. e637, 2007.

[24] T. N. Lal, M. Schröder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Schölkopf, "Support vector channel selection in BCI," *Biomed. Eng. IEEE Trans.*, vol. 51, no. 6, pp. 1003–1010, 2004.

[25] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2006, pp. 5392–5395.

[26] P. D. Heermann and N. Khazenie, "Classification of multispectral remote sensing data using a back-propagation neural network," *Geosci. Remote Sensing, IEEE Trans.*, vol. 30, no. 1, pp.

81–88, 1992.

[27] J. D. Paola and R. A. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification," *Geosci. Remote Sensing, IEEE Trans.*, vol. 33, no. 4, pp. 981–996, 1995.

[28] A. K. Qin and P. N. Suganthan, "Self-adaptive differential evolution algorithm for numerical optimization," in *Evolutionary Computation, 2005. The 2005 IEEE Congress on*, 2005, vol. 2, pp. 1785–1791.

[29] C. Brunner, M. Naeem, R. Leeb, B. Graimann, and G. Pfurtscheller, "Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis," *Pattern Recognit. Lett.*, vol. 28, no. 8, pp. 957–964, 2007.

[30] A. Saha, A. Konar, P. Das, B. Sen Bhattacharya, and A. K. Nagar, "Data-point and feature selection of motor imagery EEG signals for neural classification of cognitive tasks in car-driving," in *Neural Networks (IJCNN), 2015 International Joint Conference on*, 2015, pp. 1–8.

[31] S. B. Roy, M. Dan, and P. Mitra, "Improving adaptive differential evolution with controlled mutation strategy," in *Swarm, Evolutionary, and Memetic Computing*, Springer, 2012, pp. 636–643.

[32] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[33] J. Zhang and A. C. Sanderson, "JADE: adaptive differential evolution with optional external archive," *Evol. Comput. IEEE Trans.*, vol. 13, no. 5, pp. 945–958, 2009.

[34] A. Saha, A. Konar, A. Chatterjee, A. Ralescu, and A. K. Nagar, "EEG analysis for olfactory perceptual-ability measurement using a recurrent neural classifier," *Human-Machine Syst. IEEE Trans.*, vol. 44, no. 6, pp. 717–730, 2014.

Chapter 6

Conclusion

This chapter revisits the primary goal of the thesis and briefly highlights the findings and contributions of the work to substantiate the extent to which the objective of the thesis is accomplished. Following this, the future course of the work is discussed which are open for research to interested readers for the extension of the algorithms that have been developed as a part of this thesis.

6.1 SUMMARY OF THE WORK

The thesis has provided interesting insights into different aspects of developing modern day HCI based system. The research is primarily concentrated in exploring different computationally intelligent techniques that can be utilized judiciously to build more efficient modalities of HCI. Throughout the thesis different principles of CI discipline have been studied and analysed to find new avenues of innovative applications. While traditional pattern recognition and optimization algorithms are studied for some applications, novel algorithms and methodologies are also proposed supported by appropriate experiments and results.

In today's world research enthusiasts are extremely engrossed in exploring divergent applications of EEG signals including medical diagnosis, rehabilitative application or brain rhythm detection. The concept of utilizing the postulates of CI in EEG based BCI systems has fascinated the researchers since a long time. The present thesis enlightens about certain improvised applications of the concerned research genre which can help in enhancing system performance. A wide variety of EEG signal based features have been contemplated for different applications.

Chapter 1 provides the background for the current work by generating a rough sketch about the motivations behind each of the chapters and the possible outcomes of the research. Firstly, it starts with a brief introduction and gradual evolution of the BCI research and mentions the importance of CI discipline along with its functional definition. Next, the need of brain localization is discussed and by delving deeper into the research domain the thesis provides a tabulated list of the related trends of contemporary research that is carried out in the same discipline. After that, various brain signal measuring modalities have been surveyed and after conducting a detailed research the superiority of EEG signals has been established due to its numerous advantages over the other standard techniques. To study EEG signals it is important to know about the functionality of human brain which is addressed in the later section. Finally, amongst different EEG modalities Motor Imagery has been illustrated along with the highlighted research works for detection of MI signals.

Chapter 2 basically serves as a handbook for the remaining parts of the thesis. The next few chapters address different improvised techniques for the successful implementation of the CI postulates in EEG based BCI paradigms. To test the efficacy of the proposed scheme, it is always recommended to compare the performance outcomes with the existing best techniques. Due to space constraint it is not possible to describe each component of the existing methodologies in detail, but to understand the advantage of the proposed scheme over others it is required to know the attributes of both the schemes very well. For this reason

Chapter 2 presents the standard tools and techniques that are used later in the thesis, mainly the feature extraction methods and classification algorithms have been described thoroughly.

