

nature of the proposed framework, recognize three different scenarios, and describe how the CF/DPF handles each of them. For the worse-case scenario with the fusion filters lagging the local filters exponentially, I derive a modified-fusion filter algorithm that limits the lag to an affordable delay.

- In Chapter 5, I derive distributed expressions for computing the PCRLB for an AN/SN configured in a distributed topology referred to as the dPCRLB. I consider both full-order and reduced-order distributed estimation problems and derive algorithms for computing the dPCRLB for each case.
- In Chapter 6, I consider distributed sensor selection problem where I propose dPCRLB-based algorithms for dynamically selecting a subset of sensors.
- Chapter 7 presents an outline of the experimental methodology used to evaluate the proposed distributed estimation approaches.
- Chapter 8 concludes the thesis and provides some directions for future work.

## **Chapter 2**

# **Background and Literature Review on Brain Computer Interfacing**

### **2.1 Brain Computer Interfaces: Why and How?**

How do people with severe motor disabilities and/or speech problems manage to perform the activities of their daily lives? You may have seen someone using a set of push buttons on a computer or tablet that speaks for him/her, or those who use specialized physical devices, such as a wheelchair or a robotic arm, to help them move around and do accomplish the tasks of their day. These devices are known as Augmentative and Alternative Communication technologies (AAC), aiding those who suffer from motor disabilities or severe speech problems to improve their quality of daily life, possibly without requiring a caregiver. Despite the success and public acceptance of the conventional AAC, there have been people with certain need and requirements, due to their severe condition, who are not able to take advantage of these means of communication. For instance, those who are totally paralyzed, or “locked-in”, are restricted from both verbal and non-verbal communication, even though they are conscious and alert [1]. The inability of communicating, neither emotions and thoughts nor physical needs of one, calls for a technology capable of a deeper level of communication and of reaching out to the thoughts of the impaired users [2]. This is essentially the origin of recent upsurge in the field of Brain-Computer Interfaces (BCIs).

### **2.1.1 Introduction to the BCIs**

The BCI systems, whilst requiring no peripheral muscular activity, enable a user to use solely his/her brain activities to send commands to an electrical device. BCI can be considered as a system for which the input is the brain activity and the output is a set of device control signals, therefore, the BCI system itself functions as a translator, measuring specific features of the brain signals. Jacques Vidal [3] was the first researcher who proposed the term “brain-computer interface” in 1973, when he presented a system that could interpret brain signals into computer control signals. BCI technology initially used to be unattractive for serious scientific investigation due to false assumptions about its applications. General public, as well as academia, often used to reject the idea of successfully deciphering thoughts or intentions by means of brain activity in the past, as strange and remote. Hence, investigation in the field of brain activity has usually been limited to the analysis of neurological disorders in a clinical setting or to the exploration of brain functions in laboratories. In contrary, during the past two decades, experimental research into BCIs has expanded significantly, with promising results presented for healthy people and few successful and practical controlled clinical outcome studies for patients. BCIs are starting to prove their efficacy as assistive and rehabilitative technologies in patients who suffer from severe motor impairments. Moreover, recently, several fruitful developments and expansions of its market for both healthy and unhealthy people have emerged. This sparked progress is driven by the advancements in effectiveness and increase in the number of available technologies to record and process brain signals.

The BCI systems generally share the same principles, i.e. the detected brain signals are amplified and recorded, then filtered, smoothed, and classified according to relevant characteristics (e.g., sensorimotor rhythms over the motor cortex). After processing and decoding of the brain signals, the output of the BCI can be used to control movement of a prosthesis, orthosis, wheelchair, robot or cursor, or to direct electrical stimulation of muscles [4]. Prior to naming the potential/active practical areas of the BCI technology, it is of great importance to clarify the difference between a tool, in this case a BCI, and an application. A tool in the present context is a device which is specified by the manner in which it performs its function, and is applicable to a wide variety of applications. A tool’s performance is evaluated by its effectiveness and ease of use, and its function remains the

same, regardless of the purpose it is serving. On the other hand, an application is primarily described by the purpose it serves, and its evaluation focuses on how well it fulfills serving the target purpose, while it may also be described in terms of the tool it employs.

Having the above definitions in mind, it can be said that BCIs are tools that record and analyze brain activities, such as, Electroencephalography (EEG) signals, which will be discussed later in details in Subsection 2.2. Moving a cursor, selection among two or more possible choices shown to the subject of the study, or controling a robot, are some examples of BCIs employed as suitable and responsive tools to use. BCI applications are widely spread in various fields of research and medical industry. Geberally speaking, BCI applications can be classified into the following six main categories:

- |  |  |
|--|--|
| (1) <i>Medical</i> ;                             | (2) <i>Games and Entertainment</i> ;               |
| (3) <i>Educational and Self-Regulation</i> ;     | (4) <i>Neuroergonomics and Smart Environment</i> ; |
| (5) <i>Neuromarketing and Advertising, and</i> ; | (6) <i>Security and Authentication</i> .           |

These applications are mostly in experimental research state and not all of them have been well-established to be operable by general public. While working on either of these areas, the primary concern of BCI developers must be the needs and priorities of the anticipated user, and researchers must guard against the tendency to approach the parameters of the tools and their applications as an abstract design exercise. BCI development requiers optimized design with well-defined objectives, which should be based on not only a complete technical study, but also a thorough and comprehensive bahvioral analysis that essentially addresses the needs, desires, and incentives of the users and their possible caregivers. Satisfying these conditions requires collaborative interactions with the users, who must know how to use the technology and be persuaded that the technology is both useful and safe. Also, essential is the cooperation of the relevant health care professionals, who must be persuaded that the risk/benefit ratio is favorable, that the technology is safe and useful, and that it is equal or superior to available alternatives.

In regards to acquisition of brain activity and the modalities to do so, BCIs can be classified into the following two main categories:

- **Invasive BCIs**

Invasive BCIs are those that involve surgical implantation of electrodes, or multi-electrode grids in the brain. These systems are intended to measure patterns of neurons' activities in order to enable the researchers to decode behaviourally relevant information from the acquired data. The reason for such a risky and expensive intervention is to gain high signal-to-noise ratio (SNR) electrical responses, recorded directly from the brain. This is in contrast to the signals recorded from over the scalp, which are usually contaminated by high amount of noise, and these recordings are of low amplitude due to the nature of human scalp, resulting in low SNR. However, there is not enough evidence to justify this brutal operation, and moreover, extensive work on brain plasticity [5] has shown that a plastic change in the adult nervous system through learning is possible, if the respective neuronal circuit participates functionally in the physiological tasks of that circuit. This implies that even if the advantages of invasive BCI would outweigh its disadvantages, since the function map of the brain is subject to change in response to the learning processes the person is exposed to, this approach will not remain favorable in the long run.

In regards to implementation of Invasive BCIs, there are a few, yet of utmost importance concerns that must be addressed prior to any practical experiment.

- (1) *Possible locations of implanted electrodes, number of electrodes to implant, and the nature of the signals to record:* In most BCI applications and the majority of the cases in which BCI is used as a tool, the motor cortex of the brain is an obvious choice for recording the signals; reasons being direct relevance to motory tasks and the relatively better accessibility compared to other motor areas of the brain. In order to determine and identify the appropriate locations for implantation, functional Magnetic Resonance Imaging (fMRI), Magnetoencephalography (MEG), and other functional imaging techniques are admittedly helpful [18]. Moreover, the number of electrodes to implant has direct relation with the location of the electrodes, the minimum SNR required for the study, and the functional use of the signals, i.e., the purpose of the study/application and the rate of information transfer to fulfill the specifications of the procedure.

- (2) *User groups who might be best suited for implanted electrodes and the stability of the recordings:* Due to the severity operation required, it makes perfect sense to only proceed with invasive BCIs for patients with extreme conditions and needs. Patients who are locked-in might benefit from invasive BCI technology if it is relatively safe and effective. Selected individuals with stroke, spinal cord injury, limb prostheses and other conditions might also benefit. It is of utmost importance to keep in mind that the individuals' preferences play a key role in specifications and decisions about implantable systems, in other words, a substantial functional advantage over the conventional non-invasive systems must be proved to justify the implantation of the invasive electrodes. Extensive researches on non-human subjects [6] has shown that stable recording has been maintained over months, and in selected instances over years. Positive results regarding human subjects has also been observed sparsely in different research groups around the world, however, this field is still in its infancy, requiring more research and proven reliable outcomes. Also, prior to any implantation, tissue acceptance of the microelectrode has to be ensured.
- (3) *The ethical issues that must be considered in implanting recording electrodes in human volunteers:* It is perfectly clear that the patients must be informed of the risks and potential dangers of the operation. Also, all the potential benefits of the system need to be clearly and forcefully explained to the volunteers, especially because volunteers with severe conditions tend to overestimate the benefits of the BCI, and they must be aware of exact aid and service they will be provided with. An ethicist should be involved in the earliest phases of any human research developing or testing invasive BCI methods.

Invasive BCIs are not the focus of this thesis, however, it is worth naming the following five main types of brain activity that can be measured with invasive BCIs: (i) Local field potentials (LFPs) [7]; (ii) Single-unit activity (SUA) [8]; (iii) Multi-unit activity (MUA) [9]; (iv) Electrocorticographic oscillations recorded from electrodes on the cortical surface (electrocorticography, ECoG) [10]; and, (v) Calcium channel permeability [11].

- **Non-Invasive BCIs**

Noninvasive BCIs are implemented without any sort of surgical implantation, as they enable recording the brain signals from the external surface of the scalp. These systems are the most widely researched BCIs due to their minimal risk and the relative convenience of conducting studies and recruiting volunteers to participate in the study. Noninvasive interfaces are able to detect seven types of brain signals.

(1) *Slow cortical potentials*

Slow cortical potentials (SCPs) are shifts in the cortical electrical activity lasting from several hundred milliseconds to several seconds. These shifts might be initiated and triggered by an external event, or induced by self. Their moderating impact on information processing has been demonstrated in numerous studies, such as in [12].

(2) *P300 Event Related Potential:*

The P300 (P3) wave is an Event Related Potential (ERP) component elicited in the process of decision making. These waves' occurrence do not link to the physical attributes of a stimulus, but to a person's reaction to it. In other words, the P300 is known to reflect processes involved in stimulus evaluation or categorization. When recorded by EEG, it surfaces as a positive deflection in voltage with a latency (delay between stimulus and response) of roughly 250 to 500 ms.

(3) *Steady-State Visual Evoked Potentials:*

The Steady-State Visual Evoked Potentials (SSVEPs) are natural responses of the brain to visual stimulation at specific frequencies. The brain starts generating electrical signals at the frequencies ranging from 3.5 Hz to 75 Hz, or multiples of them, when the retina is excited by a visual stimulus at the same frequencies. SSVEPs are useful in research because of the excellent signal-to-noise ratio and relative immunity to artifacts [13].

(4) *Error-related Negative Evoked Potentials:*

The Error-related Negative (ERN) is a sharp negative going signal which begins about the same time an incorrect motor response begins, (response locked event-related potential), and typically peaks from 80-150 milliseconds (ms) after the erroneous response begins (or 40-80 ms after the onset of electromyographic activity). The ERN is observed

after errors are committed during various choice tasks, even when the participant is not explicitly aware of making the error.

(5) *Blood-oxygen-level Dependent Contrast Imaging:*

The firing of neurons causes a need for more energy to be provided quickly. Through a process called the hemodynamic response, blood releases oxygen to the active neurons at a greater rate than to inactive ones. This causes a change of the relative levels of oxyhemoglobin and deoxyhemoglobin (oxygenated or deoxygenated blood) that can be detected on the basis of their magnetic properties. This event can be measured by fMRI method.

(6) *Cerebral Oxygenation Changes:*

The Near-Infrared Spectroscopy (NIRS), as a methodology for functional neuroimaging, is based on the fact that unlike visible light, near-infrared light (wavelength from 700 to 1000 nm) easily passes through biological tissues and is mainly absorbed by few chromophores like hemoglobin with different absorption spectra for their oxygenation/deoxygenation states. The activation of brain regions causes an increased oxygen metabolic rate and to an initial deoxygenation of the tissue which is followed by increased regional cerebral blood flow (rCBF). These metabolic changes enable researchers to study emotional and cognitive tasks of brain via measurements of functional NIRS [14].

(7) *Sensorimotor rhythms*

A Sensorimotor Rhythm (SMR) is a brain wave, oscillatory idle rhythm of synchronized electric brain activity. These brain waves appear in the recordings over the sensorimotor cortex via modalities such as EEG. For most individuals, the frequency of the SMR is in the range of 13 to 15 Hz, in cortical regions outside of the motor strip. These frequencies relate to relaxed attention such as reading or engaging in a relaxing hobby such as knitting.

Out of all the aforementioned types of noninvasive BCI signals, the focus of this thesis is on *Sensorimotor Rhythms*. Such BCIs have been used with relatively satisfactory success, for instance, Wolpaw and McFarland published the results of their study in 2004 [17], in

which they showed that patients with locked-in syndrome or high spinal cord lesions were able to use sensorimotor rhythms to control cursor movements or select letters or words from a computer menu. EEG-based BCIs with the focus on sensorimotor rhythms will be fully discussed throughout the remainder of the thesis. This completes an outline of invasive and noninvasive BCIs.

As previously stated, BCIs fall into the category of communication and control systems and therefore, a BCI has an input, an output, and a translation algorithm that converts the former to the latter. BCI input consists of a particular feature (or features) of brain activities and the methodology used to measure that features. BCIs may focus on frequency-domain features (Spectral), time-domain features (Temporal), or the features measured in respect to the location of the events taken place on the scalp (Spatial). These features will be fully described and discussed later in Section ???. Each BCI uses a particular algorithm to translate its input into output control signals. Due to the high number of available techniques, the translation algorithm might include linear or nonlinear models, or neural network, to name a few, or a hybrid combination of them. In many cases, BCIs incorporate continual adaptation of important parameters to key aspects of the input provided by the user in order to improve the accuracy of capturing the intention and its translation corresponding to the cognitive learning curve the user takes while using the BCI system. BCI outputs can be cursor movements, letter or icon selection, controlling a robot arm, or another form of device control, and provide the feedback that the user and the BCI can use to adapt so as to optimize communication.

In addition to the three main components of a BCI, as a system, it has other distinctive characteristics as well, which may be the reference of BCI's evaluation, or comparison to its peer systems. These include a BCI's response time, speed and accuracy, the information transfer rate, type and extent of user training required (which can be very crucial in terms of the target population of the intended BCI application), appropriate user population, and required feedback technique by which the user would remain motionless and well-understood by the system.

The focus of this thesis, although studying and practically experimenting input and output of a BCI (Chapter 4), is mostly on the translation algorithms of a BCI. A BCI translation algorithm is essentially a series of computations, arranged and formed in a fashion to most optimally turn

BCI system input features derived by the signal processing stage into actual device control commands [18]. The diversity in translation algorithms is due in part to diversity in BCIs intended real-world applications. Nevertheless, in all cases the goal is to maximize performance and practicability for the chosen application.

## 2.2 Electroencephalography (EEG)

The BCI systems, as the translators of the brain activities, evidently require at least one modality to detect the user's intention. Either flavors of these activities, spontaneous in the absence of external stimuli, or evoked brain activity, as specific patterns elicited in the presence of external stimuli, can be captured by either "electrophysiological" or "hemodynamic"-based modalities. When information is exchanged between the electro-chemical transmitters of the neurons, a set of electrophysiological activities is generated. The neurons generate ionic currents which flow within and across neuronal assemblies. This flow, once large enough, causes electrical and magnetic fields, which can be measured by the means of Electroencephalography (EEG), Electrocorticography (ECoG), and Magnetoencephalography (MEG). On the other hand, the hemodynamic response, is a process in which the blood releases glucose to active neurons at a greater rate than in an area of inactivity. This chemical change in the blood can be monitored by neuroimaging methods such as functional Magnetic Resonance Imaging (fMRI), functional Near Infrared Spectroscopy (fNIRS), and Positron Emission Tomography (PET) [19].

Hans Berger, a German psychiatrist, was the first to systematically study the electrical activities of human brain and to invent EEG [20]. His invention brought about a revolution and to date, a considerable number of research results related to EEG-based BCIs have been reported in various international journals, covering biomedical engineering, clinical neurology, neuroscience, and neuromodulation, using EEG headsets. The reason for this popularity is that EEG, although owning a few yet significant drawbacks, outweighs other modalities due to its minimal risk and the relative convenience of conducting studies, as well as other technical distinct advantages, which will be discussed furtheron in this section.

EEG as a medical tool has become a routine clinical practice to distinguish epileptic seizures

from other types of spells, such as psychogenic non-epileptic seizures, syncope (fainting), subcortical movement disorders, and migraine variants, to differentiate “organic” encephalopathy or delirium from primary psychiatric syndromes such as catatonia, to serve as an adjunct test of brain death, to prognosticate, in certain instances, in patients with coma, and to determine whether to wean anti-epileptic medications. However, EEG has been even more trending in research-related practices, especially and extensively in neuroscience, cognitive science, cognitive psychology, neurolinguistics, and psychophysiological research. The reason behind this trending interest, despite the relatively poor spatial sensitivity of EEG, is the several advantages it possesses over other its counterparts, as briefly outlined below.

