Feature extraction from EEG signals and regularization for brain-computer interface

Mishuhina, Vasilisa

2019

Mishuhina, V. (2019). Feature extraction from EEG signals and regularization for brain-computer interface. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/141579>

<https://doi.org/10.32657/10356/141579>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 10:34:56 SGT

FEATURE EXTRACTION FROM EEG SIGNALS AND REGULARIZATION FOR BRAIN-COMPUTER INTERFACE

VASILISA MISHUHINA

School of Electrical and Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

30/12/2019

Date

Vasilisa Mishuhina

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

December 30, 2019

Date

Jiang Xudong

Authorship Attribution

Statement

This thesis contains material from paper published in the following peer-reviewed journal in which I am listed as an author:

Chapter 4 is published as V. Mishuhina and X. Jiang, "Feature Weighting and Regularization of Common Spatial Patterns in EEG-Based Motor Imagery BCI," in IEEE Signal Processing Letters, vol. 25, no. 6, pp. 783-787, June 2018. doi: 10.1109/LSP.2018.2823683

The contributions of the co-authors are as follows:

- I conducted research and wrote the draft of the manuscript.
- Prof Jiang suggested the materials area and revised the manuscript.

30/12/2019

Date

Vasilisa Mishuhina

Acknowledgements

First and foremost, I am sincerely thankful to my supervisor, Associate Professor Jiang Xudong, for his continuous guidance, motivation and encouragement that helped me throughout my PhD work.

I would like to express my gratitude to all my friends and family for their kind support. My special thanks goes to my university friends Manish Prajapati, Anna Paterova, Shilpa Manandhar, Alana Soehartono, Julia Gusakova, Alexander Matyasko, Parastoo Fahimi, and many more who made this time enjoyable as well as helped me with valuable suggestions. I am also grateful to Infinitus lab staff Ng Teng Kwee and Foo Wenney for their technical support during my research experiments.

Finally, I wish to thank my Mama, Papa and Lama for their love, support and encouragement. I dedicate my thesis to them.

Table of Contents

Statement of Originality	i
Supervisor Declaration Statement	iii
Authorship Attribution Statement	v
Acknowledgements	vii
Abstract	xiii
List of Figures	xvii
List of Tables	xx
List of Abbreviations	xxi
1 Introduction	1
1.1 Brain-Computer Interfaces	1
1.2 Motivation	2
1.3 Objectives	3
1.4 Major contribution of the Thesis	4
1.5 Organization of the Thesis	4
2 Background of Brain-Computer Interfaces	7
2.1 Introduction	7

2.2	BCI components	9
2.3	Signal acquisition	10
2.3.1	Measurement techniques	12
2.3.2	Electroencephalography	15
2.4	Neurophysiological signals	20
2.4.1	Brain	20
2.4.2	Evoked potentials	22
2.4.3	Spontaneous signals	23
2.5	Motor imagery	25
2.5.1	Homunculus	25
2.5.2	Architecture of motor imagery brain-computer interfaces . .	26
3	Signal Processing and Machine Learning for MI-BCI	29
3.1	Preprocessing	29
3.1.1	Frequency filtering	29
3.1.2	Time framing	30
3.1.3	Channel selection	31
3.2	CSP based feature extraction	32
3.2.1	Common spatial pattern	35
3.2.2	Common spatio-spectral pattern	37
3.2.3	Regularized common spatial pattern	37
3.2.4	Filter bank common spatial pattern	39
3.2.5	Sparse filter band common spatial pattern	41
3.2.6	Sparse Bayesian learning of frequency bands	42
3.2.7	Improved discriminative filter bank common spatial pattern	43
3.2.8	Common spatial pattern with tangent space mapping	45
3.2.9	Temporally constrained sparse group spatial pattern	46
3.2.10	Local region frequency-optimized common spatial pattern . .	47
3.3	Classification	49

3.3.1	Linear discriminant analysis	49
3.3.2	Naive Bayesian Parzen window	49
3.3.3	Support vector machine	50
3.3.4	Minimum Mahalanobis distance	52
4	Feature Weighting and Regularization	55
4.1	Proposed method	56
4.1.1	Feature extraction by common spatial patterns	56
4.1.2	Feature weighting	58
4.1.3	Feature regularization and classification	60
4.2	Experiments	65
4.2.1	EEG datasets	65
4.2.2	Preprocessing and experiment settings	68
4.2.3	Results and discussion	70
4.3	Summary	83
5	Time-Frequency Common Spatial Patterns	85
5.1	Proposed method	86
5.1.1	Time decomposition of EEG signal	86
5.1.2	Frequency decomposition of EEG signal	87
5.1.3	Complex common spatial patterns on decomposed frequency components	89
5.1.4	Removing indiscriminative and unreliable features	90
5.1.5	Classification in low dimensionality subspace	91
5.1.6	Classification with feature weighting and regularization	94
5.2	Experiments	98
5.2.1	EEG datasets	98
5.2.2	Preprocessing, parameter settings and visualization of pro- posed TFCSP	101
5.3	Results and discussion	109

5.3.1	Experimental results of TFCSP with PCA	110
5.3.2	Experimental results of TFCSP with FWR	113
5.4	Summary	115
6	Conclusion	117
6.1	Future work	118
	Author's Publications	121
	Bibliography	123

Abstract

Brain-computer interface (BCI) is a promising technique which analyses and translates brain signals to computer commands. BCI can be used to communicate directly from a brain to a computer instead of communication through the central nervous system (CNS) and muscles. Consequently, it has received tremendous research interest in recent years. Various methods of signal processing and machine learning supported by neuroscience, have made it possible to extract useful information from the brain. This has enabled BCI to be successfully implemented in various communication and rehabilitation applications, such as control of computer cursor, robot, prosthesis, wheelchair etc. Similarly, there are several emerging applications in other fields of gaming, security, and education.

At present, there are several techniques available to acquire brain signals. In our work, we use electroencephalography (EEG) that measures electric brain activity by electrodes located on the scalp. We choose EEG because it is a non-invasive and low cost technology with good time resolution. However, it has several difficulties with spatial resolution due to overlapping signals and distortion from the scalp. The goal of this research is to improve feature extraction and regularization of EEG signals using machine learning methods and hence achieve better results during the classification of the signals for motor imagery BCI (MI-BCI). MI-BCI establishes communication through signal generated from the human brain when a person is imagining the movement of some part of the body. The main motivation behind MI-BCI is to improve life of people with different neuromuscular disorders

such as amyotrophic lateral sclerosis, cerebral palsy, stroke, or spinal cord injury, as well as to help them with rehabilitation.

This thesis presents a detailed explanation of state-of-the-art methods of MI-BCI such as common spatial pattern (CSP) and its extensions. CSP is a widely used algorithm in BCI. However, it suffers from overfitting and sensitivity to noise and outliers. To enhance the performance of CSP, different variations of this method have been proposed in literature. However, each approach has its own limitations and does not perform well for different subjects and datasets. Therefore, in this thesis, we extend the existing approaches and propose new ones to improve classification.

Firstly, a novel feature weighting and regularization (FWR) method that utilizes all CSP features to avoid information loss is proposed. The method can be applied in all CSP-based approaches. Experiments demonstrate that the proposed FWR method enhances the classification accuracy as compared to the conventional feature selection approaches. Secondly, this thesis proposes a novel approach time-frequency CSP (TFCSP) which decomposes the EEG signal into time stages and frequency components to find the robust and discriminative features for classification. The proposed algorithm yields higher classification accuracy than other state-of-the-art methods on multiple BCI competition datasets. These improvements will eventually lead BCI to real-life applications.