Chapter 3 emphasizes on the impact of optimization in EEG based BCI system. EEG signal acquisition devices basically record brain activities through Ag/AgCl electrodes placed throughout the human scalp. From elementary neurophysiological knowledge it can be known that for specific cognitive tasks distinct regions of human brain are activated. So for particular mental state classification, it is wise to deal with relevant EEG electrodes and thus reducing the computational overhead and enhancing the accuracy as well. Similarly selection of relevant features can also be justified citing same reasons. Now, separate selection of electrodes and features is possible, but that increases the system complexity and increases the probability of carrying redundant information or incomplete information. Hence, it is always wise to select the optimal EEG features and electrodes simultaneously to deal with these issues. This chapter presents the evolutionary perspective of optimal EEG feature and electrode selection by proposing a novel variant of Firefly Algorithm termed as Self Adaptive Firefly Algorithm to optimize an objective function formulated while satisfying the constraints with SVM playing the role of the classifier. The selection of ABC for this problem is its fast convergence without the possibilities of local minima.

Chapter 4 is focussed on the extension of existing EEG feature extraction techniques in spatial domain. Due to the non-stationary behaviour of EEG signals, the spatial resolution is very poor. To address this issue spatial filtering is introduced as an alternative way to spatially represent EEG signals. The initially introduced spatial filters suffer from the drawback of overfitting and lack of robustness. Finally, Common Spatial Patterns (CSP) was introduced to overcome the limitations of basic spatial filters and CSP produced revolutionary results motor imagery detection problem. Initially, CSP was developed to detect abnormalities in EEG signal and later it was utilized as a feature extracting tool with high performance. As every other invention, gradually the drawbacks of CSP became more and more pronounced leading to the ways of regularizing CSP and thus RCSP algorithms were introduced. After exploring the details of RCSP algorithm, improvised methodologies are suggested with applications of those techniques in MI detection described in Chapter 5.

Chapter 5 is an extended portion of Chapter 4 with the highlighted applications of the methodologies proposed in the previous chapter in MI detection problems. It puts stress on two distinct cases of MI detection and in each case it emphasizes on different phases of the pattern recognition chain usually adopted for EEG based BCI applications. The first section proposes an RCSP based novel feature extraction strategy with a high classification accuracy obtained with an ensemble of k-NN classifiers. Unlike the previous section, the later section deals with enhancing the accuracy of the classifying framework adopted by incorporating Adaptive Differential Evolution in traditional Back Propagation network.

The thesis has addressed the preliminary issues for the mentioned objective. However, there is a long way to traverse before the complete system can be presented as a stand-alone unit for practical applications. The major challenges that can be addressed and combated are addressed in the next section and form an open area of research for the extension of this work.

6.2 FUTURE DIRECTIONS

- Instead of adopting a single objective approach a multi objective evolutionary algorithm sounds to be a more justified alternative in this case. More importantly, as the experiments conducted indicate binary problems so utilizing a standard genetic algorithm or heuristics based Ant Colony Optimization could have make more sense.
- The problem discussed in Chapter 3 resides in discrete search space, so it is wise to use Genetic Algorithms or other alternatives that deal with the problem in discrete search space only. But using the discussed approach basically transformed it into a real coded problem that is an inverse genotype-phenotype mapping has taken place which could have resulted in Hamming cliffs and thus making the optimization problem harder than the original one.
- All the contender algorithms considered are not well suited for black box optimization with such objective functions; instead state of the art evolutionary algorithms should have been used for real coded optimization purpose.
- In Chapter 5 instead of utilizing both the linear transforms CSP and PCA, alternative efficient measures should be preferred which singlehandedly serves the purpose.
- The channel similarity measures can be modified by introducing fuzziness into them for extending it into a fuzzy based similarity metric.
- One of the biggest disadvantages of EEG is its poor spatial resolution and hence source localization is not easy. On the other hand, devices like fNIRS have better spatial resolution and are also available at low cost. Thus, fNIRS and EEG signals can be coupled by means of some data fusion techniques to explore the advantages of both i.e. to achieve high temporal resolution as well as good spatial resolution.

Appendix A

Statistical Methods Used

This appendix describes different statistical methods adopted throughout the thesis and also illustrates about different statistical tests that have been conducted to validate the system performance.

A.1 CLASSIFICATION METRICS

This section primarily describes the confusion matrix used to measure the efficiency of a classifier and derived results from it.