EEG hardware is significantly and considerably more affordable by comparison with most other techniques. Also, immobility of modalities such as fMRI, PET, or MEG, limit the flexibility of experiment design and require a more complex, therefore costly, arrangements and setting at the data collection venue, while EEG sensors can be placed anywhere on the scalp not requiring any specific ambient conditions to work at. Moreover, EEG recordings hold a very high temporal resolution, on the order of milliseconds rather than seconds, thus, for clinical and research settings, EEG is commonly recorded at sampling rates above 250 Hz and up to 2000 Hz. Nowadays, modern EEG data collection systems are capable of recording at sampling rates above 20,000 Hz if desired. EEG, being absolutely silent while recording, enables researchers to not only study the responses to auditory stimuli, but also to investigate and track the brain changes during different phases of life, e.g., EEG sleep analysis can indicate significant aspects of the timing of brain development, including evaluation of adolescent brain maturation. Additionally, EEG, as a powerful tool to detect covert processing (i.e., processing that does not require a response), is literally non-invasive and can be used in subjects who are incapable of making a motor response. In contrast to all the useful advantages, EEG also possesses disadvantages that researchers must take into account before adopting this technique of recording as the tool by which they aim to answer the question of their study. The first drawback of EEG recordings is poor spatial resolution on the scalp as compared to techniques such as fMRI, and in order to compensate for this downside, intense interpretation is required just to hypothesize what areas are activated by a particular response. The quality of EEG signals is affected by scalp, skull, and many other layers as well as background noise. Noise is key

| Author(s)                  | Participants (condition)   | Control mechanism | Application           | Evaluation criteria  |
|----------------------------|--|-------------------|-----------------------|--|
| Birbaumer et al. (1999)    | 2 (ALS)  | SCP               | Spelling              | Selection accuracy, characters/min                         |
| Kübler et al. (1999)       | 3 (ALS), 13 able-bodied  | SCP               | Spelling              | Selection accuracy, words/min                              |
| Birbaumer et al. (2000)    | 5 (ALS)  | SCP               | Spelling              | Selection accuracy, characters/min                         |
| Donchin et al. (2000)      | 3 (complete paraplegia), 10 able-bodied<br>1 (incomplete paraplegia)   | P300              | Spelling              | Selection accuracy, characters/min                         |
| Kübler et al. (2001)       | 2 (ALS)  | SCP               | Spelling              | Selection accuracy, CRR                                    |
| Kaiser et al. (2002)       | 1 (ALS)  | SCP               | Environment control   | Selection accuracy   |
| Hinterberger et al. (2003) | 1 (ALS)  | SCP               | Spelling              | Selection accuracy (CRR)                                   |
| Sellers et al. (2003)      | 3 (ALS)  | P300              | Spelling              | Classification accuracy, selection accuracy Characters/sec |
| Müller et al. (2003)       | 1 (infantile CP)   | SMR               | Cursor control        | Classification accuracy                                    |
| Neumann & Kübler (2003)    | 11 (not specified)   | SCP               | Cursor control        | Letters/min  |
| Krausz et al. (2003)       | 4 (SCI) partial paralysis  | SMR               | Cursor (ball) control | Error rate, information transfer rate                      |
| Neumann & Birbaumer (2003) | 5 (ALS)  | SCP               | Spelling              | Selection accuracy (CRR), $r^2$                            |
| Neumann et al. (2003)      | 5 (ALS)  | SCP               | Spelling              | Self assessment, selection accuracy (CRR)                  |
| Neuper et al. (2003)       | 1 (CP)   | SCP               | Spelling              | Selection accuracy $r^2$ , letters/min                     |
| Bayliss et al. (2003)      | 1 (ALS), 9 able-bodied   | P300              | 3-choice switch       | Transfer rate  |
| Kübler et al. (2004)       | 10 (ALS), 10 able-bodied   | SCP               | Cursor control        | Selection accuracy (CRR)                                   |
| Wolpow & McFarland (2004)  | 2 (SCI), 2 able-bodied   | SMR               | Cursor control        | Correct target hit rate within 10 sec, $r$                 |
| Sellers et al. (2004)      | 15 (ALS), 1 (brain stem stroke)  | P300              | Spelling              | Selection accuracy   |
| Kübler et al. (2005)       | 4 (ALS)  | SMR               | Cursor control        | Selection accuracy (hitting the target), $r^2$             |
| Piccione et al. (2006)     | 1 (ALS), 1 (LIS post-vertebrobasilar trombosis)<br>1 (SCI), 1 (Guillain Barré syndrome), 1 (MS), 7 able-bodied | P300              | 4-choice switch       | Accuracy (hitting the target), transfer bit-rate           |

Note. ALS: amyotrophic lateral sclerosis, CRR: correct response rate, CP: cerebral palsy, DMD: Duchenne muscular dystrophy, LIS: locked-in syndrome, Lv: level, MD: muscular dystrophy, MS: multiple sclerosis,  $r^2$ : portion of the total variance due to user intent (Wolpow et al., 2002), SCI: spinal cord injury, SCP: slow cortical potentials, SMA: spinal muscular atrophy, SMR: sensorimotor rhythm.

Figure 2.1: Summary of BCI studies involving participants with disabilities (1999-2005) [23].

to EEG, insofar as it reduces the SNR and therefore the ability to extract meaningful information from the recorded signals.

All in all, considering the aforementioned pros and cons, EEG has always been a strong candidate for BCI developments, especially since changes in cortical electrical activity resulting from mental stimulation occur faster than the accompanying changes in hemodynamics. Many studies, namely, [21] and [22], have reported successful utilization of EEG-based BCIs, particularly enabling able-bodied users to generate fast and reliable control signals. In the interesting work of Moghimi *et al.* [23], 39 studies reporting EEG-oriented BCI assessment by individuals with disabilities were identified in the past decade and investigated. An interesting result of this study is shown in Table 2.1.

In the interest of standardizing the placement of EEG electrodes, an internationally recognized

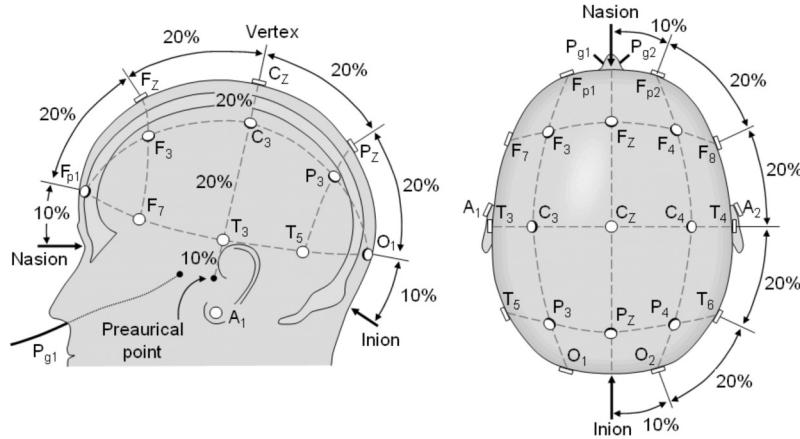


Figure 2.2: An example of 10-20 setting of EEG electrodes placement [24].

method is established, known as “10–20 system”. The system is based on the relationship between the location of an electrode and the underlying area of cerebral cortex. The distances between adjacent electrodes are either 10% or 20% of the total front-back or right-left distance of the skull; that is the reason behind the name of this system. Each site has a letter to identify the lobe, and a number to identify the hemisphere location. “F” stands for frontal, “T” for temporal, “C” for central, “P” for parietal, and “O” for occipital. Even numbers are refer to the electrodes placed on the right hemisphere, and the odd numbers are allocated to the electrodes located on the left hemisphere. To measure the distances and segmenting them, four anatomical landmarks are used: first, the nasion which is the point between the forehead and the nose; second, the inion which is the lowest point of the skull from the back of the head and is generally indicated by a prominent bump; the pre auricular points are anterior to the ears. Considering these main instructions, an example of a 10-20 system EEG electrode placement is elucidated in Figure 2.2.

To understand the studies employing EEG as a tool for recording brain signals, it is considerably important to be familiar with the EEG rhythmic activities and patterns. These activities are divided into frequency bands. These designations arose because rhythmic activity within a certain frequency range was noted to have a certain distribution over the scalp or a certain biological significance. The first frequency band commonly is considered to begin from 1 Hz and the last one is capped by 40 Hz, whereas activity below or above this range is likely to be artifactual. The following categories, provide an overview on the location and range of various meaningful and distinguishable brain

signal patterns for the purposes of data analysis.

- **Delta Patterns**

These patterns are ranged between 1–4 Hz with highest amplitudes and slowest waves. These signals are usually captured from frontal (in adults) and posterior (in children) part of the scalp. Delta waves are known as slow-wave sleep in adults and also are commonly observed in babies.

- **Theta Patterns**

These patterns are ranged between 4 – 7 Hz. Theta signals are mostly found at the locations of scalp that are not related to the task at hand, therefore, they are normally considered as “idling” waveforms. This category of signals has also been found to spike in situations where a person is actively trying to repress a response or action. Moreover, this range has been associated with reports of relaxed, meditative, and creative states.

- **Alpha Patterns**

These patterns are ranged between 7–13 Hz. These signals are often generated at the posterior regions of head (both sides) and are higher in amplitude on dominant side. That is the reason why Hans Berger named this EEG activity as the “Alpha Wave” or the “posterior basic rhythm”. This category is famously known as resting/relaxing state, due to its generation while the eyes are closed. It is interesting to know that one of the ways for EEG researchers and engineers to test their application, is to ask the subjects to close their eyes and relax; the frequency content observed during this time has to be mostly focused at Alpha rhythms. In addition to the posterior basic rhythm, there are other normal alpha rhythms such as the mu rhythm (alpha activity in the contralateral sensory and motor cortical areas) that emerges when the hands and arms are idle.

- **Beta Patterns**

These patterns are ranged between 14 – 30 Hz, and although being of high interest for capturing as a response to stimuli, this category is of low amplitude, multiple and varying frequencies is often associated with active, busy or anxious thinking and active concentration.

Low amplitude of Beta pattern makes it notably susceptible to contamination by artifacts and noise. The location of signal generation on the scalp is at both sides, symmetrically distributed, and most evidently towards the frontal side of the head. These waves are generated while the subject is actively calm and is focused and highly alert, thinking, especially, it is the dominant rhythm in patients who are alert or anxious or who have their eyes open, namely, the locked-in patients.

- **Gamma Patterns**

These patterns are seen at frequencies between 30–100 and are generated at the Somatosensory cortex of the brain. This category of waveforms are displayed during cross-modal sensory processing (perception that combines two different senses, such as sound and sight). Gamma patterns are thought to represent binding of different populations of neurons together into a network for the purpose of carrying out a certain cognitive or motor function. Also Gamma appears during short-term memory matching of recognized objects, sounds, or tactile sensations.

- **Mu Patterns**

These patterns are ranged between 8–13 Hz and partly overlapped with other frequencies. They are generated at the Sensorimotor cortex of the brain, and represent rest-state motor neurons.

In awake people, primary sensory or motor cortical areas often display 8–12 Hz EEG activity when they are not engaged in processing sensory input or producing motor output. Computer-based analyses reveal that idling waves are distinguished from each other by location, frequency, and/or relationship to concurrent sensory input or motor output. These idling patterns are usually associated with 18-26 Hz (range of Beta) rhythms. While some Beta patterns are harmonics of Mu patterns, some are separable from them by topography and/or timing, and thus are independent brain signal features [16]. Beta and Mu waves are associated with those cortical areas most directly connected to the brains normal motor output channels, therefore, are leader choices for EEG-based BCIs which enable the subject to command movements to the BCI, without any peripheral muscle movement.

As the last not least concept outlined before closing the discussion on EEG, it is important to introduce the mental process, Motor Imagery (MI). MI is one of the most popular and widely used techniques for the BCI systems to be efficient, as they are supposed to be highly accurate and capable of well-intrepretation by the quickest pace possible, regardless of the limitations of the end-user. The MI is defined as merely imagination of a limb movement, with no actual movement or peripheral (muscle) activation [25]. This mental execution of a movement is known to induce brain activity in the same way performing an acutal movement brings about the firing of neurons in the brain [26]. According to this view, the main difference between performance and imagery is that in the latter case execution would be blocked at some corticospinal level [27]. The variation in brain activity is quantified from Electrophysiological recording by EEG during the MI task. In MI-based BCIs, patients often receive visual or kinesthetic feedback in order to promote the brain response to the MI task. Increasing the accuracy of BCIs using this mental process is the focus of my thesis throughout the next chapter.

This completes a brief discussion on BCIs, EEG, and essential knowledge for understanding the applications of these two. Next, I will discuss the techniques and methods employed to process data for an EEG-based BCI system.

## 2.3 Data Dimensionality Reduction

Through the recent advances in statistics, signal processing, machine learning, and information theory, the size and the number of random variables, storing information regarding an ongoing study, are experiencing a significant growth. As a first step to proceed on the data understanding, researchers and engineers are often confronted with the problem of a proper and/or optimal selection of information among numerous patterns in a set of data. The superficial dimensionality of data, or the number of individual observations constituting one measurement vector, is often much greater than the intrinsic dimensionality, the number of independent variables underlying the significant non-random variations in the observations [28]. Therefore, to conclude the features and/or the attributes of a set of data, the first step required is to reduce the dimensionality of the data.

Dimensionality reduction is defined as the process of reducing the number of random variables under consideration by obtaining a set of key variables. Dimensionality reduction not only allows us to shift the focus of processing on the more informative variables, but also results in, generally noticeably, speeding up the processing and classification phase.

In comparison with other sorts of biomedical signals, EEG is considered to be excessively complicated for an untrained observer to understand. Raw EEG signals are also extremely burdensome to be directly processed, mostly, as a consequence of the high temporal resolution of EEG technique and the spatial mapping of functions onto different regions of the brain and electrode placement. Hence, prior to applying other processing methods to extract features and classify the recorded data, dimensionality reduction is commonly used to determine a reduced feature set, including only (more or less) the data needed for further quantifications, in respect to a negligible loss of information during this process.

Principal Component Analysis (PCA) [29] and Singular-Value Decomposition (SVD) [30] are known to be well-established methods for the purpose of dimensionality reduction. Percent Root-mean-square Difference (PRD) [31] is also a method based on SVD which has been well-regarded in Electrocardiography (ECG) processing and is applicable to EEG signals. This subsection serves to provide an overview of the aforementioned techniques.

### 2.3.1 Principal Component Analysis (PCA)

The PCA is one of the most traditionally used techniques in EEG signal processing and BCI module design [19]. This technique is a leader choice method both for dimensionality reduction and feature extraction among its similar techniques in many fields of computer science as well. In the PCA, we seek to represent a given  $n$ -dimensional data in a lower-dimensional space. This will reduce the degrees of freedom; reduce the space and time complexities. The objective is to represent data in a space that best expresses the variation in a sum-squared error sense. The PCA functions significantly better if an estimation of the number of independent components is known a-priori.

The basic approach in principal components is theoretically rather simple. First, the  $n$ -dimensional mean vector  $\mu$  and  $n \times n$  covariance matrix  $\Sigma$  are computed for the full data set. Next, the eigenvectors and eigenvalues are computed, and as the eigenvalues  $\{\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n\}$  are sorted in a

descending order, so are the eigenvectors  $\{e_1, e_2, \dots, e_n\}$  accordingly. Subsequently, a subset of eigenvectors associated with the largest eigenvalues are chosen. In practice, this is done by looking at the spectrum of eigenvectors. Often there is a clear distribution within the spectrum, implying an inherent dimensionality of the subspace governing the signal. The other dimensions are noise. Form a  $(m \times m)$  matrix  $\mathbf{A}$  whose columns consist of the  $m$  eigenvectors. Preprocessing of the data is performed as follows

$$\hat{\mathbf{X}} = \mathbf{A}^T \times (\mathbf{X} - \bar{\mathbf{X}}), \quad (1)$$

where superscript  $T$  denotes transpose operator and  $\bar{\mathbf{X}}$  is the matrix of PCA essentially rotates the set of points around their mean in order to align with the principal components. This moves as much of the variance as possible (using an orthogonal transformation) into the first few dimensions. The values in the remaining dimensions, therefore, tend to be small and may be dropped with minimal loss of information. The PCA has the distinction of being the optimal orthogonal transformation for keeping the subspace that has largest variance.

### 2.3.2 Singular Value Decomposition (SVD) and Percentage Root-mean-square Difference (PRD)

Several (if not all) data dimensionality reduction techniques are based on the decomposition of a key matrix, into its eigenvectors and eigenvalues. The reason behind this approach is the fact that through a canonical representation, there can be information retrieved that are just not obvious to a researcher by a glance at the original matrix, information such as: the underlying probability distribution of the matrix; similarities of random variables; the dependencies of random variables onto each other; and, many other useful key information [32]. In linear algebra, the SVD is a factorization of a real or complex matrix. SVD is proposed as the generalization of the conventional eigendecomposition of a positive semidefinite normal matrix (e.g., a symmetric matrix with positive eigenvalues) to any  $(m \times n)$  matrix via an extension of the polar decomposition. It has several useful applications in signal processing and statistics, namely, computing the pseudoinverse, least squares fitting of data, multivariable control, matrix approximation, and determining the rank, range and null space of a matrix. However, in biomedical engineering, this approach can be the first step taken

towards dimension reduction of the data at hand. To outline the method, suppose  $\mathbf{X}$  is an  $(m \times n)$  matrix which contains real or complex numbers. Then there exists a factorization, called a singular value decomposition of  $\mathbf{X}$ , of the form

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^H, \quad (2)$$

where

- $\mathbf{U}$  is an  $(m \times m)$  unitary matrix;
- $\Sigma$  is a diagonal  $(m \times n)$  matrix with non-negative real numbers on the diagonal;
- $\mathbf{V}$  is an  $(n \times n)$  unitary matrix and  $\mathbf{V}^H$  is the Hermitian (conjugate) transpose of  $\mathbf{V}$ .