List of Figures

2.1	General architecture of BCI	9
2.2	Invasive, partially-invasive, and non-invasive approaches	11
2.3	Illustration of non-invasive signal acquisition methods	13
2.4	10-20 standard system	16
2.5	Overview of most prominent EEG rhythms	19
2.6	Brain structure	21
2.7	Event-related synchronization and event-related desynchronization .	25
2.8	Homunculus	26
2.9	Block diagram of a conventional MI-BCI system.	27
3.1	Time framing of the motor imagery trial.	31
3.2	Channel selection	32
3.3	The architecture of filter bank CSP	40
3.4	The architecture of sparse filter bank CSP	41
3.5	The architecture of sparse Bayesian learning of frequency bands . .	42
3.6	The architecture of discriminative filter bank CSP	43
3.7	The architecture of improved discriminative filter bank CSP	44
3.8	The architecture of temporally constrained sparse group spatial pat- tern	46
3.9	The architecture of local region common spatial pattern	47

3.10 The architecture of local region frequency optimized common spatial pattern	47
3.11 Illustration of support vector machine classifier	51
4.1 Soft and step functions.	60
4.2 Eigen-values of a sample BCI training task.	63
4.3 Position of EEG electrodes in Dataset IIa	66
4.4 Timing scheme of the paradigm in Dataset IIa	67
4.5 Position of EEG electrodes in Dataset IIIa	67
4.6 Timing scheme of the paradigm in Dataset IIIa	68
4.7 Comparison of classification rate for task 1 from BCI competition III, Dataset IIa	71
4.8 Comparison of classification rate for task 2 from BCI competition III, Dataset IIa	72
4.9 Comparison of classification rate for task 3 from BCI competition III, Dataset IIa	73
4.10 Comparison of classification rate for task 4 from BCI competition III, Dataset IIa	74
4.11 Comparison of classification rate for task 5 from BCI competition III, Dataset IIa	75
4.12 Comparison of classification rate for task 6 from BCI competition III, Dataset IIa	76
4.13 Comparison of classification rate for task 1 from BCI competition III, Dataset IIIa	77
4.14 Comparison of classification rate for task 2 from BCI competition III, Dataset IIIa	77
4.15 Comparison of classification rate for task 3 from BCI competition III, Dataset IIIa	78

4.16 Comparison of classification rate for task 4 from BCI competition III, Dataset IIIa	78
4.17 Comparison of classification rate for task 5 from BCI competition III, Dataset IIIa	79
4.18 Comparison of classification rate for task 6 from BCI competition III, Dataset IIIa	79
5.1 Time stages of the proposed TFCSP method.	86
5.2 The framework of the proposed TFCSP method using PCA.	94
5.3 The overall framework of the proposed TFCSP method.	98
5.4 Position of EEG electrodes (blue) in Dataset IIb	99
5.5 Timing scheme of the paradigm in Dataset IIb	100
5.6 Visualization in topography maps.	102
5.7 Visualization in topography maps.	104
5.8 Visualization in topography maps.	105
5.9 Visualization in topography maps.	106
5.10 Visualization in topography maps.	107
5.11 Visualization in topography maps.	108

List of Tables

2.1	Comparison of non-invasive signal acquisition methods	15
4.1	Comparison of the best average performance for BCI competition IV, Data set IIa	80
4.2	Comparison of the average areas under ROC-curve (AUC) for BCI competition IV, Data set IIa	80
4.3	Comparison of the best average performance for BCI competition III, Dataset IIIa	81
4.4	Comparison of the average areas under ROC-curve (AUC) for BCI competition III, Data set IIIa	81
4.5	Comparison of the best average performance for BCI competition IV, Data set IIa	81
4.6	Comparison of the average areas under ROC-curve (AUC) for BCI competition IV, Data set IIa	82
4.7	Comparison of the best average performance for BCI competition III, Dataset IIIa	82
4.8	Comparison of the average areas under ROC-curve (AUC) for BCI competition III, Data set IIIa	82
5.1	Decomposition parameters on different datasets	101
5.2	Comparison of 5x5 crossvalidation performance for left vs right hand task from BCI competition IV, Data set IIa	110

5.3	Comparison of 5x5 cross-validation performance for foot vs tongue task from BCI competition IV, Data set IIa	111
5.4	Comparison of 5x5 cross-validation performance for left vs right hand task from BCI competition IV, Data set IIb	111
5.5	Comparison of 5x5 cross-validation performance from BCI competition III, Data set IIIa	111
5.6	Comparison of 10x10 cross-validation performance for left vs right hand task from BCI competition IV, Data set IIb	112
5.7	Comparison of 5x5 crossvalidation performance for left vs right hand task from BCI competition IV, Data set IIa	113
5.8	Comparison of 5x5 cross-validation performance for foot vs tongue task from BCI competition IV, Data set IIa	113
5.9	Comparison of 5x5 cross-validation performance for left vs right hand task from BCI competition IV, Data set IIb	113
5.10	Comparison of 5x5 cross-validation performance from BCI competition III, Data set IIIa	114
5.11	Comparison of 10x10 cross-validation performance for left vs right hand task from BCI competition IV, Data set IIb	114

List of Abbreviations

BCI	Brain-Computer Interface
BOLD	Blood Oxygenation Level Dependent
BRI	Brain-Robot Interface
CSP	Common Spatial Pattern
CSP-TSM	CSP with Riemannian Tangent Space Mapping
CSSP	Common Spatio-Spectral Pattern
DFBCSP	Discriminative Filter Bank CSP
ECoG	Electrocorticography
EEG	Electroencephalography
ERD	Event-Related Desynchronization
ERE	Eigenfeature Regularization and Extraction
ERS	Event-Related Synchronization
FBCSP	Filter Bank CSP
FBRCS	Filter Bank Regularized CSP
fMRI	Functional Magnetic Resonance Imaging
FS	Feature Selection

FWR	Feature Weighting and Regularization
IDFBCSP	Improved Discriminative Filter Bank CSP
ISSPL	Iterative Spatio-Spectral pattern Learning
Lasso	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant Analysis
LRCSP	Local Region CSP
LRFCSP	Local Region Frequency-Optimized CSP
MDRM	Minimum Distance to Riemannian Mean
MEG	Magnetoencephalography
MI-BCI	Motor Imagery Brain-Computer Interface
MIBIF	Mutual Information Best Individual Features
MMISS	Mutual Information between Spatial Spectral Features
NBPW	Naive Bayesian Parzen Window
NIRS	Near-Infrared Spectroscopy
OSSFN	Optimal Spatio-Spectral Filter Network
PCA	Principal Component Analysis
PTFBCSP	Penalized Time-Frequency Band CSP
RCSP	Regularized CSP
SBLFB	Sparse Bayesian Learning of Frequency Bands
SCP	Slow Cortical Potentials
SCSSP	Separable Common Spatio-Spectral Pattern
SFBCSP	Sparse Filter Band CSP

SMR Sensory Motor Rhythms

spec-CSP Spectrally Weighted CSP

SSEP Steady-State Evoked Potentials

SSVEP Steady-State Visual Evoked Potentials

SWCSP Spectrally Weighted CSP

SWDCSP Sliding Window Discriminative CSP

TFCSP Time-Frequency Common Spatial Pattern

TRCSP Tikhonov Regularization CSP

TSGSP Temporally Constrained Sparse Group Spatial Pattern

Chapter 1

Introduction

1.1 Brain-Computer Interfaces

Brain-Computer Interface (BCI) analyses and translates brain signals to computer commands [1, 2]. Instead of normal communication through the central nervous system (CNS) and muscles, BCI allows communication directly between a brain and a computer. The possibilities of implementing this technology are numerous, and most of them can be used for meaningful and beneficial applications.

With efficient signal acquisition technique and robust methods of signal processing and machine learning, BCI can be used to enhance life of people with health disabilities. Driving wheelchair or walking in exoskeleton by power of thoughts can return functional life to paralyzed or partially paralyzed people, as well as support caregivers. Another technology Brain-Robot Interface (BRI) allows people to operate robots by thinking. This technology also has promising applications, especially in hazard zones where humans cannot operate directly.

Another field where BCI has shown immense potential is neuro-rehabilitation. For instance it is used in motor imagery rehabilitation where people activate motor cortex by imagining movement of their body which in-turn enables real body

movements. As a result it can provide alternate means of living to disabled or paralyzed people. BCI can therefore support and motivate people in living a more capable life.

However, there are several challenges in BCI which need to be tackled before these applications and future possibilities can be properly realized.

1.2 Motivation

Brain signals are the results of millions of neurons that produce different types of changes in electrical or magnetic activity. Sensors can be used to detect these changes over different areas of the brain. Some popular non-invasive signal acquisition methods or sensors used in BCI are: magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), near-infrared spectroscopy (NIRS), and electroencephalogram (EEG). EEG is one of the most commonly used techniques due to its high temporal resolution, portable equipment, low cost and easy of use. Moreover, EEG does not require surgical procedure or special technicians to operate the system.

The focus of this thesis is on using EEG for Motor Imagery BCI (MI-BCI). It establishes communication through signals generated from human brain when a person is imagining the movement of some part of the body. A fundamental issue with non-invasive MI-BCI based on EEG is extracting right features from noisy and non-stationary EEG signal with low spatial resolution. Moreover electrodes are subjected to signals that are overlapping with each other. This limitation is intrinsic to the EEG signal and has to be dealt with methodologically.