A.1.1 Confusion Matrix and Classification Accuracy

A confusion matrix is a square matrix that signifies the association between the user intended classes termed as "Actual Classes" and the original classifier estimated classes termed as "Predicted Classes". Any element *C(i, j)* of a confusion matrix *C* basically denotes the number of samples belonging to class *i* that have been predicted as to belong to class *j* by the classifier. It is needless to say the diagonal elements represent the number of samples those are correctly classified and the off diagonal elements provide a measure of incorrect classification. A general *N×N* confusion matrix and a *2×2* confusion matrix have been shown in Figure A.1. The overall classification accuracy *CA* can be calculated from a *N×N* matrix for a *N* class problem using (1) and the one vs all classification accuracy for a particular class *i* can be obtained as *CA(i)* using (2).

$$CA = \frac{\sum_{i} C(i,i)}{\sum_{i,j} C(i,\ j)} \tag{A.1}$$

$$CA(i) = \frac{C(i,i)}{\sum_{j} C(i,\ j)} \tag{A.2}$$

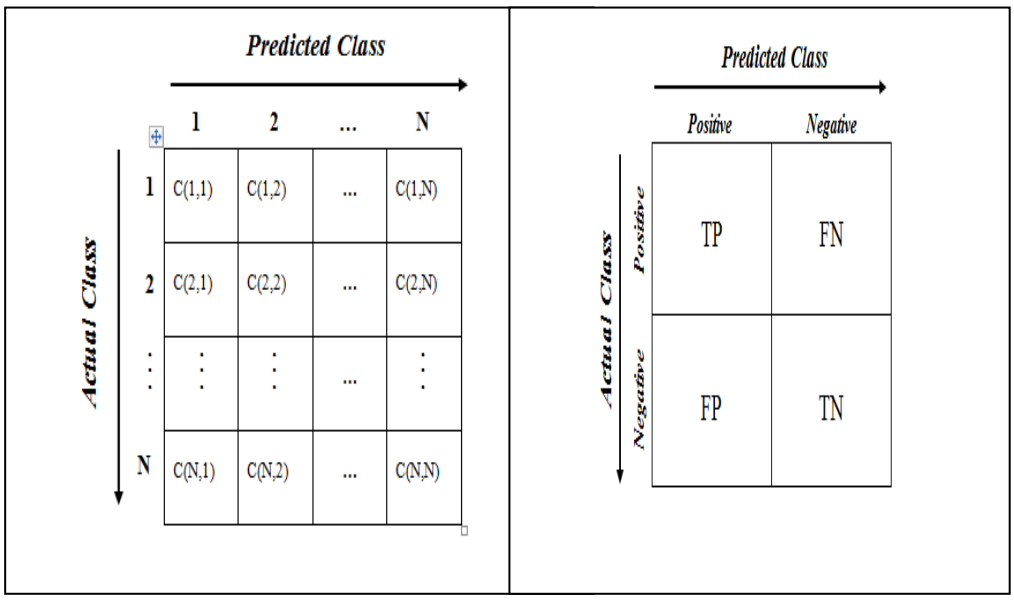


Fig. A.1(a) An N×N Confusion Matrix, (b) 2×2 Confusion Matrix

A.1.2 Type I and Type II Error

The Type I (TE_1) and Type 2 (TE_2) for class *i* are computed using (3) and (4). TE_1 denotes the false positive error rate that is the number of samples that donot belong to a class but are misclassified to belong to the class. Contrarily, TE_2 denotes the false negative error rate that is the number of samples those actually belongs to a specific class but mistakenly calssified to belong to certain another class. In both the cases, ideally these two error should be zero.

$$TE_1(i) = \left(\frac{\sum_{i,i\neq j} C(i,j)}{\sum_{i} C(i,j)}\right) \tag{A.3}$$

$$TE_2(i) = \left(\frac{\sum_{j,i\neq j} C(i,j)}{\sum_{j} C(i,j)}\right) \tag{A.4}$$

A.2 STATISTICAL TEST

In this section, McNemar's test has been illustrated for detection of the performance of two classification algorithms for correctly classifying data samples. Suppose, using a common input phylogenetic sequence, the outputs generated by the algorithm A and algorithm B be f_A and f_B respectively. A null hypothesis is stated as follows,

$$P_{rR,x}[f_A(x) = f(x)] = P_{rR,x}[f_B(x) = f(x)] \tag{A.5}$$

where $f(x)$ be the experimentally induced sign in order to map a specific point *x* onto sign classes K, such that $f(x)$ is one of the *K*=4 classes. Let n_{01} be the number of data samples misclassified by algorithm A but not by algorithm B and n_{10} be the number of data samples misclassified by algorithm B but not by algorithm A. Then, a statistic termed as Z score is defined by using the following equation,

$$Z = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}} \tag{A.6}$$

At the end of test, depending on the obtained Z scores it is decided whether the null hypothesis is accepted or the alternative hypothesis are rejected. Here *Z* has been evaluated in order to denote a comparator statistic of misclassification between the reference algorithm (Algorithm A) and any of the above mentioned competitor algorithms (Algorithm B).