Throughout the development of this thesis, I have been employing the SVD technique along with a well-regarded yet straightforward measure, Percentage Root-mean-square Difference (PRD), to evaluate the optimality of the dimensionality reduction step. PRD is widely known for its noteworthy application for Electrocardiography (ECG) signal processing techniques, for instance in [33] and [34]. While illustrating notable results in ECG, during my thesis research work, PRD turned to be practically applicable for EEG signal processing, once applied alongside the dimensionality reduction step. PRD can be considered as a quality control measure, to address the main concern of the dimensionality reduction techniques, which is the preserving the target parts of the signals while removing the redundant and irrelevant information. PRD ensures that the signals matrices are reduced in size up to a certain amount of information loss. For instance, suppose we have an estimation of irrelevant data contained within the recorded signals, which is about  $p\%$  of the entire data. PRD, utilizing SVD, decomposes the signals and begins reconstructing them variable by variable. Every time data of each variable is added, the reconstructed signals are evaluated and compared to the original one, and the error of compatibility is computed. The algorithm keeps adding variables until the error is qual or less than  $p\%$ . In order to clarify the way PRD functions as described, the Algorithm 2.3.2, presents the pseudocode of the PRD approach in details.

---

**Algorithm 1** SVD–PRD DIMENSIONALITY REDUCTION

---

**Input:** {Original EEG (OEEG) signals  $\mathbf{X}$  (Channles as the variables each containing same length of time series signals) and the Percentage of acceptable error  $p\%$ .

**Output:** {Minimum number of variables (channels) required to maintain the signals' quality in respect to the aforementioned error percentage.

1: **Decompose OEEG by SVD :**  $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}(\mathbf{X})$ . Number of columns taken into account ( $k$ ) in  $\mathbf{U}$  and  $\mathbf{V}$  is set to 1.

2: **Reconstruction Loop:**

- First  $k$  diagonal element(s) of  $\mathbf{S}$  and first  $k$  column(s) from  $\mathbf{U}$  and  $\mathbf{V}$  are chosen to reconstruct the EEG Signal as  $\hat{\mathbf{X}}$  (**REEG**), as per Equation 2.

- The PRD is computed as follows

$$PRD = \sqrt{\frac{\sum_{i,j} (\mathbf{X}_{i,j} - \hat{\mathbf{X}}_{i,j})^2}{\sum_{i,j} (\mathbf{X}_{i,j})^2}} \times 100. \quad (3)$$

• **Evaluation**

**If:**  $PRD \leq p\%$

    break;

**else:**

$k = k + 1$ ;

3: **Number of Variables Required:** Final  $k$  will be the output of the function as the minimum number of variables (channels) required for a more accurate dimension reduction.

---

## 2.4 Feature Extraction

As discussed previously, different thinking activities result in different patterns of brain signals and these mental tasks are of utmost importance while designing an experiment or an application for BCIs. From this outlook, BCI is seen as a pattern recognition system that classifies each pattern into a class according to its features. To elucidate this matter more clearly, it is important to define a “feature”. During a phenomenon which is under observation of a machine learning/pattern recognition outline, a feature is an individual measurable property or characteristic of that phenomenon under study. In other words, a feature is supposedly a variable/attribute which together with other peerly chosen features can represent the data statistically and, if selected suitably, will contribute to a rather accurate classification of the entire dataset. However, choosing informative, discriminating and independent features is a crucial step for development of effective algorithms in pattern recognition, classification and regression. Having said that, “feature extraction” is the process of starting from an initial set of measured data and building the derived values (features), aiming the features

to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations.

BCIs extract some features from brain signals that reflect similarities to a certain class as well as differences from the rest of the classes. The extracted features should be measured or derived from the properties of the signals which contain the discriminative information needed to distinguish their different types. However, the challenging issue of feature extraction in BCIs is that it is interwoven with the fact that the information of interest in brain signals is hidden in a highly noisy environment, and brain signals comprise a large number of simultaneous sources. A signal that may be of interest could be overlapped in time and space by multiple signals from different brain tasks. For that reason, in many cases, it is not enough to use simple methods such as a band pass filter to extract the desired band power.

From signal processing point of view, feature extraction is done after preprocessing and data dimension reduction, as an important step in the construction of any pattern classification and aims at the extraction of the relevant information that characterizes each class. These feature vectors are then used by classifiers to recognize the input unit with target output unit. The classifier's task is much more facilitated if it is to classify between different classes by looking at these features as it allows fairly straightforward to distinguish. However, before concatenated into a single feature vector, multiple features can be selected from different channels and from various time segments, although, it is not desirable to process high dimensional features. In several neuroimaging studies, the sample size, or the number of the subjects of the study, is often much less than the size of scanned samples. Therefore, the number of features greatly outnumbers the number of the subjects. This challenge is known as "*curse of dimensionality*" or "*small-n-large-p*" [35]. In order to choose the most suitable features, one may attempt to examine all the possible subsets for the features, although, as the number of possibilities grows exponentially, this approach becomes more and more impractical and exhaustive method of search. Obviously, there are more efficient and optimal feature extraction methods to replace this exhaustive search.

Before moving forward to describing the feature extraction technique used in this thesis, it is worth viewing feature extraction and classification from a big picture. The classification step, as the final step of recognizing the patterns, can be done in three different ways: (i) Supervised; (ii)

Unsupervised; and, (iii) Semi-Supervised, which will be discussed more in details throughout the next section. Note that the feature extraction techniques utilized for different parts of my thesis are all chosen with regards to a supervised classification method. In order to select the most appropriate classifier for a given BCI system, it is essential to clearly understand what features are used, what their properties are and how they are used. The next two subsections are allocated to provide a brief overview on Common Spatial Patterns (CSP) method.

#### 2.4.1 Common Spatial Patterns

The Common spatial patterns is a particularly popular and powerful signal-processing technique used for feature extraction in EEG-based BCIs. Originally, CSP has been designed for the analysis of multichannel data belonging to 2-class problems. Nevertheless, some extensions for multiclass BCIs have also been proposed, e.g., [36], which is not the focus of this thesis, thus, is not explained.

The CSP, as a mathematical algorithm, computes spatial filters that aim at achieving optimal discrimination by separating a multivariate signal into additive subcomponents which have maximum differences in variance between two class. Hence, the signal-to-noise ratio is increased and adverse effects of volume conduction is reduced [37]. In other words, CSP projects multichannel EEG signals into a subspace, where the differences between classes are highlighted and the similarities are minimized. It aims to make the subsequent classification much more effective. The main idea of the CSP approach is to employ a linear transform for the purpose of projecting the multi-channel EEG data onto a low-dimensional spatial subspace. The rows of the projection matrix which serves to this goal consist of the associated weights of the channels. The CSP method is based on simultaneous diagonalization of the covariance matrices of both classes. The Algorithm 2 describes the steps of implementation of CSP.

The aforementioned algorithm is constructed for the case where it is prefered to utilize feature vectors each including two elements. In the case where more elements are required, at Step 4, symmetrically, equal number of eigenvectors are retrieved and put together to build the whitening matrix.

This completes a brief outline over feature extraction. Throughout the next chapters, the additional feature extraction methods will be described. Next section is allocated to explanation of

---

**Algorithm 2** COMMON SPATIAL PATTERNS STEP BY STEP (FOR FEATURE VECTORS OF SIZE 2)

---

**Input:** *{(a) Original EEG (OEEG) signals as a tensor of size  $N_c \times N_t \times N_e$ ;  $N_c$  being the number of EEG channels,  $N_t$  denotes the number of time samples within each trial of performing the task by the subject, and  $N_e$  represents the number of times subject performed the task; (b) Lables of OEEG, that indicate to which class each trial belongs.}*

**Output:** *{Feature Vectors, ready to be classified}*

- 1: **Covariance Matrices:** Computing the sample covariance matrix corresponding to each trial  $\mathbf{X}_i$  as follows:

$$\mathbf{C}_i = \frac{\mathbf{X}_i \mathbf{X}_i^T}{\text{Trace}(\mathbf{X}_i \mathbf{X}_i^T)} \quad (4)$$

The superscript  $T$  indicates the transpose of the matrix.

- 2: **Class Distinguisher Loop:** Trial by trial, each  $\mathbf{C}_i$  is assigned to its corresponding class, using the lables.

- 3: **Class Averages:**

- The covariance matrices of each class are averaged resulting in  $\bar{\mathbf{C}}_1$  and  $\bar{\mathbf{C}}_2$ .
- The composite spatial covariance  $\mathbf{C}_c$  is computed as follows.

$$\mathbf{C}_c = \bar{\mathbf{C}}_1 + \bar{\mathbf{C}}_2.$$

- 4: **Whitening Matrix:**

- The composite spatial covariance is decomposed to its eigenvalues and eigenvectors , in respect to the average covariance matrix of the first class, i.e.,

$$[\mathbf{EVec}, \mathbf{EVal}] = \text{eig}(\bar{\mathbf{C}}_c, \bar{\mathbf{C}}_1).$$

- Whitening matrix is composed as follows.

$$\mathbf{W}(:, 1) = \mathbf{EVec}(:, 1);$$

$$\mathbf{W}(:, 2) = \mathbf{EVec}(:, end);$$

- 5: **Finalizing the feature vectors:**

- The Whitening matrix is applied to each trial of the original data, i.e.,

$$\mathbf{Z}_i = \mathbf{W}^t \times \mathbf{X}_i \times \mathbf{X}_i^t \times \mathbf{W}.$$

- Feature vector corresponding to each trial's  $\mathbf{Z}$  is computed as follows.

$$\mathbf{f}_i = \log \frac{\text{diagonal}(\mathbf{Z}_i)}{\text{trace}(\mathbf{Z}_i)}. \quad (5)$$


---

classification, the last step of processing data in any desired BCI.

## 2.5 Clasification

As stated previously, the primary goal of a BCI is to translate the intent of a subject directly into control commands for a computer application, a neuroprosthesis, or any other external device.

In BCIs which take advantage of training data to build a model for translation of the features (supervised learning), users are provided with instructions on how to perform a task as a response to a stimulus. Thereafter, an often significant number of trials are required to calibrate a BCI and prepare it for successful further analysis and interpretation. In most existing BCIs, this identification relies on a classification algorithm, i.e., an algorithm that aims at automatically estimating the class of data as represented by a feature vector. This objective can be accomplished by a statistical analysis of a calibration measurement in which the subject performs well-defined mental acts, such as imagined movements [38].

Classification is defined as the problem of statistically identifying to which of a set of categories a new observation belongs. This problem attempts to learn the relationship between a set of feature variables and a target variable of interest. Since many practical problems can be expressed as associations between feature and target variables, this provides a broad range of applicability of this model [39]. In the importance of clarification of the classification terminology, it is essential to define a few terms used throughout this thesis:

- **Trial & Epoch:** During a run/experiment with an EEG-based BCI, the subject is asked to perform the task related to the stimulus a certain number of times, each of these observations is called a trial. The set of signals recorded during a trial is called an epoch.
- **Label:** In the event of knowing the intent of the subject associated with a trial beforehand, this information is considered as a label.
- **Training data/trials:** The training data, as a part of collected data, consists of labeled trials utilized to construct a classification model. These trials together are also called the training data/dataset.
- **Test data/trials:** The test trials, the rest of collected data, are those that are evaluated using the generated model via the training trials, and the classification model assigns a label to each unlabeled test trial.
- **Classification accuracy:** In the case that the ground truth, i.e. labels of the test data, is available, it is compared against the estimated label and overall performance of the classification

model is reported as the classification accuracy, often in percentage.

Based on the stated definition, a *classifier* is defined as a function or an algorithm that maps every possible input available in the calibration (training) dataset to a finite set of decisions. In other words, given a set of training data points along with their associated training labels, a classifier determines the class label for an unlabeled test instance. Classification algorithms use the extracted features as independent variables to define boundaries between the different targets in the feature space. Building classification algorithms which are traditionally calibrated by users using a labeled data set, are also known as supervised learning.

Classification algorithms can be developed via either offline, online or both kinds of sessions. The offline session involves the examination of datasets after the experiment is carried out. The statistics of the data may be estimated from observations across entire sessions and long computational processes may be performed. While offline data analysis is valuable in terms of studying the behavior of brain signals and the effectiveness of the BCI processing algorithm, it does not address real-time requirements. However, online sessions provide a means of BCI system evaluation in a real-world environment. The data is processed in a causal manner and for higher efficiency of the processing algorithm, it is implemented as a closed-loop system. Evidently, online analysis can yield solid evidence of BCI system performance, and that is why offline simulation and cross-validation can be valuable methods to develop and test new algorithms [19].

Regardless of the model selected as to define the classifier, there are four main steps to construct a suitable and responsive classification model for the desired BCI system.

- (1) **Choosing a model:** There are several models of classifiers suggested and developed by researchers over time, some are well-suited for image processing, some other are known to work best for text-based data, and likewise for many other applications, also, the chosen model needs to be compatible with the feature vectors. A few classification models that are known to work best for BCI systems are described further in this section.
- (2) **Training:** Using the training data, the parameters of the chosen model are determined (learned). In this step, the training epochs and their labels are employed to incrementally improve the

classification model’s ability to predict to which class a desired test epoch belongs. The training process is initialized with some random values for the parameters of the model and then attempting to predict the output with those values. Then, the model continues to learn and tunes its parameters to improve the comparison results, trial by trial until it finds the best fitting model. In the interest of determining how many trials out of the entire experiment should be utilized as for the training purpose, a conventionally common rule of thumb is to split into training-test by 70%–30% or 80%–20%; certainly, much of the applicability of this convention depends on the size of the original source dataset.

- (3) **Evaluation:** Once the training is complete, the trained model has to be evaluated. This is where that dataset that we set aside earlier comes into play. Evaluation enables the designer of the BCI to test the constructed model against data that has never been used for training, referred to as the test data. This is meant to be representative of how the model might perform in the real world. The evaluation result is generally reported as ‘validation accuracy’ in percentage.
- (4) **Prediction:** Prediction is the step where the question about the subject’s intent for each trial is answered. The trained and evaluated model, if performing acceptably, is applied to the test dataset, and for each trial, a lable is predicted by the classifier. The comparison between the grand truth and the estimated lables indicates the effectiveness of the classifier.

Training an algorithm and evaluating its statistical performance on the same data yields an overoptimistic result. Therefore, it is beneficial to know how a model would perform when it is applied to new data, beforehand. However, the test dataset is set aside to assess the predictive performance of the models and to judge how they perform outside the samples the parameters are learned from. In this case specifically, it is noteworthy to analyze the capability to accurately translate the intent of the BCI user. In most real applications, only a limited amount of data is available, which leads to the idea of splitting the data: Part of data (the training sample) is used for training the algorithm, and the remaining data (the validation sample) are used for evaluating the performance of the algorithm. The validation sample can play the role of “new data”. “Cross validation” is the approach which is mainly used to serve this purpose and to estimate how accurately

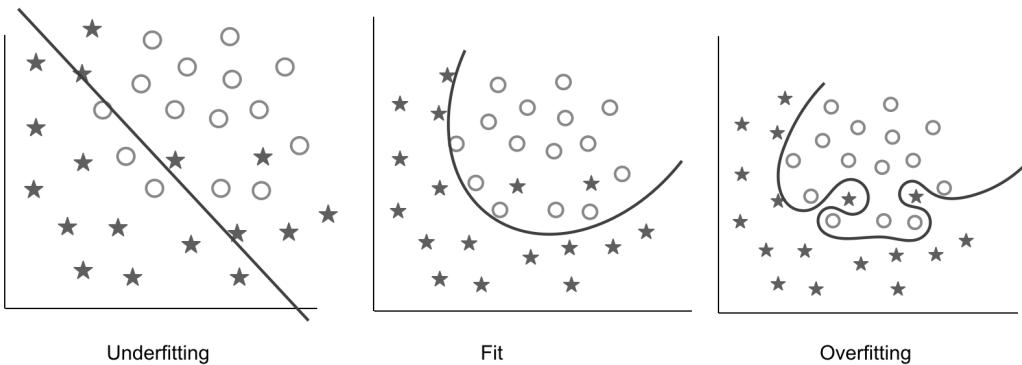


Figure 2.3: The problem of Underfitting and Overfitting.

a predictive model will perform in practice. The cross validation technique used throughout this thesis is called “*k*-fold cross validation”. In *k*-fold cross-validation, the original training dataset is randomly partitioned into *k* equal sized subsets. Out of the *k* subsets, a single subset is preserved as the validation data for evaluation of the model, and the remaining *k* - 1 subsets are used as to train the classification model and tune the parameters. The cross-validation process is then repeated *k* times, equal to the number of the folds, with each of the *k* subsets used exactly once as the validation data. The *k* results from the folds can then be averaged to serve as a single performance estimation. This way, all observations are used for both training and validation, and each observation is used for validation exactly once.

Classifiers commonly face a main problem related to the pattern recognition task, the problem of bias-variance tradeoff. This problem represents the natural trend of the classifiers towards a high bias with low variance and vice versa. Stable classifiers are characterized by high bias with low variance, while unstable classifiers show high variance with low bias. To achieve the lowest classification error, bias and variance should be low simultaneously. A set of stabilization techniques such as the combination of classifiers or regularization can be used to reduce the variance. In other words, low-performance classifiers may occur as a result of not proper fitting of the classification model, maybe because, the model is too simple to describe the target, or maybe model is too complex to express the target. *Underfitting* and *Overfitting* both are issues that lead to poor predictions on new data sets, as shown in Fig. 2.3.

*Underfitting* occurs when a statistical model cannot capture the underlying trend of the data.

Intuitively speaking, underfitting occurs when the model or the algorithm does not fit the data well enough. Specifically, underfitting occurs if the model or algorithm shows low variance but high bias. Underfitting is often a result of an excessively simple model. On the other hand, overfitting occurs when a statistical model captures the noise of the data. Intuitively speaking, overfitting occurs when the model or the algorithm fits the data too well. Specifically, overfitting occurs if the model or algorithm shows low bias but high variance. Overfitting is often a result of an excessively complicated model, and it can be prevented by cross-validation. This completes a general overview of the concept of classification and its essential terms related to it. Throughout the thesis, I have taken advantage of classification models such as “Linear/Quadratic Discriminant” and “Support Vector Machines”; a brief outline of each of these models is provided below.