The motivation of this research is to improve feature extraction and regularization of EEG signal for MI-BCI using machine learning methods to overcome non-stationarity and low spatial resolution of EEG signal and hence achieve better results during classification of motor imagery tasks. This will lead BCI a step

closer to real-life applications.

1.3 Objectives

The research objectives of the thesis can be summarized as follows:

1. Enhance classification rate for motor imagery BCI.

Despite several different methods and enhancements, there is still no single method that can work consistently the best for all subjects, datasets, and setups. Therefore, the main goal of this research is to enhance classification and develop a better method suitable for different cases.

2. Develop methods for regularization and feature extraction.

Since a fundamental issue in MI-BCI is extracting right features from noisy and non-stationary EEG signal with low spatial resolution, the next objective is to develop efficient methods for regularization and feature extraction.

3. Avoid loss of discriminative information from EEG signal

To alleviate the problems of overfitting or poor generalization and hence boost the classification accuracy, different approaches in motor imagery BCI select some features and remove the others. However, this will result in the loss of discriminative information. So, the third objective of the thesis is to extract more discriminative information by replacing the feature selection with feature weighting and regularization.

4. Using temporal information for rich informative features, and applying new time paradigm for motor imagery.

Different motor imagery may show different temporal pattern. To capture the coarse temporal pattern of different motor imagery, we propose to decompose the full length of signal trial into reaction, action, and completion

stages. It helps to discriminate different motor imagery tasks better and leads to sharper spatial filters, as well as generates rich informative features.

1.4 Major contribution of the Thesis

The major contribution of the thesis can be summarized as follows:

- Feature weighting and regularization (FWR) method:
 - Utilizing all features to avoid information loss.
 - Can be applied to any common spatial patterns (CSP) methods.
 - Regularizing features from non-stationary EEG signal.
 - Enhances classification rate for CSP methods.
- A novel time-frequency common spatial patterns (TFCSP) method:
 - Using time windows and adopting subject reaction time paradigm.
 - Applying FFT to extract frequency components.
 - Performs complex CSP.
 - Utilizes all CSP features to avoid information loss.
 - Yields a higher classification in comparison to the performance of the existing novel method.

1.5 Organization of the Thesis

Chapter 2 introduces the topic of brain-computer interfaces. It discusses on neurophysiological explanation of brain processes, which enables us to build BCI, and on the architecture of the BCI system. Furthermore, the types of neurophysiological signals and features to extract, techniques to measure brain signals, and motor imagery are discussed in this chapter.

Chapter 3 provides a detailed literature review of methods used in motor imagery BCI. This is followed by explanation on signal processing methods for preprocessing and machine learning methods for feature extraction, feature selection, and classification. All the methods used later for experiments in this thesis, as well as methods that were compared to the proposed method are explained in the chapter.

Chapter 4 proposes a novel feature weighting and regularization method that utilizes all features to avoid information loss. The experiments described in this chapter show the effect of the proposed method applied in the standard methods. Results on two publicly available datasets, BCI Competition III Dataset IIIa and IV Dataset IIa, demonstrate that the proposed FWR method enhances the classification accuracy in comparison to the conventional Feature Selection (FS) approaches.

Chapter 5 focuses on improving classification in MI-BCI. A novel approach time-frequency common spatial patterns is proposed which utilizes time windows and frequency components to find the best and most informative features. The chapter demonstrates experiments performed on three publicly available datasets BCI competition IV Dataset IIa, Dataset IIb and BCI competition III Dataset IIIa. The proposed algorithm TFCSP can yield a higher classification for the same BCI competition datasets based on comparison to the performance of the existing novel method.

Chapter 6 summarizes all the work in the thesis and proposes new directions for future work.

Chapter 2

Background of Brain-Computer Interfaces

2.1 Introduction

Since the first time electrical signals of the animal brain were recorded by Vladimir Pravdich-Neminsky in 1912 [3], and later of the human brain by Hans Berger in 1924 [4], an enormous amount of applications, devices, and methods have been developed for discovering brain. With the development of computers, researchers started to search how to connect brain to a computer in order to understand our brain more and if possible enhance it.

First BCI based on visual-evoked potentials was built by Jacques Vidal [1]. In his paper in 1973, he described a BCI as: “a first attempt to evaluate the feasibility and practicality of utilizing the brain signals in a man-computer dialogue while at the same time developing a novel tool for the study of the neurophysiological phenomena that govern the production and the control of observable neuroelectric events”.

Recently the definition and the goal of a BCI has been modified and specialized.

The most recent definition is given by Wolpaw: “A BCI is a system that measures CNS activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural CNS output and thereby changes the ongoing interactions between the CNS and its external or internal environment” [2].

People who have lost control over some parts of the body or limb can use a BCI to operate an exoskeleton or motorized wheelchair, therefore, BCI can *replace* natural output that has been lost as a result of injury or disease. After spinal cord injury when limbs are paralyzed, people might use a BCI to stimulate the paralyzed muscles so that the muscles move the limbs, therefore, a BCI output might *restore* lost natural output. If task requires continuous attention over a prolonged period, BCI can *enhance* the natural output by detecting lapses in attention and alerting the person. A BCI can *supplement* natural neuromuscular output with an additional, artificial output as controlling cursor. Finally, a BCI output is able to *improve* natural CNS output by using neurofeedback during rehabilitation.

Currently, there are plenty of different approaches and realizations of a BCI system and the number is continuously growing. It is a promising assistive technology which have been successfully used for various applications, e.g communication [5–8], rehabilitation approaches [9–13] and controlling devices, such as wheelchair [14–16], prosthesis [17–19], robot [20–22] and computer cursor control [23, 24]. Previous systems were mainly developed for medical and health applications. Now the applications have broadened to non-medical sector such as gaming [25–28].

In sections below, architecture of whole BCI systems is described based on neurophysiological explanation.

2.2 BCI components

BCI system translates raw brain signals (input) into various classes of these signals (output). The architecture of BCI is presented in Figure 2.1:

Figure 2.1: General architecture of BCI [29].

The main components of a BCI are:

- Signal acquisition: The first step is to detect physiological signals and transform them into voltages that can be amplified and digitized for processing by a computer. It is an input of the system.
- Preprocessing: The goal of the second step is to reduce noise and artifacts from the brain signals in order to clean data as much as possible before processing.
- Feature extraction: The goal of the third step is to extract the most discriminative characteristics of the data which differ among the classes.

- Classification unit. The features obtained from the training data are recognized as certain classes. During training phase this step is used for learning and calibrating the system.
- Translation into a command/application. The results from classifier are translated to computer commands or some output device.
- Feedback: A feedback helps the user to learn how to control the change in neural behavior and adapt accordingly.

Each of the components is important for overall performance of the BCI system. This thesis focuses on the feature extraction step and developing methods to enhance it. In sections below each step is described in more details.

2.3 Signal acquisition

The first step required to operate a BCI consists of measuring the subject's brain signals. The brain signals are the results of movement of millions of neurons that produce different types of changes in electrical or magnetic activity. Sensors detect these changes over different areas of the brain. Signal acquisition methods or sensors can be divided according to invasiveness, temporal resolution, spatial resolution, and electrical or hemodynamic origin as explained below.

The following three approaches according to **invasiveness** have been used in the BCI literature:

- Invasive: Intracranial microarrays are implanted directly into the grey matter of the brain.
- Partially-invasive: Sensors are implanted inside the skull but outside the brain. The partially-invasive BCIs employ Electrocorticography (ECoG), which requires the implantation of the electrodes between the cortex and the dura mater.

- Non-invasive: Sensors are located outside the skull, and there is no need for surgical implantation. Non-invasive BCIs can be developed using Electroencephalography (EEG), Magnetoencephalography (MEG), Functional Magnetic Resonance Imaging (fMRI) and Near-Infrared Spectroscopy (NIRS).

Figure 2.2: Invasive (intracortical recordings), partially-invasive (ECoG), and non-invasive (EEG) approaches [30].

Temporal resolution corresponds to the smallest time-lapse of neural activity that can be observed by the considered sensors. It is important to have higher temporal resolution in real-time application, since the signal and changes in the brain can be catch faster.

Spatial (topographic) resolution corresponds to the smallest neuronal area that can be accessed by the sensors on focus. Better spatial resolution provides a higher precision in understanding where the signal is coming from. It is essential for good classification.