Appendix B

User Guide to Run the Source Codes

This appendix provides a step by step guide to run the source codes from MATLAB to execute different works described throughout the thesis.

B.1 Associated Software Installation

- Install MatlabR2012b software for all computations.
- EEGLAB (eeglab 13.5.4b) software has been used for EEG analysis carried out throughout the thesis for deriving scalp source/sink components, activation regions in brain, ERD/ERS etc. To use this software EEGLAB path needs to be added to Matlab path. Apart from the data files, it is necessary to have additional event files containing electrode locations(.locs), sampling frequency, class labels, duration of stimulus of time stamps and so on.

B.2 Preprocessing and Feature Extraction

- The *Feature.m* file carries out the entire feature extraction task and another function named *Filtering.m* performs the filtering task. For using the EEG data acquired from different subjects, the recordings must be transformed into a proper format no. of channels×no. of time samples. The data is accompanied with necessary event files stored in associated *Target* folder. In this case, the code is written for19 channels, if fewer channels or Emotiv has to be used the channel information needs to be organized properly to get the desired output.
- Filtering has been executed by a tenth order butterworth filter, followed by Common Average Referencing (CAR). The filtered signal outputs can be easily obtained by running the codes after modifying the necessary parameters as per requirement.

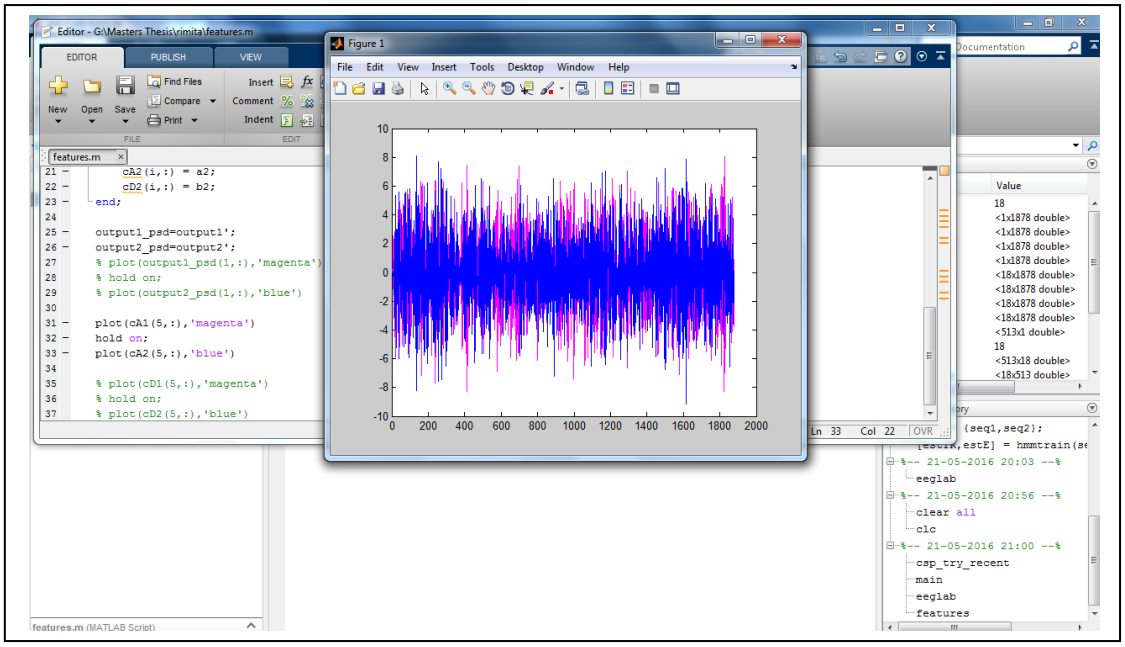


Fig. B.1 Feature Extraction (DWT)