- **Linear and Quadratic Discriminant Analysis:**

Linear Discriminant Analysis (LDA) is a simple classification model with rather acceptable accuracies, [40] for instance, without requiring highly complex computations. LDA is usually applied to classify patterns into two classes, although it is possible to extend the method to multiples classes [41]. For a two-class problem, LDA assumes that the two classes are linearly separable. In respect to this assumption, a linear discrimination function is defined in a fashion that it represents a hyperplane in the feature space in order to distinguish the classes. The class to which the feature vector belongs will depend on the side of the plane where the vector is found. The decision plane can be elucidated mathematically as follows

$$g(\mathbf{f}_i) = \mathbf{a}^t \mathbf{f}_i + b_0, \quad (6)$$

where,  $\mathbf{a}$  is the weight vector,  $t$  is the transpose operator,  $\mathbf{f}_i$  is the input feature vector extracted from the training sets, and  $b_0$  is the threshold. The input feature vector is assigned to one class or the other on the basis of the sign of  $g(\mathbf{f}_i)$ . There are several methods to compute  $\mathbf{a}$ , e.g.,  $\mathbf{a}$  may be calculated as follows,

$$\mathbf{a} = \mathbf{C}_c^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1), \quad (7)$$

where the matrix  $\mathbf{C}_c$  is as follows,

$$\mathbf{C}_c = \frac{\bar{\mathbf{C}}_1 + \bar{\mathbf{C}}_2}{2}, \quad (8)$$

and,  $\bar{\mathbf{C}}_i$  is the same as calculated in the Algorithm 2, and  $\boldsymbol{\mu}_i$ , the estimated mean of class  $i$  is calculated as below:

$$\boldsymbol{\mu}_i = \frac{1}{n} \times \sum_{i=1}^n \mathbf{F}_i, \quad (9)$$

where,  $\mathbf{F}_i$  is a matrix containing  $n$  feature vectors of class  $i$ , as  $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \dots, \mathbf{f}_n \in \mathbb{R}^{N_e}$ .

LDA serves to create a new variable, which would be a combination of the original predictors (features of the train set), which is accomplished through maximizing the differences between the features of the predefined classes. As a result, a discriminant score is a weighted linear combination of the predictors. The weights are estimated in a way that the differences between mean discriminant scores of each class has the maximum distance from all others. Generally, those predictors which have large dissimilarities between class means will have larger weights, at the same time weights will be small when class means are similar.

LDA assumes that the observations within each class are drawn from a multivariate Gaussian distribution and the covariance of the predictor variables are common across all levels of the responses (lables), however, Quadratic discriminant analysis (QDA) provides an alternative approach. While utilizing QDA, it is assumed that the measurements from each class are normally distributed. Unlike LDA, in QDA there is no assumption that the covariance of each of the classes is identical. In the case where the feature vectors consist of two classes, QDA seeks for surfaces separating the classes among conic sections (i.e. either a line, a circle or ellipse, a parabola or a hyperbola). In this sense, we can state that a quadratic model is a generalization of the linear model, and its use is justified by the desire to extend the classifier's ability to represent more complex separating surfaces. The QDA is very similar to the LDA except that because the covariance matrix is not identical, quadratic terms are also involved. This approach allows for more flexibility for the covariance matrix, therefore, tends to fit the data better than LDA, in contrast, it has more parameters to estimate.

- **Support Vector Machines:**

Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side. SVM is a classifier similar to LDA classifiers, However, in contrast, SVM selects the hyperplanes that maximize the margins, i.e., the distance between the nearest training samples and the hyperplanes [42]. Due to highly accurate results, SVMs have been successfully used in several BCI applications, namely, [29, 38, 43]. The majority of the similar classifiers use hyperplanes to separate classes, based on a flat plane within the predictor space. Whereas SVMs broaden the concept of hyperplane separation to data that cannot be separated linearly, by mapping the predictors onto a new, higher-dimensional space in which they can be separated linearly.

The method is called as it is, for the support vectors, a subset of training points in the decision function, are lists of the predictor values, taken from cases that lie closest to the decision boundary separating the classes. Computationally, finding the best location for the decision plane is an optimization problem that makes uses of a kernel function to build linear boundaries through nonlinear transformations, or mappings, of the predictors. The intelligent component of the SVM algorithm, however, is that it locates a hyperplane in the predictor space which is stated in terms of the input vectors and dot products in the feature space without ever representing the space explicitly. SVM chooses one particular solution: the classifier which separates the classes with maximal margin, see Figure 2.4. The margin is defined as the width of the largest ‘tube’ not containing samples that can be drawn around the decision boundary.

It should be noted that it is possible to create nonlinear decision boundaries, with only a low increase of the classifiers complexity, by implicitly mapping the data to another space, generally of much higher dimensionality, e.g., [45].

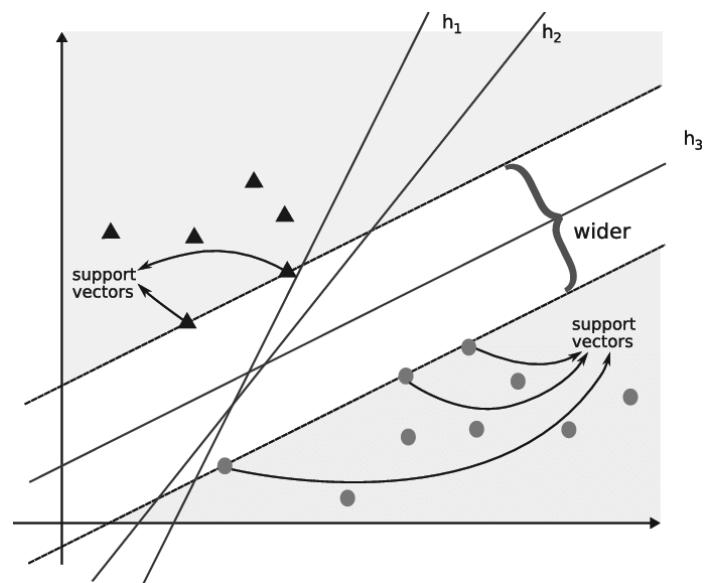


Figure 2.4: An example of possible and optimal hyperplanes [44]

## **Chapter 3**

# **Graph-based Frameworks for Spatio-Temporal Filtering and Dimentionality Reduction in MI EEG-based BCIs**

In this chapter, two frameworks, GD-BCI and GDR-BCI, are proposed to address the problem of processing EEG signals for motor imagery brain-computer interfaces. The goal is to identify the pattern of the brain activity using a robust method for pre-processing, processing, and classification of the EEG signals. To this end, the GD-BCI framework is a new graph-based approach, proposed to spatio-temporally filter the data by taking into account not only the geometrical structure of the channels/electrodes, but also the correlation between the EEG signals. The most significant feature vectors required for classification of EEG signals are adaptively selected through spectral decomposition of the data using the graph Laplacian matrix. The tangent space mapping method is then applied to bring the captured data into Euclidean space. In order to classify the dimensionally-reduced EEG signals, the linear support vector machine algorithm is employed. Experiments are conducted on Dataset IVa from BCI Competition III, including data from five different subjects consisting of

right hand and right foot motor imagery actions. The results show that the proposed GC-BCI framework provides higher classification accuracy as compared to the other existing methods. However, as the impressive accuracies come with the price of exhaustive search for a pair of constants required for adjustment of the graph, I sought for a solution to adaptively and automatically configure the graph, hence, the second framework the GDR-BCI is developed.

In the GDR-BCI, similar to the one proposed in GD-BCI, by capitalizing the fact that functionality of different connectivity neighborhoods varies based on the intensity of the performed activity and concentration level of the subject, an initial functional clustering of EEG electrodes is built by designing a separate adjacency matrix for each identified functional cluster. Using a collapsing methodology based on total variation measures on graphs, the overall model will eventually be reduced (collapsed) into two functional clusters. The proposed framework offers two main superiorities over its state-of-the-art counterparts and the GD-BCI: (i) First, the resulting dimensionality reduction is subject-adaptive and respects the brain plasticity of subjects, and; (ii) Second, the proposed methodology identifies active regions of the brain during the motor imagery task, which can be used to re-align EEG electrodes to improve accuracy during consecutive data collection sessions. The experimental results based on the same Dataset IVa from BCI Competition III show that the proposed method can provide higher classification accuracy as compared to its counterparts

### 3.1 Introduction To Graph Signal Processing

In mathematics, graph theory proposes a “graph” as a structure corresponding to a set of objects in which some pairs of the objects are in some sense “related”. It consists of a set of vertices or nodes (objects) and a set of edges or connections indicating the presence of some interaction (relation) between the vertices. Therefore, a graph can be defined as an ordered pair of  $G = (V, E)$  comprising a set  $V$  of vertices, together with a set  $E$  of edges. Moreover, for a graph representing a real life model, e.g., when a graph is an abstract representation of a network of sensors, the definition may be extended to  $G = (V, E, K)$ , where a number is assigned to each edge, together summarized in matrix  $K$ . Due to this simple yet powerful and flexible structure, graphs offer the ability

to model massive amounts of data and complex interactions among them in a systematic, organized, and processing-friendly manner. Graphs are generic data representation forms that are useful for describing the geometric structures of data domains in numerous applications, including social, energy, transportation, sensor, and neuronal networks. The weight associated with each edge in the graph often represents the similarity between the two vertices it connects. The connectivities and edge weights are either indicated by the physics of the problem at hand or inferred from the data. On the other hand, a growth of interest is widely observed in efficient signal processing techniques for representation, analysis and processing of large datasets (big data) emerging in various fields and applications. These datasets share common traits: their elements are related to each other in a structured manner, e.g., through similarities or dependencies between data elements, and the conventional methods to deal with them are inadequate. Hence, as a solution to address this issue, the field of Graph Signal Processing (GSP) [46] has emerged to merge computational analysis of the aforementioned signals/data with the graph theory.

Biological networks have proved to be a popular application domain for graph signal processing, with recent research works focusing on the analysis of data from systems known to have a network structure, especially, the human brain. The growing number of publications studying brain activity or brain network features from a GSP perspective, namely, [47–50], indicates that these are promising applications in this field. Despite recent advances in the GSP field, however, its application for EEG processing is still in its infancy.

As described in Chapter 2, EEG is widely used to capture brain waves and due to volume conduction, unprocessed EEG signals are known to have poor spatial resolution and a rather blurred image of the brain activity is often obtained from multichannel EEG signals due to low SNR. The field of GSP makes it possible to non-invasively infer the anatomical connectivity of distinct functional regions of the cerebral cortex via utilization of a regular/weighted graph with the vertices corresponding to different EEG channels. The connectivity between the nodes of the graph (channels of the EEG headset) can also be expressed through the physical distance between the electrodes. During my research work in this thesis, I employed two GSP models for the purpose of dimensionality reduction. The explanation of each and the final results are presented in the following sections.

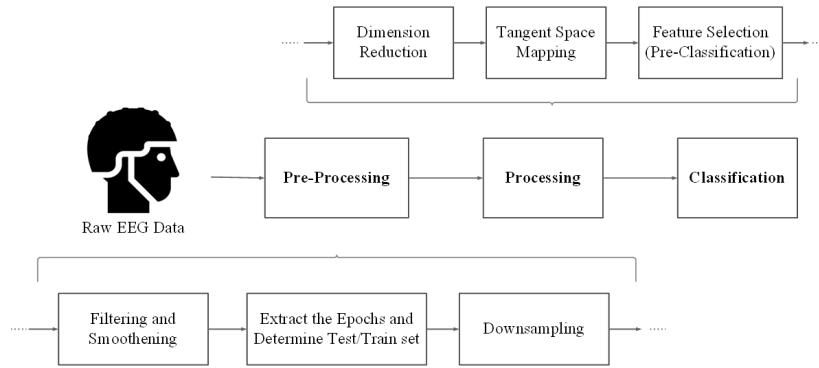


Figure 3.1: Block diagram of the proposed graph-based spatio-temporal filtering framework for brain-computer interface (GD-BCI).

## 3.2 The GD-BCI Framework

In this framework, a new graph-based approach for spatio-temporal filtering and feature selection in motor imagery brain-computer interfaces is proposed, referred to as the GD-BCI.

### 3.2.1 Framework Outline

The GD-BCI framework is realized by taking into account not only the geometrical structure of the electrode channels, but also the correlation coefficients obtained from the EEG signals. In brief, the graph similarity and Laplacian matrices are computed, the spatio-temporal filtering is performed in the graph Fourier transform domain, which is obtained in a spectral decomposition of the graph which is later applied to the EEG signals for dimensionality reduction purpose. A tangent space mapping technique is employed to project data from the Riemannian to the Euclidean domain. The PCA method is applied for selecting the most significant features for the classification task. Finally, the a linear SVM is used to solve a two-class classification problem.

We consider supervised learning from EEG signals based on the available set of EEG epochs (trials) denoted by  $\mathbf{X}_i \in \mathbb{R}^{N_{\text{ch}} \times N_t}$ , for  $(1 \leq i \leq N_{\text{Trial}})$ , where  $N_{\text{Trial}}$  is the total number of trials used for processing;  $N_{\text{ch}}$  is the number of EEG channels (electrodes), and;  $N_t$  is the number of time samples collected from each electrode in one trial. The training dataset is denoted by  $\{(\mathbf{X}_i, l_i)\}$ , for  $(1 \leq i \leq N_{\text{Trial}})$ , where  $l_i$  represents the label corresponding to the  $i$ th trial, e.g.,  $l_i$  could be “right foot” or “right hand”. Parameters rTR and rTS are the number of eigenvectors required for the

dimensionality reduction process of training and test sets, respectively. For vector  $\mathbf{X}_i$ , the sample covariance matrix is defined as

$$\mathbf{C}_i = \frac{1}{N_t - 1} (\mathbf{X}_i - \boldsymbol{\mu}_i)(\mathbf{X}_i - \boldsymbol{\mu}_i)^T, \quad (10)$$

where  $\boldsymbol{\mu}_i$  is the column-wise mean of  $\mathbf{X}_i$ . The Eq. 10 above and the equation used in Step 2 of the Algorithm 2.4.1 for computing the covariance are equivalent. The proposed GD-BCI framework consists of the following main tasks, (as shown in Fig. 3.1): (i) Pre-processing; (ii) Spatio-temporal filtering; (iii) Mapping to tangent space; (iv) Feature selection; and (v) Classification. Below and in each sub-section, the aforementioned tasks are described in details.

### 3.2.2 Pre-Processing

Before processing EEG signals for classifying MI tasks, typically, a pre-processing step is required. At this stage, bandpass filtering is applied to extract specific frequency contents of the signal, containing brain patterns regarding the motory tasks. This step is then followed by downsampling in order to reduce the load of processing for the algorithm. The signal used for processing is extracted from specific period of each trial time interval, i.e., as per the instructions given by the data collectors, the time interval during which the subject is responding to a stimulus shown. This step is conventionally done by selecting a predefined time interval after a visual cue and selecting one sample value out of  $n$  samples. In order to take into consideration the most of subject's response to each stimulus, I have tested the use of other methods for data smoothing in advance to the down-sampling step, namely, simple averaging, simple moving average, weighted moving average and moving median. More specifically, instead of choosing randomly one value, i.e., random selection, within the time interval between the two consecutive visual cues, we apply the above methods. It is experimentally, as can be viewed in Table 3.1, found that using weighted moving average as a smoothening filter provides a lower reconstruction error after dimensionality reduction stage and results in a better classification accuracy.

The results in the first column of Table 3.1 are achieved by decomposing each trial utilizing SVD, and then reconstructing the matrix of trials using 41 eigenvectors corresponding to the 41

highest eigen values, and finally, measuring the difference (error) between the original and rebuilt signals in percentage. The number 41 is chosen considering the method employed in [48] and the downsampling approach they had is presented as **RS**. The second column, shows the measurement of similarity of the original and reconstructed signals on the scale of 0–1. For this purpose, the signals are reconstructed using a certain number of eigenvectors, corresponding to a measure known as ‘expected variance’. It can be proved [51] that when the covariance of a matrix is computed and then decomposed to its eigenvectors and eigenvalues, the eigenvector with the largest eigenvalue will be the direction in which the most variance occurs. Consequently, if all the eigenvectors are put together, the entire variance in the target matrix can be explained. This means, as a surrogate for measuring the absolute value of variance explained, the simple fraction below can be an indicator of the variance covered by the reconstructed matrix. For a covariance matrix of size ( $N_{\text{ch}} \times N_{\text{ch}}$ ), Expected Variance is defined as follows

$$EV = \frac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^{N_{\text{ch}}} \lambda_i} \times 100, \quad (11)$$

where the largest  $r$  eigenvalues, and their associated eigenvectors, are put together for the purpose of reconstruction of the data matrix, and  $\lambda_i$  is an eigenvalue retrieved from the decomposition of the covariance matrix. Having that said, the second column of Table 3.1 is composed by first, determining the number of eigenvectors/eigenvalues ( $r$ ) required to cover  $EV \geq 80\%$ , and then the matrix is reconstructed in the same manner as for the first column of Table 3.1, only by  $r$  eigenvectors instead of 41. The last column of Table 3.1 shows the number of eigenvectors required to ensure  $PRD$  of 2% or less. In view of this, I use this filter in the smoothening filter block of our proposed framework. Next, I present my proposed graph-based dimensionality reduction technique.