Origin of activity refers to the measurement of a brain signal can be done based on its neurophysiological origins. It can be either electrophysiological (measure exchange information between the neurons in the form of ionic currents) or hemo-

dynamic (measure changes in the blood flow delivering glucose and oxygen) [31]. Electrophysiological sensors are used in EEG, ECoG, MEG, and Intracortical neuron recording, while hemodynamic in fMRI and NIRS.

Invasive recording techniques combine excellent signal quality, good spatial resolution, and a higher frequency range. However, invasive techniques require surgical procedures to implant the sensors. In the short term, the initial mechanical trauma of electrode insertion and the presence of a foreign body lead to scar formation surrounding the implant, inflammation, and a decrease in neuron density surrounding the electrode. Over time, implanted electrodes may suffer from mechanical failures (broken connectors, insulation fracture, and delamination) [32]. Moreover, such methods tend to be more complex and expensive than the non-invasive methods. It has been used mostly on people who suffer from neurophysiological disorders and are left with no other alternative means of treatment. It is interesting to note, that the first signal Hans Berger recorded was, in fact, from the invasive technique - intracranial EEG.

Non-invasive methods allow the development of simpler and more accessible BCIs. Significantly reduced risk for the user of non-invasive BCIs is the main advantage of such a technique. Therefore, in this thesis non-invasive method is considered. For better resolution, non-invasive methods need better acquisition systems, including software techniques that can overcome these drawbacks and distinguish patterns of brain activity from the noise.

2.3.1 Measurement techniques

Non-invasive techniques used in BCI applications including magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), near-infrared spectroscopy (NIRS), and electroencephalogram (EEG) are briefly described here:

- Magnetoencephalography measures the magnetic fields produced by neurons

Figure 2.3: Illustration of non-invasive signal acquisition methods [33].

that can be observed outside of the head. MEG uses an array of SQUID (Superconducting Quantum Interference Devices) as magnetometer to measure the magnetic field. From neurophysiological perspective, MEG signals are similar to EEG signals. In the context of BCI, MEG is an important tool to train human subjects to modulate their neural activity for movement control [34].

This recent non-invasive data acquisition technique has good spatial and temporal resolutions. MEG is more accurate and better at locating sources of the signal than EEG since the signal in MEG is less distorted by the skull. However, the equipment is extremely expensive and non-portable. Moreover, as there is a significant magnetic noise in a traditional setup, it requires a shielding room.

- Functional magnetic resonance imaging provides information about brain activity connected to the changes in the blood flow and oxygenation. After applying strong magnetic resonance pulses to the brain Blood Oxygenation Level Dependent (BOLD) changes are detected in fMRI . The use of fMRI

in BCI technology is relatively recent. Real-time fMRI-based BCI (rtfMRI-BCI) can be used in neurofeedback when a subject can voluntarily change brain activity [35]. Traditionally, fMRI recordings were analyzed offline for long duration.

The advantage of this popular non-invasive technique is good spatial resolution as it can discriminate two regions which are few millimeters apart based on the type of activity. However, BOLD changes can only be recorded after some seconds of neural activity. Therefore temporal resolution is comparatively low in fMRI. Besides, fMRI devices are very bulky, expensive, and require magnetic shielding.

- Functional near-infrared spectroscopy is a relatively new technology that uses the near-infrared region of the electromagnetic spectrum to analyze the hemodynamic response through the level of Infrared (IR) light absorption in the scalp optodes. The level of IR light absorption is different for oxygen-rich (oxygenated) blood and the oxygen-depleted (de-oxygenated) blood. fNIRS showed an advantage while identifying the prefrontal cortex cognitive tasks like mental arithmetic, music imagery, emotion induction [36].

fNIRS has relatively low cost and noise compared to fMRI. It is also portable, safe and easy to use. Since it is neuroimaging technology, fNIRS is not affected by electrical and magnetic noises. One of the major disadvantages of fNIRS is due to the nature of the hemodynamic response. As in fMRI, neural activity appears in vascular changes after a certain number of seconds and therefore the temporal resolution is limited.

- Electroencephalography measures electric brain activity caused by neuron firing [37, 38]. EEG signals are recorded by electrodes located on the scalp. The number of electrodes may vary from 1 to 256 and they are generally attached using an elastic cap. EEG is the most widely used brain signal

Table 2.1: Comparison of non-invasive signal acquisition methods

Type	Origin of activity	Temporal resolution	Spatial resolution	Portability	Price range (SGD)
Electroencephalography (EEG)	Electrophysiological	<0.05s	~10 mm	Portable	~250–70 000
Magnetoencephalography (MEG)	Electrophysiological	<0.02s	~5 mm	Not portable	~2.7–4 million
Functional magnetic resonance imaging (fMRI)	Hemodynamic	~1s	~1 mm	Not portable	>1.4 million
Near-infrared spectroscopy (NIRS)	Hemodynamic	~1s	~5 mm	Portable	>27 000

acquisition technique and is the principal noninvasive BCI method [2,39,40].

The advantage of EEG is high temporal resolution, as well as small size of equipment and low cost. The major drawback of EEG is the relatively poor spatial resolution. Since EEG sensors measure focal brain activation as well as distant activity, EEG signals become spatially noisy. As this limitation is intrinsic to the EEG signal, it has to be dealt with methodically.

Table 2.1 summarizes all signal acquisition techniques discussed above. Advantages such as high temporal resolution, portability, ease of use and low cost makes EEG a popular choice for BCI. Besides, EEG is non-invasive and does not require a special technician to operate the system. Furthermore, EEG does not need a shielding room compared to fMRI and MEG. We therefore consider the use of EEG signals for MI-BCI in this thesis.

2.3.2 Electroencephalography

Generally an EEG recording system contains electrodes, A/D converter, amplifiers, and a recording device. The electrodes record the analog signals from the scalp, the amplifiers amplify the signals and the A/D converter converts the signals to their digital form. The signals are then recorded and displayed if needed. Usually EEG signal is measured between an active electrodes and a reference electrode

as the potential difference over time. A ground electrode is used to define the differential voltage between the active and the reference points. Therefore in minimal configuration, an EEG recording contains one active, one reference, and one ground electrode.

In order to make results of the measurements comparable along a timeline and among each other, standard systems are being used to locate the sensors on the scalp at specific locations. One of the popular international methods is the ten-twenty system [41], which is based on proportional measurements between easily identified skull landmarks such that it provides adequate coverage of all parts of the head. Internationally standardized ten-twenty system electrode setup is showed at Figure 2.4.

Figure 2.4: 10-20 standard system [30].

EEG systems can be varied according to the type and number of electrodes. Dry [42], quasi-dry [43] and gel electrodes can be used in EEG systems. Among them gel electrodes are mostly used. Such electrodes need a local conductive paste application in order to reduce the electric impedance at the scalp-electrode contact points before each measurement. It is therefore time-consuming and requires

experienced staff. In this regard, dry electrodes were proposed. A dry electrode makes use of an inert, conductive material that mechanically couples with the skin for signal transduction, eliminating the use of gel application or skin preparation. However, such electrodes are bulkier, more expensive and more susceptible to movement artifacts due to the absence of a liquid contact. Finally, quasi-dry electrodes have also been proposed which achieves localized skin hydration effect at the electrode contact point by releasing a small amount of a moisturizing agent from a reservoir inside the electrode, triggered by the electrode/scalp adduction. It is a relatively new concept of electrodes and has not yet been properly tested and applied in devices.

The number of electrodes in an EEG device may range from 1 to about 256. In order to increase spatial resolution of EEG signal more electrodes can be used. For clinical settings, normally from 16 to 32 electrodes are used, for research-oriented applications from 64 to 256. In our research, we use 3 EEG databases with 3, 22 and 60 number of electrodes.

For real-life BCI applications, EEG device should be portable, preferably wireless and compact, with a smaller amount of electrodes, and preferably without the need of gel. However, the data that we utilize in this research are based on recordings from gelled electrodes with wires and experimental EEG setup. Such a setup provides better connection and more reliable signals. However even with such experimental setup, the classification results of the EEG signals are still far from perfect. Therefore, our research focuses on solving the problem of classification performance of EEG signals in such a system. The next step would be to use machine learning and signal processing to overcome problems of wireless, portable EEG devices with fewer electrodes for real-life applications.

After measuring, amplifying, digitizing, and storing, EEG signals can be expressed in terms of a number of rhythmic activities, each of which oscillates within a different frequency band. In order to study these rhythmic activities, the frequency

spectrum of EEG signals is usually divided into different frequency bands as shown in Figure 2.5. These frequency bands are termed delta (δ), theta (θ), alpha (α), beta (β), mu (μ), and gamma (γ).