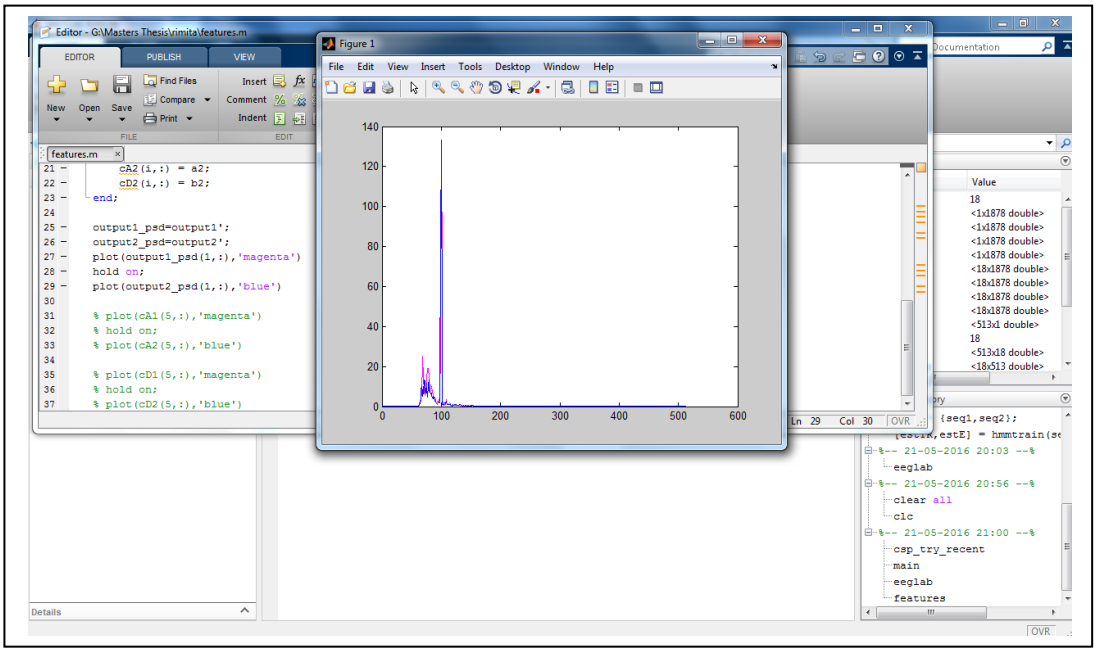


Fig. B.2 Feature Extraction (PSD)

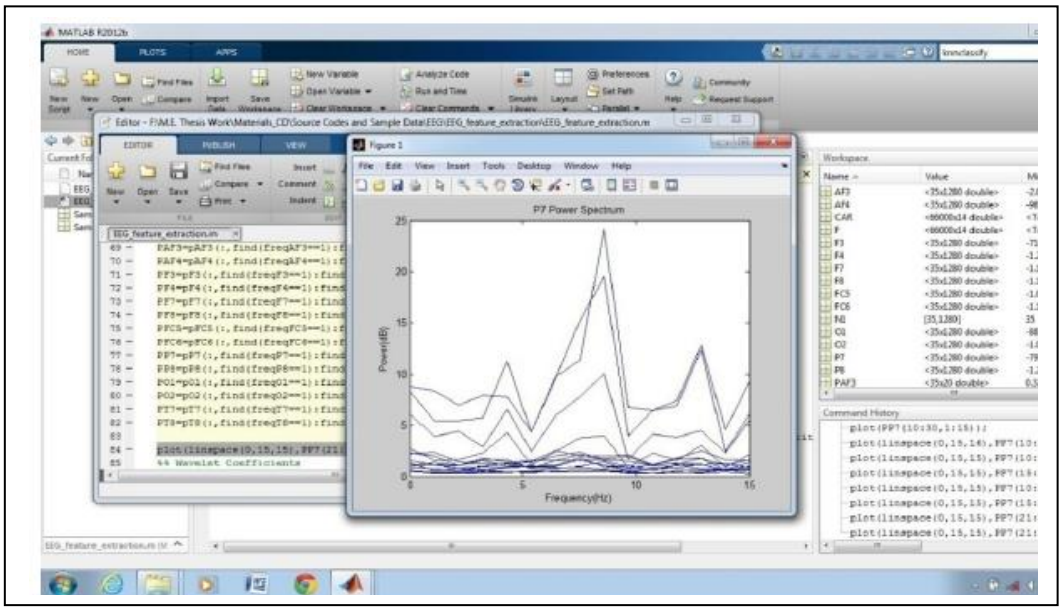


Fig. B.3 Feature Extraction (PSD)

- The features that are extracted include Adaptive Autoregressive Parameter, Hjorth parameter, Power Spectral Density (Fig. B.3), CSP (Fig. B.4) and Discrete Wavelet Transform (Fig. B.1).
- For AAR the update coefficient and the order can be changed as per requirement. Similarly, for PSD, the frequency range, the width of the Hamming window and percentage overlap is chosen by the user as per requirement. For CSP, the number of lowest and largest Eigen vectors is also determined by the user only. For DWT, the

mother wavelet and the order of decomposition need to be modified as per desired frequency range.