### 3.2.3 Graph-based Spatio-Temporal Filtering

The aforementioned weighted graph  $G = (V, E, K)$  consists of a finite set  $V$  of vertices (electrode channels) and a finite set  $E$  of edges with the corresponding weights  $[k_{pq}]_{n \times n} \in K$ . The

Table 3.1: Average reconstruction error obtained using various data smoothening methods for training datasets.

|     | FF: rTR = 41 | EV: 80% | PRD: 2% |
|-----|--------------|---------|---------|
| AA  |              |         |         |
| RS  | 4.32         | 0.90    | 69      |
| SA  | 4.21         | 0.87    | 68      |
| SMA | 4.21         | 0.86    | 68      |
| WMA | 4.08         | 0.83    | 67      |
| MM  | 4.33         | 0.91    | 69      |
| AW  |              |         |         |
| RS  | 3.39         | 0.68    | 59      |
| SA  | 3.27         | 0.66    | 57      |
| SMA | 3.27         | 0.65    | 57      |
| WMA | 3.12         | 0.62    | 55      |
| MM  | 3.39         | 0.69    | 59      |
| AL  |              |         |         |
| RS  | 3.63         | 0.68    | 58      |
| SA  | 3.51         | 0.66    | 57      |
| SMA | 3.50         | 0.63    | 57      |
| WMA | 3.37         | 0.62    | 55      |
| MM  | 3.62         | 0.69    | 58      |
| AY  |              |         |         |
| RS  | 4.20         | 0.68    | 64      |
| SA  | 4.06         | 0.65    | 63      |
| SMA | 4.06         | 0.65    | 63      |
| WMA | 3.90         | 0.62    | 62      |
| MM  | 3.19         | 0.69    | 64      |
| AV  |              |         |         |
| RS  | 4.39         | 0.83    | 67      |
| SA  | 4.34         | 0.81    | 66      |
| SMA | 4.34         | 0.81    | 66      |
| WMA | 3.27         | 0.79    | 65      |
| MM  | 3.41         | 0.85    | 67      |

weights  $k_{pq}$  can be defined as a function of proximity between vertices (electrodes)  $p$  and  $q$ , as

$$\mathbf{K}_{\text{PG}} = \exp \left( -\frac{\mathbf{D}(p, q)^2}{2\sigma_d^2} \right), \quad (12)$$

where  $p$  and  $q$  are the electrode positions, and  $\mathbf{D}(p, q)$  denotes the distance between the two electrodes. In this framework, in order to take into account the dependencies of the data captured at

each electrode, a weight matrix is proposed which is a function of both the electrode proximity and correlation coefficients obtained from the EEG signals.

$$\mathbf{K}_{\text{VPG}} = \exp\left(-\frac{\mathbf{D}(p, q)^2}{2\sigma_d^2}\right) \cdot \exp\left(-\frac{(1 - \|\boldsymbol{\rho}(p, q)\|)^2}{2\sigma_\rho^2}\right), \quad (13)$$

where  $\sigma_d$  and  $\sigma_\rho$  specify the amount of exponential decay rate, and

$$\boldsymbol{\rho}(p, q) = \frac{c_{pq}}{\sqrt{c_{pp}c_{qq}}}, \quad (14)$$

obtained using the elements of the covariance matrix  $\mathbf{C}$ , given in (1). Accordingly, the degree matrix  $\mathbf{D}$  is defined using the weight matrix as

$$\mathbf{D} = \text{diag} \left\{ \sum_q k(1, q), \dots, \sum_q k(n, q) \right\}. \quad (15)$$

The graph Laplacian matrix is derived from  $\mathbf{K}$  and plays an important role in describing the underlying structure of the graph signal. The graph Laplacian and its normalized version are defined as  $\mathbf{L} = \mathbf{D}_i - \mathbf{K}$  and  $\mathbf{L}_{\text{normal}} = \mathbf{I} - \mathbf{D}_i^{-1/2} \mathbf{K} \mathbf{D}_i^{-1/2}$ , where  $\mathbf{I}$  is the identity matrix. Spectral graph theory studies the graph properties in terms of eigenvalues and eigenvectors associated with the Laplacian matrix of the graph. The set of eigenvectors of  $\mathbf{L}_{\text{normal}}$  constitute the basis function for the underlying signal defined on graph, and its eigenvalues are known as the corresponding graph frequencies. The eigen decomposition of the real and symmetric normalized Laplacian is given by

$$\mathbf{L}_{\text{normal}} = \sum_i \boldsymbol{\lambda}_i \mathbf{u}_i \mathbf{u}_i^T, \quad (16)$$

where  $\{\boldsymbol{\lambda}_i\}_{i=1,\dots,n}$  is the set of eigenvalues and  $\{\mathbf{u}_i\}$  the set of orthogonal eigenvectors used for dimension reduction.

Let  $\mathbf{U}$  contain the  $\mathbf{L}$ 's first  $r$  eigenvectors corresponding to the first  $r$  eigenvalues of  $\mathbf{L}$  sorted in ascending order. The proposed spatio-temporal filtering on graph spectral theory employs matrix

$\mathbf{U}$  to represent the EEG signals with lower number of features  $\mathbf{F}_r$  as given by

$$\mathbf{F}_r = \mathbf{U}_r^T \mathbf{X}. \quad (17)$$

It should be noted that the first  $r$  eigenvectors correspond to the first  $r$  low-frequency basis functions in graph spectral domain. To perform dimensionality reduction, the value of  $r$  can be adaptively determined for different subjects using the PRD method. The spatio-temporal filtering stage is followed by mapping the data from the existing manifold to the Euclidean space. To this end, tangent space mapping method is employed as a bridge operation to enable us to treat the data transferred to Euclidean space as vectors. This mapping method is discussed next.

### 3.2.4 Tangent Space Mapping and Feature Extraction

A ‘manifold’ is a topological space that locally resembles Euclidean space near each point. In other words, each point of an  $n$ -dimensional manifold has a neighbourhood that is homeomorphic to the Euclidean space of dimension  $n$ . Likewise, a Riemannian manifold is defined as a smooth manifold with a smooth section of the positive-definite quadratic forms on the tangent space. Having that said, it is known that the sample covariance matrices belong to the Riemannian manifold of the symmetric and positive definite matrices. However, several significant and commonly used state-of-the-art methods of machine learning, and specifically, classification techniques, such as those introduced in Chapter 2, are mostly designed to be applied to datasets in the Euclidean space. In view of this, we employ the tangent space mapping technique [52] and the vectorizing approach that follows it as the feature extraction method for GD-BCI framework and then project data to Euclidean space as vectors to proceed with the classification step.

To explain the utilized approach, let a Riemannian manifold  $\mathcal{S}(n)$  be a space of  $(n \times n)$  symmetric positive definite matrices given by  $\mathcal{S}(n) = \{\mathbf{S} \in \mathcal{M}(n), \mathbf{S}^T = \mathbf{S}\}$ , where  $\mathcal{M}(n)$  is the space of all square real matrices. The set of all the matrices is denoted as  $\mathcal{C}(n) = \{\mathbf{C} \in \mathcal{S}(n), \mathbf{u}^T \mathbf{C} \mathbf{u} > 0\}$ , which is not Euclidean. Tangent space mapping provides a Euclidean tangent space,  $\mathbf{T}_Q \mathcal{C}(n)$  at the point  $Q$ , which approximates the aforementioned Riemannian manifold through the following steps:

- Compute the set of sample covariance matrices for each trial as given in 10.
- Compute the mean Riemannian distance as

$$\bar{\mathbf{C}} = \frac{1}{N_{\text{Trial}}} \sum_{i=1}^{N_{\text{Trial}}} \mathbf{C}_i. \quad (18)$$

- Compute the map  $\mathbf{s}_i$  from  $\mathbf{C}$  to  $\mathbf{T}_Q \mathbf{C}(n)$  as

$$\mathbf{s}_i = \text{Upper} \left( \log \left( \bar{\mathbf{C}}^{\frac{-1}{2}} \mathbf{C}_i \bar{\mathbf{C}}^{\frac{-1}{2}} \right) \right), \quad (19)$$

where *Upper* is used to weigh the upper triangular half of a matrix and vectorize it. In particular, it assigns 1 as the weight for main diagonal and  $\sqrt{2}$  for off-diagonal entries. The resulting feature vectors are further trimmed and the most relevant ones are selected for classification purpose.

As per described in Chapter 2, the high dimensional features may lead to poor classification performance. This is due to the fact that large number of irrelevant features not only degrades the generalization of the model, but also imposes computational cost. In view of this and in order to determine the most significant feature vectors of  $\mathbf{s}_i$  in the tangent space mapping process, which are maximally related to the desired classes, we use a graph-based feature selection method similar to what is given in 17. To this end, a weighted graph is built from  $\mathbf{s}_i$  and the first 10 eigenvectors of the corresponding Laplacian matrix are selected. The elected eigenvectors are then fed into the classifier as its input. In addition, for  $PRD = 2\%$ , eigenvectors are adaptively selected for each subject. It should be noted that using the PRD method, one can adaptively compact feature vectors for a better classification result.

### 3.2.5 Classification

In order to classify the selected feature vectors as representative of right hand or right foot MIs, we employ the support vector machine (SVM) algorithm. The SVM uses training feature vectors to learn a decision boundary that separates these two classes by projecting data into a higher dimensional space using a kernel function. Once the decision boundary is learned, the SVM determines the class membership of a newly-observed feature vector according to the side of boundary that the

vector falls.

### 3.2.6 Simulations

The proposed framework is benchmarked on the dataset IVa from the BCI competition III taken from <http://www.bbci.de/competition/iii/>. The EEG positioning is based on 10–20 standard system. The dataset is composed of EEG recordings of 118 electrodes. The experiment is a classical cue-based MI paradigm in which each of 5 subjects, namely, AA, AL, AV, AW, and AY, perform 280 trials of right hand and right foot MIs. In the pre-processing step, the EEG signals are bandpass filtered in the frequency band [8 – 30] Hz by a 5th order *Butterworth* filter. The time interval is restricted to the segment located from 0.5s to 4s after the cue. The weighted moving average filter is then applied to smoothen the data.

In order to obtain the most significant features for each subject, PRD is used to adaptively determine the required number of eigenvectors from which data can be reconstructed with a predefined error. The corresponding size of datasets for different subjects are given in Table 3.2.

Table 3.2 illustrates the classification accuracy and its corresponding standard deviation averaged over 400 runs, obtained using the proposed GD-BCI method using two graph construction method, namely, physical graph (PG) and value-physical graph (VPG). It is seen from this table that proposed method using GD-BCI-VPG outperforms its GD-BCI-PG counterpart by almost 10%. This is due to the fact that the GD-BCI-VPG is built using the electrode channel proximity (PG case) as well as the correlation coefficients of the EEG signals, i.e., spatio-temporal filtering of the EEG signals.

We now compare the performance of the proposed GD-BCI method to that obtained from the other existing methods, in terms of the classification accuracy. Table 3.3 gives the comparison results of the proposed method using VPG and PG and that provided by [48] and [53], when constant or adaptive number of features are selected. It is seen from this table that the proposed method provides higher classification accuracy for various subjects as compared to those yielded by [48] and [53]. In addition, the standard deviation of the classification accuracy obtained using the proposed method is lower than that provided by [48].

Table 3.2: Accuracy performance for predicting two classes and the corresponding standard deviation obtained using the proposed GD-BCI framework with two graph construction methods: PG and VPG

| Subject AA (168 Train + 112 Test) |            |                |            |
|-----------------------------------|------------|----------------|------------|
| PRD<2.5%                          |            | PRD<4.15%      |            |
| rTR=rTS=60                        |            | rTR=41, rTS=43 |            |
| VPG                               | PG         | VPG            | PG         |
| 82.90±1.23                        | 80.99±1.64 | 76.29±1.61     | 74.38±2.14 |
| Subject AL (224 Train + 56 Test)  |            |                |            |
| PRD2%                             |            | PRD<3.5%       |            |
| rTR=59, rTS=56                    |            | rTR=41, rTS=39 |            |
| VPG                               | PG         | VPG            | PG         |
| 97.37±0.30                        | 97.62±0.23 | 97.62±0.27     | 97.38±0.31 |
| Subject AW (56 Train + 224 Test)  |            |                |            |
| PRD<2%                            |            | PRD<3.3%       |            |
| rTR=rTS=58                        |            | rTR=rTS=41     |            |
| VPG                               | PG         | VPG            | PG         |
| 92.58±2.82                        | 91.30±2.75 | 93.70±2.21     | 90.83±3.14 |
| Subject AV (84 Train + 196 Test)  |            |                |            |
| PRD<2%                            |            | PRD<4.29%      |            |
| rTR=rTS=66                        |            | rTR=41, rTS=42 |            |
| VPG                               | PG         | VPG            | PG         |
| 65.79±2.97                        | 66.87±2.69 | 66.99±2.83     | 64.21±3.32 |
| Subject AY (28 Train + 252 Test)  |            |                |            |
| PRD<2%                            |            | PRD<4%         |            |
| rTR=rTS=66                        |            | rTR=41, rTS=38 |            |
| VPG                               | PG         | VPG            | PG         |
| 83.14±5.76                        | 68.82±5.67 | 80.29±4.64     | 84.32±4.61 |

### 3.2.7 Conclusion

In this framework, a new dimensionality reduction and feature selection technique is proposed for analyzing EEG signals obtained from an EEG-based BCI during MI tasks. The GD-BCI has been established by leveraging the recent advances in the field of graph signal processing. The proposed method is composed of an efficient graph-based dimensionality reduction technique followed by tangent space mapping of the EEG signals to the Euclidean space and a graph-based feature selection to. Experiments have been conducted on a set of EEG signals obtained from the BCI

Table 3.3: Performance comparison of the proposed GD-BCI method in two-class classification problem with that provided by [48] and [53].

|         | GD-BCI-PG     |             | GD-BCI-VPG    |             | [48]        | [53]     |
|---------|---------------|-------------|---------------|-------------|-------------|----------|
|         | PRD: rTR = 41 | PRD: 2%     | PRD: rTR = 41 | PRD 2%      | rTR = 41    | rTR = 10 |
| AA      | 74.38± 2.14   | 80.99± 1.64 | 76.29± 1.61   | 82.90± 1.23 | 81.43± 10.9 | 74.1     |
| AL      | 97.38± 0.31   | 97.62± 0.23 | 97.62± 0.27   | 97.37± 0.30 | 97.50± 2.98 | 98.2     |
| AW      | 90.83± 3.14   | 91.30± 2.75 | 93.70± 2.21   | 92.58± 2.82 | 98.57± 0.79 | 77.7     |
| AV      | 64.21± 3.32   | 66.87± 2.69 | 66.99± 2.83   | 65.79± 2.97 | 69.29± 5.56 | 59.2     |
| AY      | 84.32± 4.61   | 68.82± 5.67 | 80.29± 4.64   | 83.14± 5.76 | 93.93± 4.30 | 80.6     |
| Average | 82.22± 2.70   | 81.12± 2.60 | 82.98± 2.31   | 84.35± 2.69 | 88.14± 4.90 | 78.0     |

competition. The results have shown that the proposed method produces encouraging results providing high recognition accuracy for two-class classification. However, all these great advantages come with an undesirable price: to determine the constants  $\sigma_d$  and  $\sigma_\rho$  to configure the best-suited graph for each subject’s data, an exhaustive search is required. The trial and error solution to find the proper  $\sigma_d$  and  $\sigma_\rho$  is not preferable in a real-world application. Therefore, I sought for an adaptive and systematic approach as explained in the next section, to adjust the graph for each subject’s dataset without an excessive trial and error involved.

### 3.3 GDR-BCI: Dimensionality Reduction of EEG Signals via Functional Clustering and Total Variation Measure

In this framework, similar to the one proposed in GD-BCI, the observations obtained from the EEG channels, a non-uniformly distributed sensor field, are taken into account in a manner that a representation graph is formed using geographical distances between sensors to form connectivity neighborhoods. However, by capitalizing the fact that functionality of different connectivity neighborhoods varies based on the intensity of the performed activity and concentration level of the subject, an initial functional clustering of EEG electrodes was formed by designing a separate adjacency matrix for each identified functional cluster.

### 3.3.1 Framework outline

This framework contributes to detection of main EEG electrodes that capture the intention of the subject due to the exposed stimulus. This purpose is achieved through an uncomplicated yet straightforward and systematic method which separates the EEG electrodes in a desired number of clusters. These clusters are formed under a biological constraint and then collapsed in an intelligent fashion to identify the two clusters that best spread over the active part of the brain during the motor imagery task. In order to better perceive the impact of our proposed GSP-based approach for processing EEG signals, widely used feature extraction/classification algorithms are coupled with the proposed framework deliberately. The incentive of the proposed method roots back in the tendency of reducing the dimension of the dataset before processing it and without losing crucial information, and also, in order to address challenges that previously proposed graph-based methods typically face, i.e., formerly graph transforms applied to EEG datasets for dimensionality reduction were formed empirically through trial and error for determining/tuning the constants required for selecting the number of neighbors of each node. Via a change in the perspective of graph-based EEG signal processing, I rather focus on deciding which nodes are capturing the intended activity, and thereafter, are showing most relative activity to the purpose of the BCI, at the time of the experiment. It is important to note that this approach is subject-adaptive, technique-adaptive, applicable to multiclass experiments, and has practical advantages for experiments with limited means of measurement at hand.

The electrodes of EEG headsets are often distributed over the subject's skull in a consistent geometrical structure, i.e., the arrangement or location of the electrodes of EEG recording device. Hence, the neighborhood of each channel and their distances from other channels can be treated as the required information for constructing a graph signal.

The GD-BCI framework involves a graph defined as a triplet  $G = (V, E, K)$ , in which  $K$  is a weighted adjacency matrix. Intuitively speaking, each element of  $K$  represents the weight of the corresponding element in  $E$ . Matrix  $K$  is formulated as an exponential term including the Euclidean distance as in (12) where as stated previously, the Euclidean distance between any two

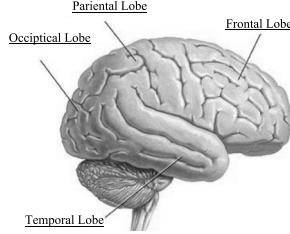


Figure 3.2: Functional clustering of human brain [?].

channels is computed via the following equation.

$$D_{pq} = \sqrt{{X_{pq}}^2 + {Y_{pq}}^2} \quad (20)$$

The matrix  $\mathbf{K}_{\text{VPG}}$  takes into account the physical distance, as well as the correlation of the data in each trial [54]. However, the main obstacle on the way of constructing  $\mathbf{K}_{\text{VPG}}$  above, is how to determine the threshold the constants  $\sigma_d$  and  $\sigma_\rho$ . During the previous framework, the neighborhood and the structure of the graph is highly sensitive to the values of  $\sigma_d$  and  $\sigma_\rho$ . This challenge inspired us to investigate an alternative approach to define the graph and reduce the dimension of the EEG data.