Relevant characteristics of these bands are as follows:

- Delta (δ) rhythms can be observed in EEG signals of adults during deep sleep. It is the slowest recorded brain waves in human beings. The frequency of delta rhythms is below 4 Hz.
- Theta (θ) rhythms appear during low brain activities, drowsiness and sleep. The frequency of theta rhythms is between 4 Hz and 8 Hz.
- Alpha (α) [4] were among the first rhythmic waves documented and named by Hans Berger. It can be observed in EEG signals when the subject is relaxed or has closed eyes. Alpha rhythms are located in the 8-13 Hz frequency band and appear mainly in the occipital lobe of the brain.
- Mu (μ) rhythms are located as alpha rhythms in the 8-13 Hz frequency band, mostly in motor and sensorimotor cortex. It can be observed during movement and motor imagery. This rhythm is our main focus for motor imagery BCI. Mu rhythm is also known as the “sensorimotor rhythm” [44].
- Beta (β) rhythms are related to consciousness, brain activities, and motor behaviors when the eyes of a subject are open. The frequency of beta rhythms is between 12 Hz and 30 Hz. Motor imagery BCI uses beta rhythms as well.
- Gamma (γ) rhythms are associated with attention, perception, and cognition. It is the fastest recorded brain waves in human beings. The frequency of gamma rhythms is between 30 Hz and 100 Hz.

Figure 2.5: Overview of most prominent EEG rhythms [45].

In the next section more information on brain processes, neurophysiology and neurotransmission are presented.

2.4 Neurophysiological signals

2.4.1 Brain

The activity of the human brain can be recorded by EEG due to neurotransmission process. The core component of the brain is the neuron. The average human brain is composed of about 86 billion neurons [46]. A neuron is an electrically excitable cell that processes and transmits information by electrochemical signaling. A typical neuron consists of a soma (cell body), an axon (a special, extra-long, branched cellular filament, which can be thousands of times the length of the soma), and the dendrites (long, feathery filaments attached to the cell body in a complex branching “dendritic tree”). A cell body carries out the metabolic activities of the cell, and the dendrites receive signals from neighboring neurons. An axon sends out signals to other neurons through a gap between an axon and the receptor dendrites - the synapse.

Every neuron maintains a voltage gradient across its membrane, due to metabolically-driven differences in ions of sodium, potassium, chloride and calcium within the cell, each of which has a different charge. In the resting state of the cell, such concentrations lead to a voltage difference of approximately 70mV across the membrane - which is called “resting potential”. After crossing a certain threshold (typically -55mV) due to stimulus (either electrical or sensory) the sodium ion channels open and sodium ions flow into the cell thus triggering the rapid rise and fall in the membrane potential termed as “action potential”. This pulse travels rapidly along with the cell’s axon and is transferred across the synapse to a neighboring neuron, which receives it through its feathery dendrites.

The flow of positively or negatively charged ions into the postsynaptic or presynaptic cell results in excitatory or inhibitory postsynaptic potentials (EPSP or IPSP) and induces an extracellular current in the opposing direction [47]. This

current accumulates temporally and spatially and is often strong enough to admit measurement on the scalp. EEG signal is the potential differences caused by the EPSP and IPSP currents between two scalp locations. Vertically oriented cortical pyramidal neurons are the principal contributors to EEG.

Figure 2.6: Brain structure. Motor areas, somatosensory cortex, posterior parietal cortex, and visual cortex are particularly important for BCIs. [30].

The brain consists of five structures: hypothalamus, thalamus, brain-stem, cerebellum, and cerebral cortex. The cerebral cortex is the most important part of the brain for BCI. It is divided into four lobes frontal, temporal, parietal and occipital as in Fig 2.6.

- The frontal lobe is responsible for memory, language, motor functions, behavior. Prefrontal cortex is activated during conscious thinking. Premotor cortex and the primary motor cortex are responsible for movement and motor imagery. The spinal cord which controls muscles is directly connected to neurons in motor areas.

- The temporal lobe is responsible for audio data processing. Audio processing happens in primary auditory cortex. Amygdala and hippocampus which are located in temporal lobe as well are activated during the processing of emotions and memory.
- The parietal lobe is responsible for processing the sensory information and interpreting the visual information. Primary somatosensory cortex is responsible for the sense of touch.
- The occipital lobe is associated with visual data recognition. Primary visual cortex is responsible for basic visual information and is relayed directly from the eye to the cortex.

In order to develop BCI, knowledge about spatial distribution of signal when specific brain parts activated is necessary. This knowledge allows us to look for the signal in the right part of the brain. As was discussed previously, each part of the brain is responsible for different functions. For motor imagery BCI we concentrate on motor cortex area in the middle of the head towards the edge of frontal lobe.

Patterns of brain activity acquired from specific brain parts in order to drive BCI are called **neurophysiological signals**. It should be relatively easy to identify and to control by the user. Two main categories of neurophysiological signals are presented in literature:

- Evoked signals are produced as a reaction to external stimulations.
- Spontaneous signals are generated without any external stimulus.

2.4.2 Evoked potentials

Steady-State Evoked Potentials (SSEP) are brain activity patterns that appear when a subject perceives periodic stimuli such as repeated sound or flickering

light. Various kinds of SSEP are used for BCI, such as Steady-State Visual Evoked Potentials (SSVEP), which are by far the most used [48], somatosensory (touch) SSEP [49] or auditory SSEP [50]. Depending on the sort of the stimulation, SSEP appear in certain brain areas which corresponds to the nature of stimulation. SSEP generate brain signals of the same frequency as stimulus frequency.

An advantage of SSEP is that it can be used within a BCI without training. Moreover, since stimulus with different frequencies will lead to change of SSEP frequencies, it is possible to use a large number of different stimulus and commands. For example, a BCI application based on SSVEP can use flickering buttons with unique frequencies displayed on a screen, so that by focusing on certain button mainly located in the parietal area of the brain and used for communication BCI such as spellers.

P300 are positive waveforms that appear approximately 300 ms after auditory, visual, or somatosensory stimulus which is infrequent and significant. P300 is mainly located in the parietal area of the brain and used for communication BCI such as spellers. For example, the speller device with a 6x6 matrix containing 36 symbols [51]. When the rows and columns are randomly highlighted, the user is asked to focus on the intended symbol. In response to this focusing, a potential P300 is elicited in the brain. P300 does not require training sessions, but can be very tiring for the user and therefore not so practical.

2.4.3 Spontaneous signals

Slow Cortical Potentials (SCP) are slow voltage changes generated in the cortex [52]. SCPs are detectable in every human brain, even if the motor periphery is completely disconnected from the central nervous system. Subjects can learn to self-regulate their SCPs by producing either cortical positivity or negativity according to the task requirement. Positive shifts in SCPs are associated with the mental inhibition and slow negative shifts are associated with the mental

preparation.

SCP was used for spelling where subject could select letters or words with their SCPs [53]. The training sessions might take a long time for the subject to learn how to self-regulate SCPs.

Event-related synchronization and **event-related desynchronization** in Sensory Motor Rhythms (SMR) are brain rhythms related to motor actions. These rhythms are mainly located in mu (8-13 Hz) and beta (14-30 Hz) frequency bands over the motor cortex. Movement or preparation for movement is typically accompanied by a decrease in SMR [44, 54–57]. This decrease has been labeled as Event-Related Desynchronization (ERD). Its opposite, SMR increase or Event-Related Synchronization (ERS) occurs after movement and with relaxation.

Experiments in [58] when subjects were performing left and right hand motor imagery clearly shows ERD and ERS patterns in frequency bands 9-13 Hz and 10-12 Hz. It can be seen in Fig 2.7 that ERD/ERS patterns are recognizable during right (black) and left hand (gray) imagery movement. EEG signal was recorded using channels C3 and C4 (channels over the motor cortex) from the left and right sensorimotor areas.

The most important medical application for SMR or ERD/ERS-based BCI is to restore or replace the lost motor function [14–19]. Someone who has lost limb control could use a BCI to operate a powered wheelchair. Someone with a spinal cord injury whose arms and hands are paralyzed could use a BCI to control stimulation of the paralyzed muscles with implanted electrodes so that the muscles move the limbs. Other than that ERD/ERS-based BCI has received increasing attention for rehabilitation [9–13].