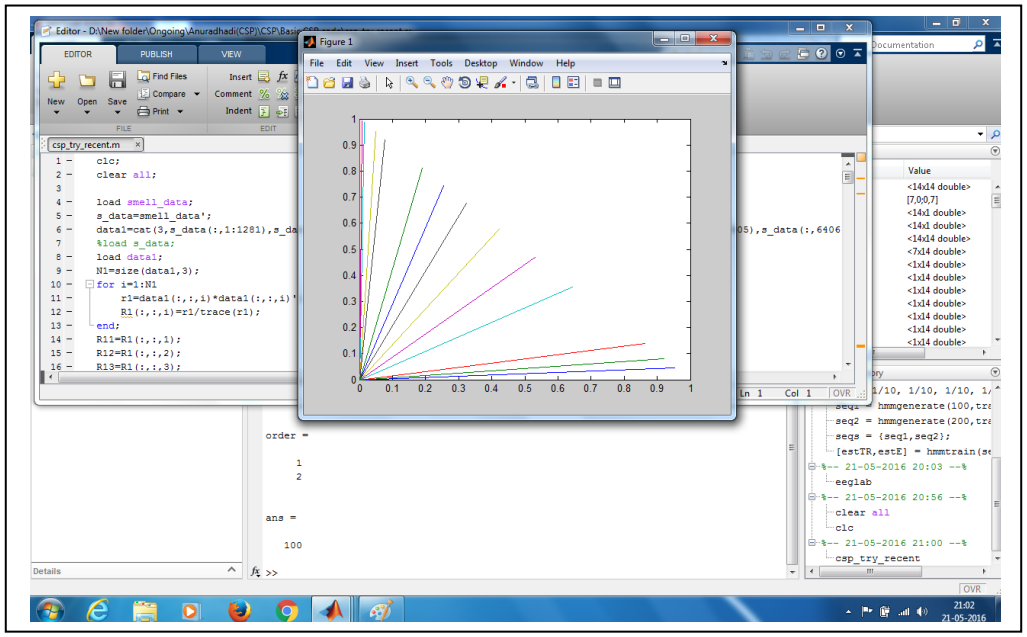


Fig. B.4 Feature Extraction (CSP)

B.3 CLASSIFICATION

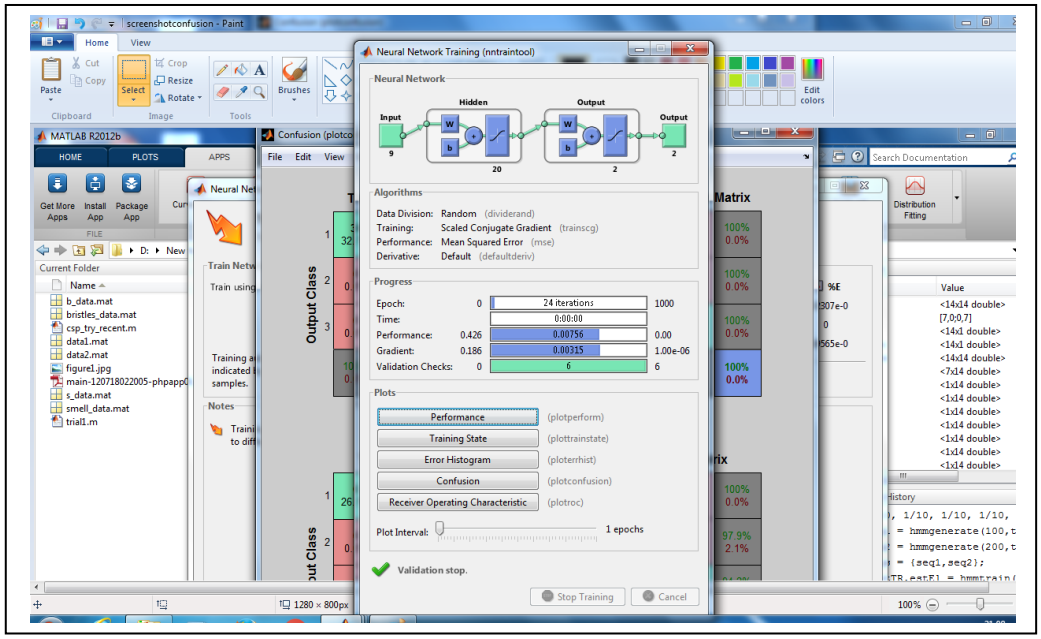


Fig. B.5 Classification (BPNN)

- The Classifiers.*m* file contains the code of all the classifiers including SVM, k-NN, LDA, Naïve Bayes and Neural Network (Fig. B.5). To select any one classifier the function call lines of the other classifiers need to be commented off. If there are no

different datasets for Training and Testing, then separate chunks of available data must be assigned to the folders Train1 and Test1, using cross validation. Another file containing the class labels according to the training samples in proper format is required to be fed to the classifier for proper training.