Inspired by Reference [55], we decided to define our graph  $G$  as a twin of  $\mathbf{G} = (\mathbf{V}; \mathbf{A})$  in which the matrix of nodes, i.e.,  $\mathbf{V}$ , is similarly defined as the set of EEG electrodes. However, the second term  $\mathbf{A}$  is the adjacency matrix. In this work,  $\mathbf{A}$  is constructed via the Euclidian distances of electrodes from each other as described in detailed further on.

### 3.3.2 Defining the Adjacency Matrix

As is described in Reference [55], in the case of measurements from a non-uniformly distributed sensor field (in our context, EEG channels), data recorded from each sensor is a separate time series. In their example, a representation graph was constructed using geographical distances between sensors and considering  $L$  nearest sensors as the connectivity neighborhood.

However, motivated by existence of different regions of the human brain with different functionality as shown in Fig. 3.3.2, and the fact that these functions might have various intensities of activity due to the status of the person and her/his concentration, I formed the idea of grouping the

electrodes and proposing an individual corresponding adjacency matrix for each cluster independent from others. In this fashion, the problem of determining/extracting the constants ( $\sigma_d$ , and  $\sigma_p$ ) through trial and error is tackled, as the biological function of brain regions would be the constraint of the neighborhood for forming the graphs.

Fig. 3.3 illustrates the sparsity of the electrodes (projected onto 2-dimension) of the EEG headset used to collect the dataset and the groups (clusters) assigned to this structure. Number of electrodes considered in each cluster is defined as:

- Cluster{1} = [1 2 3 4 5 6 7 8 9];
- Cluster{2} = [11 12 17 18 19 25 26 27 28 35 36 37 45 46];
- Cluster{3} = [10 14 15 16 23 24 33 34];
- Cluster{4} = [13 20 21 22 29 30 38 39];
- Cluster{5} = [31 32 42 50 59 67 68];
- Cluster{6} = [40 41 49 56 58 76 77];
- Cluster{7} = [43 44 51 52 60 61 69 70];
- Cluster{8} = [47 48 56 57 64 65 74 75];
- Cluster{9} = [53 54 55 62 63 71 72 73 81 82 90 91 92 99 100];
- Cluster{10} = [78 79 80 86 87 88 89 98];
- Cluster{11} = [83 84 85 93 94 95 96 101];
- Cluster{12} = [97 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118];

whereas each cluster is detached and defined separately, each of them can be considered as an isolated graph,  $G_{(c_i)} = \{V_{(c_i)}, A_{(c_i)}\}$ . The adjacency matrix for the  $i$ -th cluster is formed as a square matrix of Euclidean distances the nodes of the  $i$ -th cluster see each other from. It is worth mentioning that at the beginning of this work, a general adjacency matrix was considered to model the connectivity of each electrode to all others in the structure of the headset. The following steps

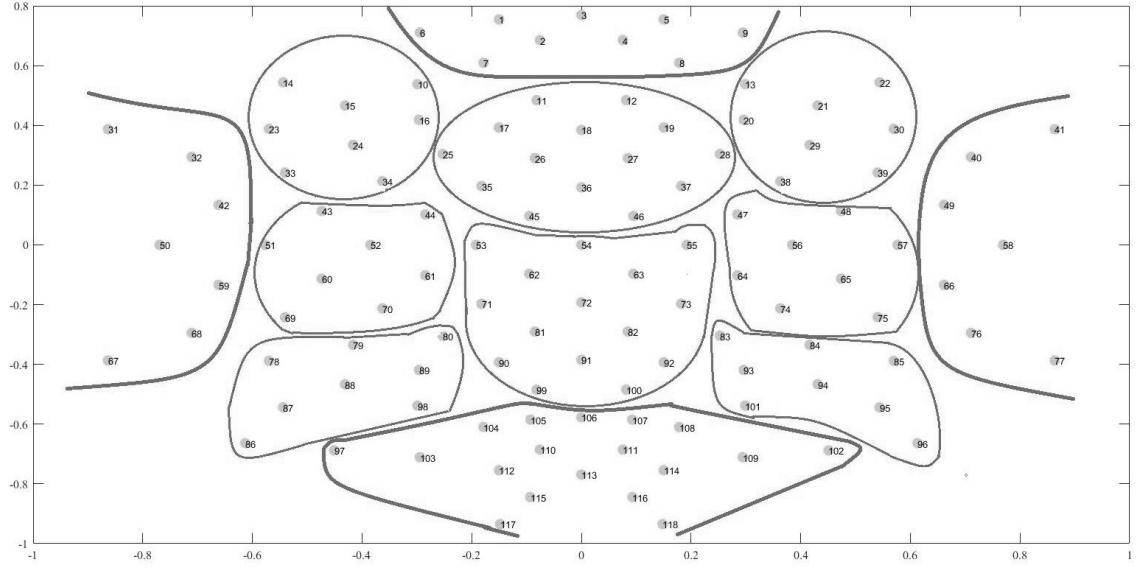


Figure 3.3: Sparsity pattern of the EEG electrodes (2-dimensional projection).

carried out throughout the rest of the framework were similar to the clustering approach, however, the accuracies were not satisfying.

### 3.3.3 Total Variation Graph and Selection of best Clusters: Dimensionality Reduction

In Reference [56] an interesting concept is introduced referred to as the Total Variation Graph. This concept denotes the signal variations on graphs and is formulated as follows.

$$\text{TV}_G(\mathcal{S}) = \frac{1}{\|\mathcal{S}\|_2^2} \left\| \mathcal{S} - \frac{1}{|\lambda_{\max}|} \mathbf{A} \mathcal{S} \right\|_2^2. \quad (21)$$

Intuitively speaking, Eq. (21) evaluates the variation of the graph from a time sample to another. Term  $\mathcal{S}$  is the vector of values produced from all the nodes of the graph at a certain time sample; Term  $\mathbf{A}$  is the aforementioned adjacency matrix, and;  $|\lambda_{\max}|$  is the largest-magnitude eigenvalue of matrix  $\mathbf{A}$ , which satisfies the condition  $|\lambda_{\max}| \geq |\lambda_m|$ , for  $(0 \leq m \leq N_{\text{ch}} - 1)$ . We employed this concept to extract the features and feed a LDA classifier with the features generated via Total Variation Graph. In other words, the set of data collected from each cluster of electrodes was separately fed to Total Variation Graph model (Eq. (21)), and the vector of features for each trial

(epoch) would be from the size of (number of time samples  $\times$  1).

The same sub-steps of feature extraction and classification are carried out for all the clusters, and the accuracies of the classification are then compared to each other. The two clusters with maximum accuracy are selected and their electrodes are used for the final feature extraction (through CSP) and classification with three different kernels of classifiers.

It is important to note that this step of this framework has two main superiorities over similar works. First, the dimensionality reduction is subject-adaptive and respects the brain plasticity of subjects, by finding the two clusters matching the most active regions of the brain, in regards to the task the subject was asked to do. Therefore, the proposed approach is not limited to motor imagery and can be employed within the processing kernel of BCIs using other techniques as well. The second advantage is that the proposed approach can be taken to another level of implementation in practical experiments. In the case that the EEG cap would have the location of electrodes, yet not all/enough electrodes would be available to record the data, using steps outlined in Sub-sections 3.3.2 and 3.3.3, the researchers can find out which regions of the brain are active during the time that the subject is performing the tasks. Then, they can place the electrodes in other regions close to the active area and increase the accuracy of their data collection. This advantage adds flexibility to practical experiments regardless of the technique being used.

### 3.3.4 Feature Extraction and Classification

As mentioned previously, the feature extraction method is a conventional CSP. The reason for using this technique of feature extraction is the fact that this method is one of the most known and well-regarded techniques of extracting features from 2-class datasets in EEG signal processing. The target of this framework is to keep the non-graph part of the approach simple and as common as possible, so that the effect of the previous two steps could be observed and validated more clearly. As for the classification, two well-known classifiers, i.e., LDA and QDA are used with 5-fold validation technique.

### 3.3.5 Simulations

For the purpose of evaluating the proposed graph-based dimensionality reduction (GRD) framework, the dataset IVa from BCI Competition III was employed. This dataset includes the information regarding the location of electrodes with respect to the center of the scalp. An order five Butterworth bandpass filter is used in the pre-processing step to filter each raw of the dataset and then a moving average function is utilized to smoothen the values. Next, the time series were down-sampled by the factor of 10, meaning that only one time sample is selected out of each batch of 10 samples.

As for the feature extraction section, we considered two different scenarios of using two or four eigenvectors to construct the whitening matrix, and the impact of the increase in the size of whitening matrix can be easily noticed by comparing the table of results. Tables 5.1(a)-(c) illustrate the accuracy of classifications of features extracted by a whitening matrix constructed by two eigenvectors, i.e., the ones corresponding to the max and min eigenvalues, with two different classifiers. Table 3.4(a) illustrates the classification accuracies obtained with 100 training samples (50 per class). Table 3.4(b) shows the accuracies for the classifiers trained with only 60 samples (30 per class), which is suitable for BCIs with limited number of available training trials. Table 3.4(c) shows the classification accuracies with only 200 samples (100 per class), which is a conventional number of training trials for this dataset. Based on these results, it is observed that the proposed approach is significantly and effective for motor imagery classification problem. Besides, improved accuracies through both classifiers clearly reflect the positive impact of our proposed GDR framework. Tables 3.5(a)-(c) exhibit the accuracies for the scenarios of extracting features with a whitening matrix of 4 eigenvectors, where other items are kept similar to the scenario in Table 3.4. These results again elaborate the effectiveness of the proposed GDR framework and illustrates its great potential for improving the overall motor imagery classification accuracy when coupled with more complex feature extraction methods.

### **3.3.6 Conclusion**

As a novel graph-based dimensionality reduction, (GDR) framework for processing EEG signals is proposed based on functional clustering of EEG electrodes and a collapsing step via total variation on graphs. The proposed methodology initially forms connectivity neighbourhoods and constructs a global block-diagonal graph representation of the EEG channels. By capitalizing on the fact that functionality of different connectivity neighborhoods varies, a collapsing step is introduced based on total variation measures on graph, to reduce the overall graph model into two functional clusters. The experimental results based on Dataset IVa from BCI Competition III show that the proposed method can provide higher classification accuracy as compared to the other existing methods.

## **3.4 Summary**

Throughout this chapter, the methods and techniques suggested in graph theory and graph signal processing were applied to the processing core of an MI EEG-based BCI system. Two frameworks were proposed, GD-BCI and GDR-BCI and the effectivity of them is well-elaborated via the results of the implementation. This completes my research work on developing the theoretical approaches to improve the performance of MI EEG-based BCI systems.

Table 3.4: Accuracy comparison of the proposed GDR framework coupled with two different classifiers and with two CSP features.

(a)

| CSP with 2 Eigen<br>Vecors |                  | With 100 Training Trials |      |       |       |       |      |
|----------------------------|------------------|--------------------------|------|-------|-------|-------|------|
|                            |                  | LDA                      |      |       | QDA   |       |      |
|                            |                  | Train                    | Mean | STDV  | Best  | Train | Mean |
| AY                         | Conventional CSP | 84.14                    | 0.43 | 85.00 | 53.89 | 83.98 | 1.27 |
|                            | GDR + CSP        | 85.00                    | 1.00 | 87.00 | 81.00 | 81.20 | 1.76 |
| AW                         | Conventional CSP | 91.84                    | 0.43 | 93.00 | 66.67 | 92.06 | 0.72 |
|                            | GDR + CSP        | 89.47                    | 0.80 | 91.00 | 81.11 | 89.82 | 1.12 |
| AV                         | Conventional CSP | 76.08                    | 1.35 | 80.00 | 52.22 | 67.86 | 0.20 |
|                            | GDR + CSP        | 72.30                    | 0.75 | 74.00 | 64.44 | 72.72 | 0.17 |
| AL                         | Conventional CSP | 94.55                    | 0.55 | 95.00 | 86.11 | 93.94 | 0.79 |
|                            | GDR + CSP        | 86.14                    | 0.99 | 89.00 | 91.11 | 86.75 | 1.32 |
| AA                         | Conventional CSP | 78.54                    | 1.14 | 82.00 | 51.11 | 73.43 | 1.25 |
|                            | GDR + CSP        | 75.60                    | 1.09 | 79.00 | 64.44 | 75.21 | 1.77 |

|         |                  |       |         |                  |       |
|---------|------------------|-------|---------|------------------|-------|
| Average | Conventional CSP | 62.00 | Average | Conventional CSP | 63.11 |
|         | GDR + CSP        | 76.42 |         | GDR + CSP        | 78.00 |

(b)

| CSP with 2 Eigen<br>Vecors |                  | With 60 Training Trials |      |       |       |       |      |
|----------------------------|------------------|-------------------------|------|-------|-------|-------|------|
|                            |                  | LDA                     |      |       | QDA   |       |      |
|                            |                  | Train                   | Mean | STDV  | Best  | Train | Mean |
| AY                         | Conventional CSP | 88.71                   | 1.48 | 91.67 | 53.64 | 88.52 | 1.36 |
|                            | GDR + CSP        | 59.95                   | 2.47 | 65.00 | 50.91 | 54.68 | 1.45 |
| AW                         | Conventional CSP | 85.86                   | 1.87 | 90.00 | 47.27 | 77.48 | 1.37 |
|                            | GDR + CSP        | 82.67                   | 0.82 | 83.30 | 55.91 | 81.58 | 1.68 |
| AV                         | Conventional CSP | 82.69                   | 1.70 | 86.67 | 56.36 | 74.99 | 2.95 |
|                            | GDR + CSP        | 82.67                   | 0.82 | 83.30 | 55.91 | 62.66 | 1.88 |
| AL                         | Conventional CSP | 80.20                   | 1.52 | 85.00 | 49.55 | 79.12 | 2.16 |
|                            | GDR + CSP        | 85.97                   | 1.01 | 88.33 | 89.55 | 90.94 | 1.34 |
| AA                         | Conventional CSP | 81.32                   | 1.49 | 85.00 | 56.36 | 76.03 | 1.34 |
|                            | GDR + CSP        | 78.71                   | 1.58 | 81.67 | 71.36 | 75.99 | 1.54 |

|         |                  |       |         |                  |       |
|---------|------------------|-------|---------|------------------|-------|
| Average | Conventional CSP | 52.64 | Average | Conventional CSP | 54.09 |
|         | GDR + CSP        | 64.73 |         | GDR + CSP        | 63.36 |

(c)

| CSP with 2 Eigen<br>Vecors |                  | With 200 Training Trials |      |       |       |       |      |
|----------------------------|------------------|--------------------------|------|-------|-------|-------|------|
|                            |                  | LDA                      |      |       | QDA   |       |      |
|                            |                  | Train                    | Mean | STDV  | Best  | Train | Mean |
| AY                         | Conventional CSP | 74.66                    | 0.44 | 76.00 | 56.25 | 71.55 | 0.75 |
|                            | GDR + CSP        | 85.96                    | 0.69 | 88.00 | 86.25 | 82.09 | 0.28 |
| AW                         | Conventional CSP | 67.96                    | 0.52 | 70.00 | 45.00 | 68.20 | 1.32 |
|                            | GDR + CSP        | 88.01                    | 0.42 | 89.00 | 88.75 | 88.21 | 0.44 |
| AV                         | Conventional CSP | 66.56                    | 0.88 | 68.50 | 50.00 | 55.74 | 1.03 |
|                            | GDR + CSP        | 72.31                    | 0.88 | 74.50 | 65.00 | 71.25 | 1.48 |
| AL                         | Conventional CSP | 93.81                    | 0.43 | 94.50 | 90.00 | 91.53 | 0.56 |
|                            | GDR + CSP        | 91.84                    | 0.32 | 92.50 | 93.75 | 92.42 | 0.21 |
| AA                         | Conventional CSP | 78.97                    | 0.64 | 81.50 | 57.50 | 65.69 | 0.67 |
|                            | GDR + CSP        | 78.94                    | 0.81 | 81.00 | 68.75 | 78.31 | 0.95 |

|         |                  |       |         |                  |       |
|---------|------------------|-------|---------|------------------|-------|
| Average | Conventional CSP | 59.75 | Average | Conventional CSP | 59.25 |
|         | GDR + CSP        | 80.50 |         | GDR + CSP        | 80.25 |

Table 3.5: Similar to Table 5.1 except that four CSP features are utilized.  
(a)

| CSP with 4 Eigen<br>Vecors |                  | With 100 Training Trials |      |       |       |       |       |
|----------------------------|------------------|--------------------------|------|-------|-------|-------|-------|
|                            |                  | LDA                      |      |       | QDA   |       |       |
|                            |                  | Train                    |      | Test  | Train |       | Test  |
| AY                         | Conventional CSP | Mean                     | STDV | Best  | Mean  | STDV  | Best  |
|                            | GDR + CSP        | 95.08                    | 0.88 | 97.00 | 47.78 | 94.46 | 1.00  |
| AW                         | Conventional CSP | 83.02                    | 1.40 | 86.00 | 85.00 | 83.80 | 1.35  |
|                            | GDR + CSP        | 96.04                    | 0.85 | 98.00 | 51.67 | 96.32 | 0.66  |
| AV                         | Conventional CSP | 94.18                    | 1.17 | 96.00 | 78.33 | 94.17 | 0.7.  |
|                            | GDR + CSP        | 78.91                    | 1.63 | 84.00 | 51.11 | 78.46 | 2.18  |
| AL                         | Conventional CSP | 70.79                    | 1.66 | 76.00 | 65.00 | 69.59 | 2.15  |
|                            | GDR + CSP        | 96.90                    | 0.36 | 98.00 | 53.89 | 97.38 | 1.15  |
| AA                         | Conventional CSP | 92.35                    | 0.58 | 94.00 | 93.89 | 90.57 | 1.06  |
|                            | GDR + CSP        | 86.80                    | 1.02 | 89.00 | 49.44 | 85.49 | 1.31  |
| <b>Average</b>             |                  | 81.39                    | 0.92 | 84.00 | 68.89 | 81.11 | 1.12  |
|                            |                  | 50.78                    |      |       | 84.00 | 84.00 | 65.56 |
|                            |                  | GDR + CSP                |      |       | 78.22 | 78.00 |       |