ERD/ERS patterns are widely used in EEG-based BCI during motor imagery. It is easy to detect ERD/ERS patterns in motor cortex area during motor imagery as well as it is easy for the user to control it. In this thesis, we focus on ERD/ERS

Figure 2.7: Event-related synchronization and event-related desynchronization during right (black) and left hand (gray) imagery movement [58].

brain activity patterns since the goal of this research is to improve the classification accuracy of MI-BCI.

2.5 Motor imagery

Motor imagery is a mental task when the subject is imagining the movement of some part of the body. Recently four parts of the body during motor imagery were explored: left hand, right hand, feet, and tongue.

2.5.1 Homunculus

Each part of the human body is represented in certain regions of the sensorimotor area of the cortex in the brain. It is the concept known as homunculus [59]. Topography in Fig 2.8 shows homunculus mapping from the body to the brain.

Figure 2.8: Homunculus, map of the body in the brain [40].

BCI is able to recognize different imaginary movement because of the homunculus. Since the signal is spread spatially according to the particular movement of the body, we use machine learning to teach a computer to recognize particular spatial activation in the brain. While some body parts located are too near to each other in the motor cortex, parts such as tongue, hands and feet are quite distinguishable. This neurocortex activation enabling to use more than two motor imagery. With better signal acquisition together with methods of machine learning, motor imagery BCI in the future can recognize most of the body movements.

2.5.2 Architecture of motor imagery brain-computer interfaces

The architecture of Motor Imagery Brain-Computer Interface (MI-BCI) is presented in Figure 2.9. The core components of a MI-BCI are as follows:

Preprocessing: Firstly, frequency filtering reduces noise which is present in the brain signals. Since typical EEG artifacts appear in low frequency ranges, frequency filtering removes artifacts as well. Secondly, continuous EEG signal is

Figure 2.9: Block diagram of a conventional MI-BCI system.

divided into trials after filtering using time markers of certain motor imagery tasks for training.

Feature extraction: The goal of this step is to extract the most discriminative characteristics of the data which differ among the classes. The foot, hand movement or tongue motor imageries produce distinguishable patterns in the motor cortex. Therefore features have to be extracted properly. A great variety of features have been used in the literature such as: amplitude values of EEG signals [60], band powers [61], power spectral density [62] values, autoregressive [63], adaptive

autoregressive [64] parameters, time-frequency features (wavelets) [65], principal component analysis (PCA) [66] or independent component analysis (ICA) [67]. The most popular approach is the common spatial pattern (CSP) [2, 40, 68–78]. As such, the thesis is concentrated on this method and it will be described in more details hereafter.

Classifier training: The features obtained from the training data are then used to train a classifier and thus end the training phase.

The testing phase: It is composed of a pre-processing stage similar to the one used in training followed by feature extraction. After this, the features are fed to the classifier that was previously trained in the training phase, producing a classification of the test input data.

Each of the components together with the methods used in this research is described in details in the next chapter.

Chapter 3

Signal Processing and Machine Learning for MI-BCI

3.1 Preprocessing

EEG signal is known to be very noisy as it is easily affected by electrical activity of different non-targeted parts of the body, such as eyes or muscles. After the data have been acquired, the signal is generally preprocessed in order to clean or de-noise it and to enhance relevant information. The goal of the preprocessing step is to increase the signal-to-noise ratio of the input signals.

3.1.1 Frequency filtering

The reason for frequency filtering comes from neurophysiological models which suggest that the signal is mainly located at a specific frequency. Therefore, it is advisable to reduce the frequency content of the EEG signal to some frequency band of interest. Since the ERD/ERS can be found at the mu and beta rhythms, applying frequency filter with this particular bandpass (8-30 Hz) to the signals before extracting bandpower features is very useful. Most of the methods therefore

use this frequency band.

In many MI-BCI systems, a bank of digital bandpass filters will be applied to extract different EEG rhythms from the raw data. In general, both finite impulse response (FIR) and infinite impulse response (IIR) filters can be used for the EEG signals. IIR filters are significantly more computationally efficient than FIR and at the same time introduce a relatively low delay in the system. As a result, in most motor-imagery BCI systems, IIR filters are used for filtering the data. Among them, Butterworth, Chebyshev, and elliptic filters are presented [76]. In this work, the 5th order Butterworth filter is used as a preprocessing step.

3.1.2 Time framing

In offline synchronous BCI, time framing is essential. Cue-paced or synchronous BCI is BCI where the subject is asked to perform a mental task for certain time from the moment of given cue or stimuli. Such a case is called a trial. Usually, time from 0.5 to 2.5 seconds from the beginning of the trial is considered as useful and considered as an informative signal for training and testing phases. The first 0.5 seconds are excluded as reaction time. After that only 2 seconds are usually taken even when the subject has been asked to perform a mental task for 4 seconds.

Figure 3.1: Time framing of the motor imagery trial. The trial begins at 2s after the visual cue.

In this work, we apply different time framing. In Chapter 4 we use conventional time framing 0.5 to 2.5 seconds. While in Chapter 5 whole trial time is used.

3.1.3 Channel selection

Channel selection is applied in order to focus only on certain spatial areas of the brain and then improve the classification accordingly. In most cases, channels around the motor cortex are chosen. Another reason for channel selection is to understand which channels can be eliminated in EEG systems of 60 or more electrodes without loss of information. This is useful for real-life applications where fewer electrodes are preferable.

Figure 3.2: Channel selection [79]. 14 EEG channels are selected (green color) from the 64 channels.

Many papers show improvements in classification rate by applying channel selection [80–82]. In this thesis, out of the 3 datasets we have tested, dataset IIIa which contains 60 electrodes can be considered for channel reduction. However, the other 2 datasets were recorded with 22 and 3 electrodes. Channel reduction for such small numbers of electrodes may cause loss of information. The goal of this work is to avoid loss of information and therefore we utilize all available EEG channels and do not apply channel reduction.

3.2 CSP based feature extraction

Common Spatial Pattern (CSP) [40] is one of the most popular feature extraction algorithms in MI-BCI. It finds spatial filters that maximize the variance of

bandpass-filtered EEG signal from one class while minimizing the variance from the other class. Spatial filters are used to isolate the relevant spatial information embedded in the signals by selecting or weighting the contributions from the different electrodes (different spatial regions). In spite of its good performance, CSP suffers from overfitting and sensitivity to noise and outliers. To enhance the performance of CSP, many variations of this method have been proposed in literature where regularization and spatio-spectral methods show good results. Common Spatio-Spectral Pattern (CSSP) [70] method applies CSP with finite impulse response (FIR) filter that uses a one-time delayed sample. Sub-band CSP [72] and Filter Bank CSP (FBCSP) [73] utilizes the EEG signal in multiple subbands, which significantly improves the performance of CSP.

A few variations of FBCSP were proposed in the papers [83–85]. Filter Bank Regularized CSP (FBRCS) [86] combines two methods - FBCSP and RCS. Discriminative Filter Bank CSP (DFBCSP) [87, 88] selects the subject-specific frequency bands by comparing fisher ratio of filtered EEG signal. Sliding Window Discriminative CSP (SWDCSP) [89] uses overlapping frequency filtering and clustering algorithm to find the best combination of CSP features. Sparse Filter Band CSP (SFBCSP) [90] exploited sparse regression for automatic overlapped frequency band selection. Local Region CSP (LRCSP) and Local Region Frequency-Optimized CSP (LRFCS) [91] consider local CSPs generated from individual channels and their local regions optimized by using filter banks.

To optimize spatial and FIR spectral filters Spectrally Weighted CSP (spec-CSP) [92] and Iterative Spatio-Spectral pattern Learning (ISSPL) [93] were presented. Optimal Spatio-Spectral Filter Network (OSSFN) [94] uses gradient-based learning algorithm and optimizes the filters by maximizing the mutual information between the feature vectors and the corresponding class labels. The aim of probabilistic Bayesian framework [95] is to find an optimized frequency band by particle-based Bayesian spatio-spectral filter optimization algorithm. Maximizing

the Mutual Information between Spatial Spectral Features (MMISS) method [96] proposes to optimize the spatial and spectral filters simultaneously in a unified mutual information-based framework. Sparse Bayesian Learning of Frequency Bands (SBLFB) [97] method finds the discriminative features from multiple overlapping subbands by sparse Bayesian learning. Separable Common Spatio-Spectral Pattern (SCSSP) [98] uses matrix-variate Gaussian model and jointly processes the data in both spectral and spatial domains.