Fig. B.6 Classification (Confusion Matrix)

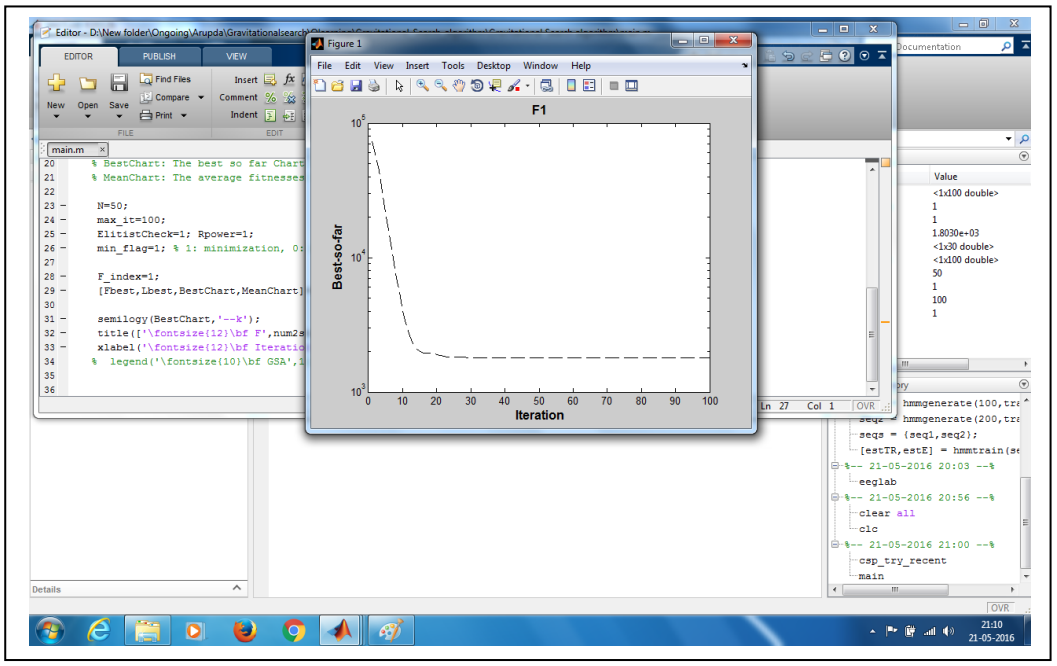


Fig. B.7 Optimum Electrode Selection (Firefly Algorithm)

- The file *Feat_select.m* employs PCA that transform the input features to the desired dimension of output before feeding to the classification step.