(b)

| CSP with 4 Eigen<br>Vecors |                  | With 60 Training Trials |      |       |       |       |       |
|----------------------------|------------------|-------------------------|------|-------|-------|-------|-------|
|                            |                  | LDA                     |      |       | QDA   |       |       |
|                            |                  | Train                   |      | Test  | Train |       | Test  |
| AY                         | Conventional CSP | Mean                    | STDV | Best  | Mean  | STDV  | Best  |
|                            | GDR + CSP        | 94.43                   | 0.86 | 96.67 | 45.45 | 94.93 | 0.33  |
| AW                         | Conventional CSP | 84.02                   | 2.17 | 90.00 | 82.27 | 82.60 | 2.65  |
|                            | GDR + CSP        | 93.96                   | 0.83 | 95.00 | 47.73 | 91.08 | 1.41  |
| AV                         | Conventional CSP | 83.66                   | 2.25 | 88.33 | 58.64 | 87.88 | 1.36  |
|                            | GDR + CSP        | 85.49                   | 2.48 | 91.67 | 52.73 | 87.62 | 2.64  |
| AL                         | Conventional CSP | 75.06                   | 1.86 | 78.33 | 65.00 | 70.38 | 2.18  |
|                            | GDR + CSP        | 90.78                   | 1.59 | 95.00 | 45.00 | 87.92 | 3.20  |
| AA                         | Conventional CSP | 94.60                   | 0.89 | 95.00 | 89.55 | 92.49 | 1.75  |
|                            | GDR + CSP        | 86.89                   | 1.82 | 90.00 | 47.27 | 84.49 | 1.49  |
| <b>Average</b>             |                  | 71.97                   | 2.53 | 78.33 | 76.36 | 69.54 | 2.70  |
|                            |                  | 47.64                   |      |       | 75.00 | 75.00 | 73.64 |
|                            |                  | GDR + CSP               |      |       | 74.36 | 70.91 |       |

(c)

| CSP with 4 Eigen<br>Vecors |                  | With 200 Training Trials |      |       |       |       |       |
|----------------------------|------------------|--------------------------|------|-------|-------|-------|-------|
|                            |                  | LDA                      |      |       | QDA   |       |       |
|                            |                  | Train                    |      | Test  | Train |       | Test  |
| AY                         | Conventional CSP | Mean                     | STDV | Best  | Mean  | STDV  | Best  |
|                            | GDR + CSP        | 94.49                    | 0.67 | 96.50 | 45.00 | 93.12 | 0.71  |
| AW                         | Conventional CSP | 86.34                    | 0.73 | 88.00 | 86.25 | 86.00 | 0.62  |
|                            | GDR + CSP        | 92.76                    | 0.35 | 93.50 | 57.50 | 91.60 | 0.62  |
| AV                         | Conventional CSP | 88.34                    | 0.52 | 89.50 | 80.00 | 86.58 | 0.97  |
|                            | GDR + CSP        | 73.49                    | 1.33 | 76.50 | 51.25 | 67.85 | 1.97  |
| AL                         | Conventional CSP | 76.76                    | 0.98 | 79.00 | 67.50 | 74.63 | 1.30  |
|                            | GDR + CSP        | 94.91                    | 0.49 | 96.00 | 50.00 | 93.05 | 0.29  |
| AA                         | Conventional CSP | 92.04                    | 0.19 | 92.50 | 93.75 | 92.31 | 0.60  |
|                            | GDR + CSP        | 84.19                    | 0.72 | 86.00 | 50.00 | 75.65 | 0.71  |
| <b>Average</b>             |                  | 81.27                    | 0.59 | 83.00 | 73.75 | 79.40 | 0.91  |
|                            |                  | 50.75                    |      |       | 81.50 | 81.50 | 72.50 |
|                            |                  | GDR + CSP                |      |       | 80.25 | 80.75 |       |

## **Chapter 4**

# **Practical Solutions to Improve the Performance of EEG-based BCI systems**

As stated in Chapter 2, the techniques and methods proposed in theory to serve the end goal of advancing the BCI systems, as well-structured and proven to be highly effective as can be, are not the sole approach mitigate the real-world problems that BCI systems struggle with during implementation and utilization. Therefore, should a BCI researcher desire to truly move teh edges of knowledge in thsi field, her/his primary task is to get involved in practical experiments and implementations of BCIs, however basic and pilot, while developing the theoretical methods and more complicated approaches of these systems. This way, challenges such as the comfort of the subject/patient/end-user of the BCI, the ethical issues, the feasibility and practicability of a designed system, and so many other similar aspects of implementing a BCI system will have a real sense and substantial meaning to the aforesaid researcher. Motivated by this outlook, this chapter is allocated to the practical experiments I carried out for my thesis and the solutions provided and tested. The first solution is “Progressive Fusion of Multi-rate MI Classification for BCIs” to address the issues arised in the case of limited number of training trials at hand. The second solution is “Improving the Accuracy of MI EEG-based BCIs Through Trimming the Epochs” to readjust the recorded epochs in a manner that most informative parts of the signals are extracted and the segments of the epochs which do not include the respose of the subjects to the stimuli would be discarded.

## 4.1 Progressive Fusion of Multi-rate MI Classification for BCIs

The work presents a practical implementation of an EEG-based BCI system developed based on the Emotiv EPOC headset [57]. The focus of this approach is on development/implementation of a synchronous BCI system, i.e., EEG signals are analyzed during pre-defined periods of time, initiated by an interface. The objective is to research promising ideas in the design, development, and implementation of signal processing technologies that contribute to the advancement of robust, real-time, and adaptive EEG-based BCI system. In particular, the developed EEG-based framework consists of two filters running in parallel namely: (i) *The Progressive Filter*: An efficient filter that performs both feature extraction and classification (CSP is employed) steps based on the set of all arriving epochs to re-train progressively over time. (ii) *The Active Filter*: A simplified CSP-based feature extraction approach running online based on pre-trained classifiers, i.e., a lighter version of the Progressive Filter that runs faster than its counterpart. After each trial and during the rest period, the Active Filter produces the classification results and communicates its decision to the next processing module in the BCI pipeline (e.g., a connected Arduino microcontroller). In the meantime, the Progressive Filter incorporates new EEG epochs to adopt and re-train. Once the Progressive Filter is trained, it checks its output with the Active Filter and if the two are not in consensus, it updates the models of the Active Filter, i.e., the coupling of the two filters to improve active classification performance. The proposed framework is evaluated both based on dataset IVa from the BCI competition III, and through real data collected via the Emotiv Epcoc headset.

### 4.1.1 Progressive Multi-rate MI Classification Outline

In an EEG-based BCI, as explained previously, several channels (sensors) are used to pick the potential differences detected at the scalp produced from combined activity of several neural impulses. Once the signal is recorded, the next steps involve pre-processing the recorded data, extracting discriminating features, and training a classifier to detect different MI classes, similar to the aforementioned approaches. The following framework is motivated by the fact that throughout a session of EEG recording, as a cognitive process, subjects become habituated towards the stimuli

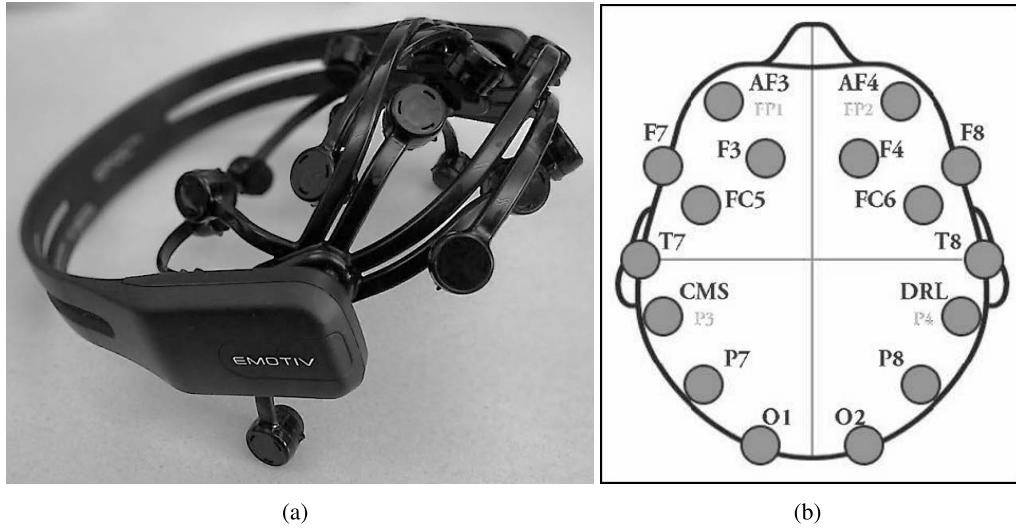


Figure 4.1: (a) Emotiv Epoç headset. (b) Electrode placement and activation.

triggering their response. Therefore, in the case that there is a limited number of training trials available to initiate the classification model, the BCI system fails to adapt itself to the ongoing changes happening in the patterns of data. Having that said, the proposed framework consists of two parallel pathways for classification, among which, one of them functions as an observer. The observer takes action once the recorded epochs by can improve the performance of the previous main classifier. The algorithm by which this idea is implemented is described as follows.

The proposed progressive and multi-rate classification framework is developed based on the CSP feature extraction method and consists of two filters, referred to as the Active Filter and the Progressive Filter, which are partially coupled at the consensus epochs based on the individual classification results. The Active Filter is a simple MI classification feature extraction method (e.g., with two eigenvectors taken into account for the construction of whitening matrix) which forms the CSP features and uses a pre-trained classification model to assign them to the two MI classes. The Active Filter produces results at the end of each epoch while the Progressive Filter uses several consecutive epochs for re-training and adaptation. In other words, after each trial and during the rest period, the Active Filter produces the classification results and communicates its decision to the next processing module in the BCI pipeline (e.g., a connected Arduino microcontroller). In the meantime, the Progressive Filter, in an offline fashion, performs the processing, adaptation, and

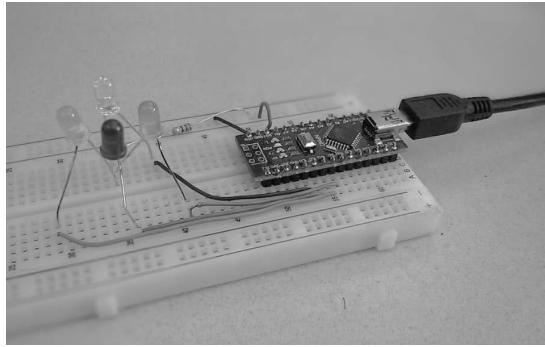


Figure 4.2: Arduino microcontroller used to turn BCI classifications into external actions like moving motors or controlling LEDs.

re-training tasks and it couples its models with the Active Filter in case the two filter are not in consensus. This completes the outline of the proposed multi-rate MI classification framework, next, the development of the experimental setup and report results obtained from implementation of the proposed framework will be discussed.

#### 4.1.2 Experimental Implementation of the Multi-rate MI Classification

In this subsection, the proposed progressive and multi-rate MI classification framework and its practical implementation using the Emotiv EPOC headset is further elaborated. Fig. 4.1.1 illustrates the headset and electrode placements. EMOTIV company describes their headset as a high resolution, multichannel, portable system designed for practical research applications [57, 59]. It has 14 channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4) and two reference points (P3 and P4). Sampling is done sequentially at 2048 Hz internally, but the signal sent to the computer is already processed and reduced to 128 Hz with a resolution of 14 bits (1 LSB = 0.51V), a bandwidth of 0.2 - 45Hz, and filtered with digital notch filters at 50Hz and 60Hz. The built-in filters are digital 5th order sinc filters. The input dynamic range is 8400  $\mu$ V. The connection to the computer is wireless over the 2.4GHz band, and uses a proprietary protocol. The headset itself is wireless and powered with a LiPoly battery with 12 hours working time per full charge. Signal quality is determined internally with proprietary system and impedance measurements. Several recent research studies [60, 61] used the Emotiv headset reporting promising results which were the motivation to use it to implement and develop the proposed framework as described further on.

Several online and offline experimental scenarios were performed to develop/implement the proposed multi-rate classification framework. In the first scenario reported below, I briefly outline different experimental setups developed/implemented and observations made that led to the setup used in Scenarios 2 and 3 where the proposed framework were implemented/tested.

**Scenario 1:** In the experiments, originally, MI was considered during trials that last 3-5 seconds with 1-3 seconds break between trials and then the following two classes were used: (i) Forward arrow → stimuli provided to imagine lifting one's right hand, and; (ii) Backward arrow ← stimuli provided to imagine lifting one's right foot continuously during each trial. At first, total of 200 trials were performed by each subject and data was collected with the headset tilted back to approximately cover C3 and C4 motor cortex area as suggested by [62]. From the feedback obtained from the subjects, it is concluded that the experiment was too long and headset placement was also uncomfortable. Therefore, we reduced the total trials and performed two sessions of 30 trials with an intermission to make the experiment easier on the subjects. The headset was also worn normally and the experiments were done in the middle of an empty conference room, far from other electronics (other than the laptop). Besides, a feedback (shown in Fig. 4.1.2) was tried out by the subject, using an Arduino microcontroller which is used to turn BCI classification results into external lighting of different-colored LEDs. This feedback is removed later on to keep the setup simpler. Finally, the classes were changed to “left and right hand” movements. The primary attempt, originally, was to avoid this to prevent dyslexic confusion of the stimuli, yet, the undertaken effort was to simplify the cognitive process such that the cues are more simply linked with the action. As a result, the stimuli were chosen as black and white line drawings of a left or right hand offset on the left or right side of a central cross, Fig. 4.1.2.

**Scenario 2:** Before applying the proposed framework to real datasets collected from Emotiv headset, in this scenario, the proposed algorithm was used to classify BCI Competition III IVa datasets. The experiment is performed based on the following steps: (i) Raw data is filtered to remove the DC gain and to pick the information within 7-30Hz. Thereafter, the filtered data is chopped into epochs and then smoothed using the weighted moving average method by a window size of 10 time samples. After this step, the dataset is downsampled to keep one sample out of each batch of 10;

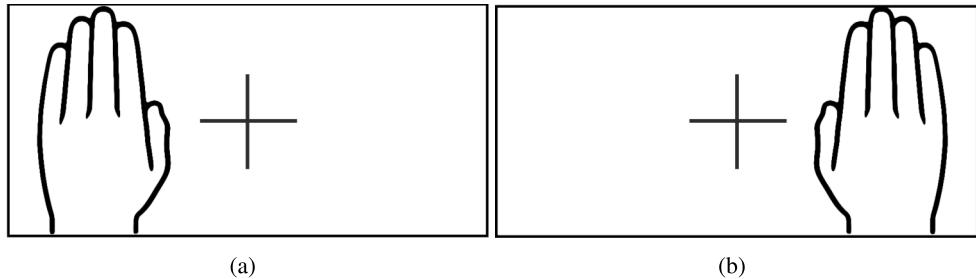


Figure 4.3: (a) The visual stimulus shown to the subject as ‘Left Hand’. (b) The visual stimulus shown to the subject as ‘Right Hand’.

Table 4.1: Performance of different models based on real experimental data sets.

|             | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|-------------|---------|---------|---------|---------|---------|
| CSP(2d) LDA | 52.00%  | 62.50%  | 66.67%  | 50.00%  | 20.00%  |
| CSP(2d) QDA | 54.00%  | 60.00%  | 70.00%  | 55.00%  | 30.00%  |
| CSP(4d) LDA | 48.00%  | 47.50%  | 50.00%  | 60.00%  | 80.00%  |

(ii) The CSP method is used to extract four features,  $M$  first trials were selected to train the classifier; (iii) A Quadratic Discriminant (QD) classifier by 10-fold cross validation is trained; (iv) The trained classifier is used by the Active Filter to classify incoming epochs, as the Progressive Filter was collecting the new epochs to update and re-train itself; (v) After  $L$  epochs, the Progressive Filter, re-trained by  $M + L$  trials and fused with Active Filter to be used as the new Active Filter for classifying the next  $L_1$ -epochs. Steps (ii) to (v) are repeated for each  $L$ -epoch batch under the condition of improving the classification accuracy. As an observation, the dataset of subject “aa” in BCI Competition II-Iva, was evaluated by the aforementioned algorithm, by setting  $M = 90$  and  $L = 20$ . The QD classifier is trained by the accuracy of 93.3% and classifies the next 20 epochs, as the Active Filter, by the accuracy of 55%. Thereafter, Progressive Classifier is updated using 110 epochs, trained classifier’s accuracy is 98.2%, and is fused with the Active classifier, which classifies the next 30 epochs by the accuracy of 60%. The same procedure is done to train the Progressive Filter by the accuracy of 93.6% using 130 epochs. The updated Active Filter then classifies next 40 new epochs by the accuracy of 65%. Scatter plots of these two Progressive Filters are shown in Fig. 4.1.2.