Time-frequency feature optimization was applied for MI-BCI [99–101]. For the simultaneous optimization of filter bands and time window within CSP Temporally Constrained Sparse Group Spatial Pattern (TSGSP) [102] was presented. CSP/AM-BA-SVM [103] applies an optimum selection of time interval for each subject. Spectrally Weighted CSP (SWCSP) [104] proposes to solve optimization of time segment, channel configuration, and frequency bands jointly. Penalized Time-Frequency Band CSP (PTFBCSP) [105] selects automatically optimal subject-specific time-frequency bands.

To reduce overfitting of CSP, regularization was used. Regularized CSP (RCSP) [74] applies Tikhonov regularization to enforce the smoothness of the entries in the weighted spatial filters. The probabilistic CSP framework uses RCSP and statistical inference algorithms such as maximum a posteriori (MAP) estimation and sparse Bayesian learning [75]. CSP with quantized minimum error entropy criterion (CSP-QMEE) [106] structures a new cost function instead of the L2-norm in the conventional CSP.

Information geometry has become more popular in MI-BCI field. Riemannian geometry has been successfully used for motor imagery [107]. Minimum Distance to Riemannian Mean (MDRM) [108] uses geometric properties of the covariance matrix as the EEG descriptor and adopts the Riemannian distance to discriminate among them. CSP with Riemannian Tangent Space Mapping (CSP-TSM) [109] proposes to extract Riemannian tangent space features from the spatial filtered

EEG signal.

Deep Learning (DL) was recently applied in the BCI [110]. A deep learning approach with data augmentation method was proposed using two models — Convolutional Neural Network (CNN) and Wavelets Neural Network (WNN), to extract features and classify motor imagery signals in [111]. CNN architecture with a variational autoencoder (VAE) for classification [112] was proposed for motor imageries. Time stacked spatial EEG signal has been employed as the input to the Temporal Convolutional Network (TCN) in [113]. Despite the popularity of DL in many other areas, according to the comprehensive review [114] of 154 papers that apply DL to EEG, many challenges remain to be solved such as lack of new available EEG datasets and poor generalization between subjects.

3.2.1 Common spatial pattern

Common spatial pattern (CSP) [40] finds spatial filters that maximize the variance of one class and minimize the variance of the other class. Let a matrix $\mathbf{X} \in \mathbb{R}^{ch \times T}$ capture a trial of band-pass filtered EEG signals that characterizes one imaginary movement, where ch is the number of channels of EEG signals and T the signal length of a trial. The covariance matrix of a trial is computed by

$$\mathbf{\Sigma} = \mathbf{X}\mathbf{X}^T. \quad (3.1)$$

If the training set contains N_j trials for class Ω_j , the average normalized covariance matrix of class Ω_j is computed by

$$\mathbf{C}_j = \frac{1}{N_j} \sum_{\mathbf{\Sigma} \in \Omega_j} \frac{\mathbf{\Sigma}}{tr(\mathbf{\Sigma})}. \quad (3.2)$$

We can find a spatial filter $\mathbf{w} \in \mathbb{R}^{ch \times 1}$ that maximizes the variance of Ω_1 and minimizes it of Ω_2 by solving

$$\begin{aligned} \max J(\mathbf{w}) &= \frac{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{C}_2 \mathbf{w}}, \\ \text{or} \\ \max J(\mathbf{w}) &= \frac{\mathbf{w}^T \mathbf{C}_2 \mathbf{w}}{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}} \end{aligned} \quad (3.3)$$

Solution (3.3) is obtained by solving the generalized eigenvalue problem:

$$\begin{aligned} \mathbf{C}_1 \mathbf{w} &= \lambda \mathbf{C}_2 \mathbf{w}, \\ \text{or} \\ \mathbf{C}_2 \mathbf{w} &= \lambda \mathbf{C}_1 \mathbf{w}, \end{aligned} \quad (3.4)$$

where λ and $\mathbf{w}$ are the generalized eigenvalue and eigenvector. Equation (3.4) yields ch eigenvectors $\mathbf{w}_k$ and eigenvalues λ_k , $k = 1, \dots, ch$. The output of a spatial filter $\mathbf{w}_k$ to a trial of EEG signals is

$$\mathbf{y}_k = \mathbf{w}_k^T \mathbf{X}. \quad (3.5)$$

The logarithm of the filtered signal variance serves as a feature f_k for classification:

$$f_k = \log(\text{var}(\mathbf{y}_k)) = \log(\mathbf{y}_k \mathbf{y}_k^T) = \log(\mathbf{w}_k^T \mathbf{\Sigma} \mathbf{w}_k). \quad (3.6)$$

Hence, a feature vector $\mathbf{f} = [f_1, \dots, f_k, \dots, f_{ch}]$ produced by all spatial filters can be computed by

$$\mathbf{f} = \log(\text{diag}(\mathbf{W}^T \mathbf{\Sigma} \mathbf{W})), \quad (3.7)$$

where $\mathbf{f} \in \mathbb{R}^{ch \times 1}$ and $\mathbf{W}$ is the eigenvector matrix whose columns are eigenvectors from (3.4). All CSP based approaches in BCI select d features from $\mathbf{f}$, $d < ch$.

3.2.2 Common spatio-spectral pattern

Common spatio-spectral pattern (CSSP) [70] is an extension of CSP which uses the method of time delay embedding in order to optimize both filters: spectral and spatial. It doubles the number of signal channels by concatenating the signals $\mathbf{X}$ and their delayed version $\mathbf{X}_\tau^T$ as

$$\hat{\mathbf{X}} = [\mathbf{X}^T \ \mathbf{X}_\tau^T]^T. \quad (3.8)$$

Then the above CSP is applied to the new augmented data matrix $\hat{\mathbf{X}} \in \mathbb{R}^{2ch \times T}$ to produce features for classification. This method can be interpreted into a spatial and a spectral filter (finite impulse response (FIR) filter). The projected signal can be expressed as:

$$\begin{aligned} \mathbf{W}^T \mathbf{X} + \mathbf{W}_\tau^T \mathbf{X}_\tau &= \sum_{k=1}^{ch} \mathbf{w}_k^T \mathbf{X}_k + \mathbf{w}_{k-\tau}^T \mathbf{X}_{k-\tau} \\ &= \sum_{k=1}^{ch} \psi_k \left(\frac{\mathbf{w}_k^T}{\psi_k} \mathbf{X}_k + \frac{\mathbf{w}_{k-\tau}^T}{\psi_k} \mathbf{X}_{k-\tau} \right) \end{aligned} \quad (3.9)$$

where $\psi_k \in \mathbb{R}^{ch \times 1}$ is a pure spatial filter and $\left(\frac{\mathbf{w}_k^T}{\psi_k}, 0, \dots, 0, \frac{\mathbf{w}_{k-\tau}^T}{\psi_k} \right)$ denotes a FIR filter for each channel.

Even though the results showed an improvement of the CSSP algorithm over CSP, the flexibility of the frequency filters is still very limited. The choice of the proper time delay is another problem, which can be solved by prior information.

3.2.3 Regularized common spatial pattern

Regularized CSP (RCSP) [74] reduces the sensitivity of CSP to noise and overfitting by adding prior information and therefore regularizing it. Regularization can

be done at two levels: covariance matrix estimation level and objective function level.

Covariance matrix estimation with regularization becomes:

$$\begin{aligned} \tilde{\mathbf{C}}_j &= (1 - \alpha)\hat{\mathbf{C}}_j + \alpha\mathbf{I} \\ \text{with } \hat{\mathbf{C}}_j &= (1 - \beta)s_j\mathbf{C}_j + \beta\mathbf{G}_j \end{aligned} \quad (3.10)$$

where $\mathbf{C}_j$ is the covariance matrix as in 3.2, $\tilde{\mathbf{C}}_j$ is the regularized estimate, $\mathbf{I}$ is the identity matrix, s_j is a constant scaling parameter, $\mathbf{G}_j$ is a pre-built covariance matrix of EEG signal of considered mental state, α and $\beta \in [0, 1]$ are predefined regularized parameters. To reduce a possible estimation bias due to small training data, parameter α shrinks the initial covariance matrix toward the identity matrix. And for stable estimation, parameter β shrinks the initial covariance matrix toward a pre-built covariance matrix $\mathbf{G}_j$.