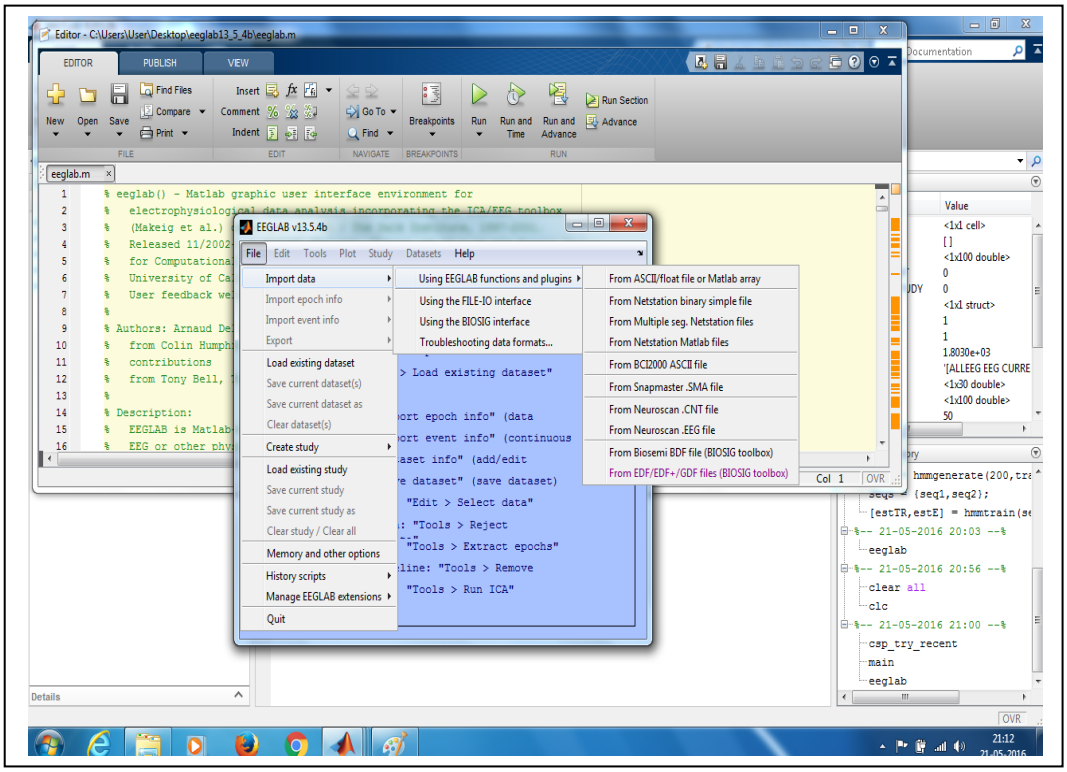


Fig. B.8 Electrode Selection (EEGLAB)

- In these works, confusion matrix (Fig. B.6) and classification accuracies have been tabulated in most of the cases. But the sensitivity, specificity and ROC curves can also be obtained.

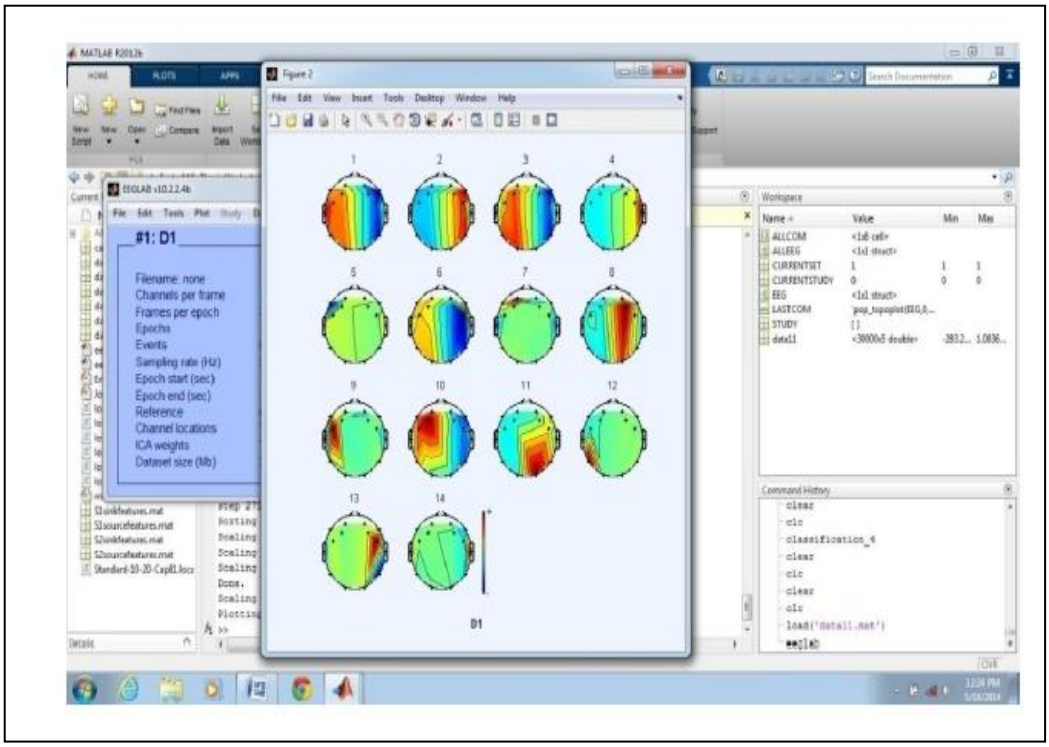


Fig. B.9 Electrode Selection (Scalp Maps)

B.4 OPTIMAL EEG ELECTRODE AND FEATURE SELECTION

- The *Optimal_Loc.m* file needs to be executed along with loading the required datasets. The objective function values with respect to function evaluations can be plotted after running the function *Fire_opti.m*, as shown in (Fig. B.7).
- From the selected scalp locations the corresponding source activations within the brain can be viewed. EEGLAB has to be initiated by running *eeglab.m* in the EEGLAB folder. The data and the channel location file have to be loaded into eeglab (Fig. B.8). Then ICA has to be run to find the source components corresponding to the channel components provided. One such sample is illustrated in Fig. B.9.