**Scenario 3:** In this scenario, the same proposed framework is implemented based on the experimental setup described in Scenario 1. The first ten epochs are incorporated to train Model 1 using CSP

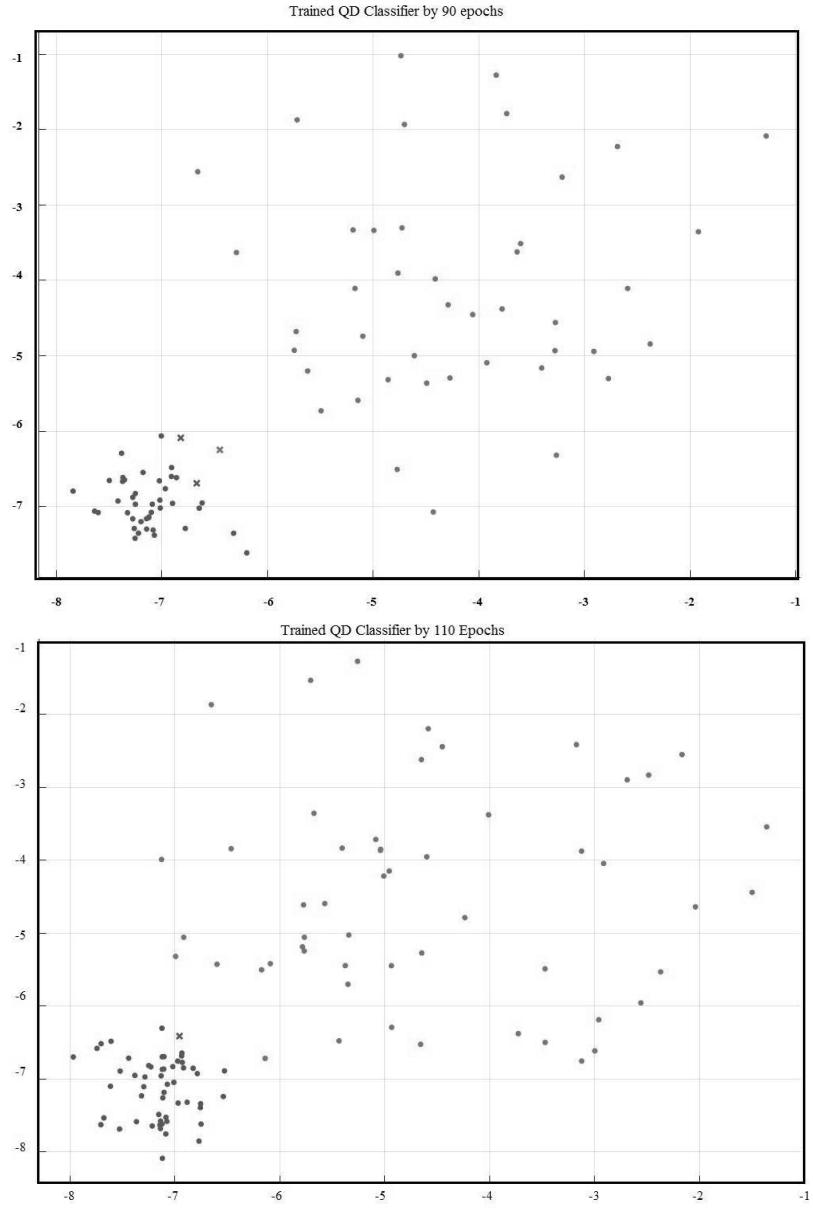


Figure 4.4: Scatter plots obtained from two Progressive Filters trained based on 90 and 110 epochs, respectively.

feature extraction, and LDA model for classification. Only two dimensions of the CSP are used. The resulting trained model is employed by the Active Filter for online analysis of the following epochs, i.e., categorizing them as they come in. While the Active Filter performs online classification based on Model 1, the Progressive Filter uses the next ten epochs, i.e., it trains another model (referred to as Model 2) based on twenty epochs and after the 20th epoch starts classifying in parallel to the

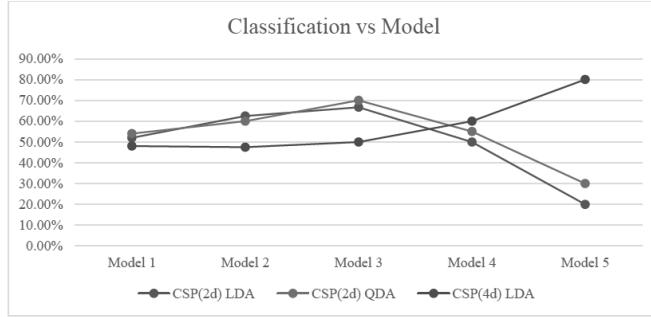


Figure 4.5: Classification results obtained from implementation of the proposed progressive and multi-rate framework based on data collected via Emotiv headset.

Active Filter which uses Model 1. This process was repeated until Model 5 is trained. Fig. 4.1.2 and Table 4.1 illustrate the results. It is observed that using more training data could result in providing improved performance. In practice, the online model (the Active Filter) should be updated to the best performing model available at the Progressive Filter. The same procedure was repeated by using a QDA classifiers, and then again for both LDA and QDA using 4 dimensions from the CSP. The results show a peak return from training in LDA and QD for 2d CSP, before more training data actually decreases the model's performance. Although 4d CSP with LDA starts out with lower performance as compared to its counterparts, it produces improving results as more epochs are taken into account. These results are significantly promising with further improvement/investigation which is one of the objectives of our future research in this direction.

#### 4.1.3 Conclusion

In this section, a novel progressive and multi-rate EEG-based MI classification framework is proposed, which consists of two separate filters. The Emotiv Epoc headset is used to develop and implement the proposed framework. The two filters (referred to as the Active Filter and the Progressive Filter) are partially coupled at the consensus epochs based on the individual classification results, i.e., the Active Filter (a simple MI feature extraction algorithm which uses pre-trained classification models) produces results at the end of each epoch while the Progressive Filter uses several epochs to re-train. The experimental implementations of the proposed framework indicate its potential for improving performance of real-time MI classification.

## **4.2 Improving the Accuracy of MI EEG-based BCIs Through Trimming the Epochs**

As previously stated, one of the most popular and commonly used techniques to satisfy the requirement for an effective and efficient BCI is Motor Imagery, which is defined as merely imagination of a limb movement, with no actual movement or peripheral (muscle) activation. However, as much as this field outlines a promising framework, the researchers dealing with MI commonly face two types of challenges. First challenge is to deal with different comprehensions of subjects from “imagination of the movement”. Some subjects imagine repeating the movement during each epoch, while some others might execute the mental imagination of the activity only once, and not necessarily within consistently equal time intervals after the stimulus is shown. In order to tackle this obstacle of various reactions, implementation of methods in which the classifier is trained subject by subject has been adopted by the researchers of the field. These approaches are adaptive to the nature of the datasets collected, for instance, should the subject react to the stimulus right after he/she sees it, the classifier of the dataset collected from this subject is trained to read the epochs’ information right after the marker. The second challenge is the fact that through a cognitive process, the brain of the subject learns to decrease the motor concentration while doing the same task. Hence, the amplitude of the signals within each epoch tends to descend over time. As a clarification to the case propounded, an actual example and the observation corresponding to that follows. Through experiments done for the research work described in the previous section, the subjects were asked to fill in surveys after they participated in each session of experiment. The outcome of these surveys was critical: it was difficult for them to keep their concentration all through the long experiments and they were not able to consistently do the tasks they were instructed for. The experiment was extremely tiring on the subjects, and, for instance, 100 trials would be too long to keep the subjects focus. The substantial difference of their performance was proved once the long experiments were splitted into sets of runs with much less number of trials: through feedbacks and computational analysis on the data collected from them, much better performance and more informative datasets were observed. Altogether, as the latency of the overall system (from headset, wireless communication and software) is unknown, and more importantly, delay in human response is inevitable, this

question was raised: where it is best start time sample within the time samples of each epoch? To answer this question, the following approach was designed and tested.

#### 4.2.1 Trimming Framework Outline and Simulation

To answer the raised question regarding the most informative time interval within each epoch, a straightforward processing was considered. As the main goal was to investigate whether or not a trimmed epoch would contribute to a better classification accuracy, the following modules were considered.

- **Preprocessing:** Raw dataset of BCI Competition III-IVa were filtered via an order 5 *Butterworth* filter, to remove the DC gain and to pick the information within 7-30Hz. Thereafter, the epochs were arranged and then smoothed using the weighted moving average method by a window size of 10 time samples. Then, the dataset is downsampled to keep one sample out of each batch of 10. Finally, within each epoch, there are 350 time samples.
- **Feature Extraction:** The CSP method is used to extract features. Two eigenvectors are taken for the purpose of construction of the whitening matrix. Also, In respect to the rule of thumb mentioned in Chapter 2, both scenarios of splitting the dataset into training and test trials is implemented: once with 60% of the trials for training, which is 168 trials out of 280, and the remaining 112 trials were tested by the trained classifier. In Another run, 196 trials (70% of the trials) were put aside for training the classifier and that leaves 84 trials in test dataset.
- **Classification:** Both LDA and QDA models with 5-fold cross validation were utilized to classify the test sets.

These three steps, without any dimensionality reduction technique applied or complexity in the algorithm were deliberately adopted for the sole purpose of enabling a better exhibition of the effects of the trimming step, which comes before feature extraction.

The trimming step is elaborated in Algorithm

The accuracies achieved from implementation of this approach are presented in Tables 4.2.1. As the tables and the Fig. 4.2.1 illustrate, trimming step impressively boosts the performance of the conventional CSP-based algorithm.

---

**Algorithm 3** TRIMMING THE EPOCHS

---

**Input:**  $\{\text{Original EEG signals } \mathbf{X}\}$

**Output:**  $\{\text{EEG signals } \bar{\mathbf{X}} \text{ which epochs are trimmed based on the best time of start } \hat{t}\}$

- 1: **Sampling the trainset:** Half of the  $\mathbf{X}_{train}$  is considered as a sample of the training set  $\bar{\mathbf{X}}_{train}$ , to investigate the best time of start.
  - 2: **Best Time of Start Loop:**
    - Step = number of time samples corresponding 0.1s
    - For:**  $\hat{t} = 1:(\text{Number of time samples between time } = 0 \text{ to } 1\text{s})/\text{Step}$ 
      - Epochs of  $\mathbf{X}_{train}$  are adjusted as  $\bar{\mathbf{X}}_{train} = \mathbf{X}_{train}(:, \hat{t}:\text{end}, :)$ ;
      - CSP is applied to  $\bar{\mathbf{X}}_{train}$  and then LDA is trained and the training accuracy is stored.
    - End For.**
  - 3: **Fiding the best  $\hat{t}$ :** The stored accuracies are evaluated, the  $\hat{t}$  corresponding to the best  $\hat{t}$  determines  $\hat{t}_{best}$ .
  - 4: **Preparing the final  $\bar{\mathbf{X}}$ :**  
Final  $\bar{\mathbf{X}} = \bar{\mathbf{X}}(:, \hat{t}_{best}:\text{end}, :)$ ;
- 

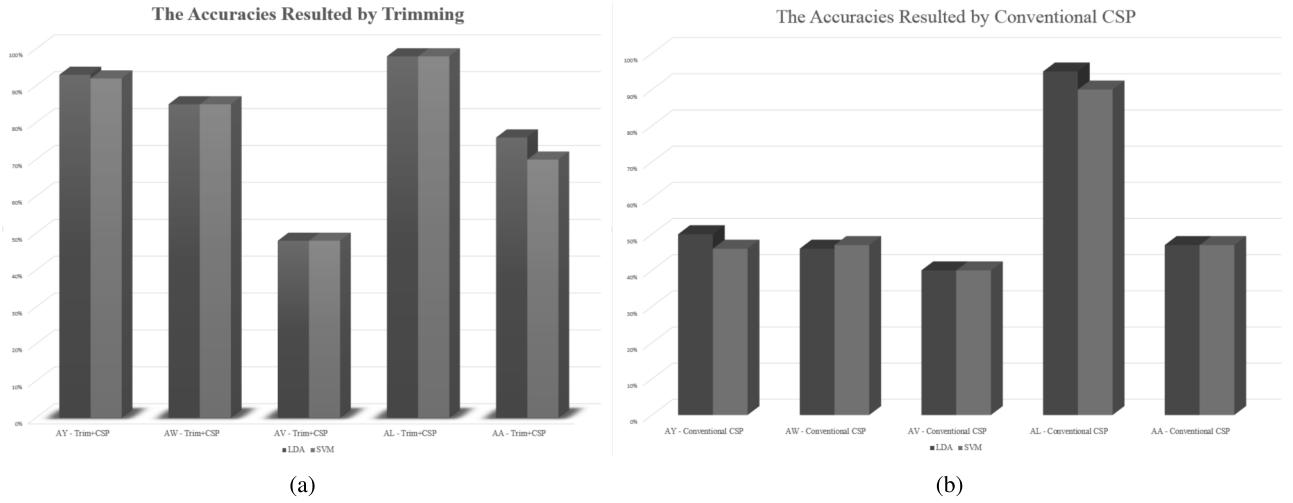


Figure 4.6: (a) The accuracies achieved via adding the trimming step to the conventional CSP algorithm. (b) The accuracies achieved via the conventional CSP algorithm.

#### 4.2.2 Conclusion

The proposed algorithm aims to help BCI developers who focus on MI EEG-based systems to better evaluate and estimate the intentions of the subject. The accuracy of classification in MI EEG-based BCIs can be significantly enhanced through trimming the epochs. This method is specifically useful for those BCIs used for Locked-in patients due the way it enriches the processing algorithm and takes the burden of performance concentration off the patients' shoulders. Implementation of this algorithm is highly recommended for any EEG-based study in order to prevent the possible

exhaustion of the subjects of the study. Our work significantly shows better results as compared to its counterpart.

### 4.3 Summary

Throughout this chapter, the practical experiments I carried out for my thesis and the solutions provided and tested were elucidated in details. The first solution was “Progressive Fusion of Multi-rate MI Classification for BCIs” which aimed to address the issues arised in the case of limited number of training trials available at the intiation phase of a BCI. The second solution was “Improving the Accuracy of MI EEG-based BCIs Through Trimming the Epochs” during which an additional step was suggested in order to trim the recorded epochs in a manner that most informative parts of the signals are extracted and the segments of the epochs which do not include the respose of the subjects to the stimuli would be discarded. Both frameworks show notable and impressive impact on the performance of MI EEG-based BCI systems. This completes the discussion on BCI systems and frameworks to enhance the end results of such systems.

Table 4.2: Accuracy comparison of the proposed trimming framework.

(a)

| Subjects | 168 Train<br>(60% - 40%) | LDA   |      |      | SVM  |       |      | Test |  |
|----------|--------------------------|-------|------|------|------|-------|------|------|--|
|          |                          | Train |      |      | Test | Train |      |      |  |
|          |                          | Mean  | STDV | Best |      | Mean  | STDV |      |  |
| AY       | Trim                     | 91%   | 0.5  | 92%  | 93%  | 94%   | 0.6  | 95%  |  |
|          | CSP                      | 65%   | 0.7  | 67%  | 50%  | 66%   | 1.7  | 70%  |  |
| AW       | Trim                     | 93%   | 0.2  | 94%  | 85%  | 93%   | 0.4  | 94%  |  |
|          | CSP                      | 59%   | 0.8  | 62%  | 46%  | 60%   | 0.9  | 62%  |  |
| AV       | Trim                     | 60%   | 0.9  | 62%  | 48%  | 58%   | 1.0  | 61%  |  |
|          | CSP                      | 57%   | 0.7  | 59%  | 40%  | 56%   | 1.8  | 60%  |  |
| AL       | Trim                     | 96%   | 0.1  | 97%  | 98%  | 96%   | 0.2  | 97%  |  |
|          | CSP                      | 93%   | 0.4  | 94%  | 95%  | 95%   | 0.4  | 96%  |  |
| AA       | Trim                     | 88%   | 0.5  | 90%  | 76%  | 90%   | 0.5  | 90%  |  |
|          | CSP                      | 0.68  | 0.6  | 70%  | 47%  | 0.68  | 1.0  | 71%  |  |

|         |                  |     |         |                  |     |
|---------|------------------|-----|---------|------------------|-----|
| Average | Trim + CSP       | 80% | Average | Trim + CSP       | 66% |
|         | Conventional CSP | 56% |         | Conventional CSP | 54% |

(b)

| Subjects | 196 Train<br>(70% - 30%) | LDA   |      |      | SVM  |       |      | Test |  |
|----------|--------------------------|-------|------|------|------|-------|------|------|--|
|          |                          | Train |      |      | Test | Train |      |      |  |
|          |                          | Mean  | STDV | Best |      | Mean  | STDV |      |  |
| AY       | Trim                     | 90%   | 0.5  | 92%  | 92%  | 95%   | 0.4  | 95%  |  |
|          | CSP                      | 65%   | 0.6  | 66%  | 55%  | 66%   | 1.5  | 69%  |  |
| AW       | Trim                     | 94%   | 0.2  | 95%  | 93%  | 94%   | 0.6  | 96%  |  |
|          | CSP                      | 94%   | 0.3  | 95%  | 90%  | 93%   | 0.4  | 94%  |  |
| AV       | Trim                     | 61%   | 0.6  | 63%  | 40%  | 63%   | 1.3  | 66%  |  |
|          | CSP                      | 58%   | 0.8  | 61%  | 46%  | 58%   | 1.0  | 60%  |  |
| AL       | Trim                     | 97%   | ~0   | 97%  | 100% | 97%   | 0.3  | 97%  |  |
|          | CSP                      | 93%   | 0.3  | 94%  | 98%  | 95%   | 0.3  | 95%  |  |
| AA       | Trim                     | 88%   | 0.3  | 89%  | 68%  | 89%   | 0.4  | 90%  |  |
|          | CSP                      | 0.71  | 0.4  | 72%  | 52%  | 0.7   | 1.0  | 72%  |  |

|         |                  |     |         |                  |     |
|---------|------------------|-----|---------|------------------|-----|
| Average | Trim + CSP       | 78% | Average | Trim + CSP       | 78% |
|         | Conventional CSP | 68% |         | Conventional CSP | 68% |