Pre-built covariance matrix $\mathbf{G}_j$ can be defined as average covariance matrix among all training data corresponding to the same mental state from other subjects.

$$\mathbf{G}_j = \sum_{i \in \Theta} \frac{N_j^i}{N_{tr,j}} \mathbf{C}_j^i \quad (3.11)$$

where Θ is a set of available subjects, N_j^i is the number of trials for class j and subject i used to estimate the covariance matrix $\mathbf{C}_j^i$, N_j is the total number of trials for the class.

At the objective function level, regularization terms can be added for penalizing non-satisfactory solutions. One of the best regularized CSP is Tikhonov Regularization CSP (TRCSP) [74]. Two objective functions with regularization are formulated as:

$$J_{r1}(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}}{\mathbf{w}^T (\mathbf{C}_2 + \delta \mathbf{K}) \mathbf{w}}, \quad (3.12)$$

$$J_{r2}(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{C}_2 \mathbf{w}}{\mathbf{w}^T (\mathbf{C}_1 + \delta \mathbf{K}) \mathbf{w}}, \quad (3.13)$$

where δ is the regularization parameter and $\mathbf{K}$ is predefined matrix. Here $\mathbf{K}=\mathbf{I}$ (identity matrix) is used as in [74]. Spatial filters are obtained by solving the generalized eigenvalue problems of J_{r1} and J_{r2} and features are produced by the spatial filters in the same way as CSP. RCSP selects features corresponding to the largest eigenvalues of J_{r1} and J_{r2} .

The algorithms use either regularization of the covariance matrices or the objective function, but not both. However, none of the combinations of the two approaches reached performances as high as that of the corresponding RCSP with a single form of regularization [74]. Therefore, RCSP with a single form of regularization are the simplest and most efficient algorithms. Regularizing CSP helps to overcome overfitting. On the other hand, it only uses spatial information and does not consider spectral and time information.

3.2.4 Filter bank common spatial pattern

Instead of temporal FIR filtering as presented in CSSP, filter bank CSP (FBCSP) [73] decompose the EEG signals into multiple frequency bands using filter bank as represented in Figure 3.3.

Overall, the method includes four stages such as frequency filtering, spatial filtering, feature selection, and classification. At the frequency filtering stage the EEG signal is bandpass-filtered into b frequency bands using zero-phase Chebyshev Type II filters. At the next stage, CSP features are extracted from each of the frequency bands (d features for each band). Following by feature selection stage where Mutual Information Best Individual Features (MIBIF) algorithm selects from a set of features $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{1 \times d}, \dots, \mathbf{f}_{b \times d}]$, where $\mathbf{f}_k \in \mathbb{R}^{N_j \times 1}$, $l = 4$ best

Figure 3.3: The architecture of filter bank CSP [73].

features for each class, so that a total number of selected features is $\tilde{\mathbf{F}} \in \mathbb{R}^{N_j \times l}$, where N_j is the number of trials for class.

The mutual information between class j and feature $\mathbf{f}_k$

$$I(\mathbf{f}_k; j) = H(j) - H(j|\mathbf{f}_k) = H(\mathbf{f}_k) - H(\mathbf{f}_k|j) \quad (3.14)$$

where $H(\mathbf{f}_k)$ is the entropy, and $H(\mathbf{f}_k|j)$ is the conditional entropy can be approximated by:

$$\begin{aligned} H(\mathbf{f}_k) &= -\frac{1}{N} \sum_{i=1}^N \log(p(\mathbf{f}_k^i)) \\ H(\mathbf{f}_k|j) &= -\frac{1}{N_j} \sum_{j=1,2} \log(p(\mathbf{f}_k^j|j)) \end{aligned} \quad (3.15)$$

where N denotes the total number of trials, and N_j denotes the number of trials of the class j , $p(\cdot)$ and $p(\cdot|\cdot)$ are probability functions estimated by Parzen window method. In the last stage, classification method Naive Bayesian Parzen Window (NBPW) is trained on the selected features.

Although FBCSP method performs well and was the winner of BCI competition, it has several disadvantages. Computationally it is a very heavy method. FBCSP uses feature selection which may cause loss of discriminative information. Additionally, time and frequency ranges used in the method are limited.

3.2.5 Sparse filter band common spatial pattern

Sparse filter band CSP (SFBCSP) [90] is the extension of FBCSP method as shown in Figure 3.4. At first, the method decomposes the EEG signals into 17 overlapping frequency subbands within 4-40 Hz using filter banks. Then CSP features are extracted from each of the frequency bands.

Figure 3.4: The architecture of sparse filter bank CSP [90].

Later, best CSP features are chosen by regression model:

$$u = \arg \min_u \frac{1}{2} \|\mathbf{G}\mathbf{u} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{u}\|_1, \quad (3.16)$$

where $\mathbf{u}$ is sparse vector from the Lasso penalty which is learned for feature selection, $\mathbf{y} \in \mathbb{R}^j$ is a vector with class labels 1,2, λ is a hyperparameter for regulating the sparsity of $\mathbf{u}$. Column vectors in $\mathbf{G}$ related to zero entries in $\mathbf{u}$ are excluded.

For classification, support vector machine (SVM) is used.

3.2.6 Sparse Bayesian learning of frequency bands

Sparse Bayesian learning of frequency bands (SBLFB) [97] is the extension of FBCSP method as represented in Figure 3.5.

Figure 3.5: The architecture of sparse Bayesian learning of frequency bands [97].

It decomposes the EEG signals into multiple frequency bands using filter bank. In total, 17 subbands from 4–40Hz with a width of 4Hz and overlapping rate of 2Hz. Then CSP features are extracted from each of the frequency bands. The

best CSP features are chosen by sparse Bayesian learning:

$$p(\mathbf{u}|\boldsymbol{\alpha}, \beta, \mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{u}, \beta)p(\mathbf{u}|\boldsymbol{\alpha})}{p(\mathbf{y}|\boldsymbol{\alpha}, \beta)} \quad (3.17)$$

where $\mathbf{u}$ is sparse weight vector, $\boldsymbol{\alpha}$ is a list of hyperparameters separately controlling the inverse variances of weights $\mathbf{u}$, β is inverse variance of zero mean Gaussian noise. Linear discriminant criterion is used for classification.

3.2.7 Improved discriminative filter bank common spatial pattern

Improved Discriminative Filter Bank CSP (IDFBCSP) [115] is extension of Discriminative FBCSP [87,88] method. Discriminative FBCSP is extension of FBCSP. The architecture of DFBCSP is shown in Figure 3.6 and the architecture of IDFBCSP is shown in Figure 3.7.

Figure 3.6: The architecture of discriminative filter bank CSP [115].

Both methods decompose the EEG signals into multiple frequency bands using filter bank. DFBCSP uses 12 bandpass filters with a bandwidth of 4 Hz in the range of 6 to 40 Hz. IDFBBCSP dividing EEG signal into 12 bandpass filters with 4–30 Hz range having bandwidth of 4 Hz with 2 Hz overlap (i.e. 4–8 Hz, 6–10 Hz, 8–12 Hz, ..., 26–30 Hz). Then CSP features are extracted from each of the frequency bands.

Figure 3.7: The architecture of improved discriminative filter bank CSP [115].

IDFBBCSP uses mutual information to choose the best filter bands with CSP for classification:

$$I(X, Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (3.18)$$

where $p(x, y)$ is the joint probability distribution function of X and Y , and $p(x)$

and $p(y)$ are the marginal probability distribution functions of X and Y , respectively. A larger mutual information value implies the corresponding feature has a greater predictive ability of the class membership (i.e. discriminating features).

The features obtained from all the bands are concatenated to form the feature vector $\mathbf{f}$. The feature matrix $\mathbf{F}$, is formed using the feature vectors of all the trials from the train data. The feature matrix is then utilized to determine the mutual information using 3.18, which gives $\mathbf{MI} = [I_1, I_2, \dots, I_K]$, where I_k is the mutual information value of the k -th feature. For feature selection, linear discriminant analysis (LDA) is applied. For classification, support vector machine (SVM) is used.

3.2.8 Common spatial pattern with tangent space mapping

CSP with Riemannian tangent space mapping (CSP-TSM) [109] uses Riemannian Tangent Space Mapping to extract features and then fusing it with CSP features. The Riemannian distance δ_R can be obtained:

$$\delta_R(\Sigma, \Sigma_i) = \|\text{Log}_{\Sigma}(\Sigma_i)\|_{\Sigma} = \|\text{upper}(\Sigma^{-1/2}\text{Log}_{\Sigma}(\Sigma_i)\Sigma^{-1/2})\|_2 = \|\mathbf{s}(i)\|_2 \quad (3.19)$$

where $\mathbf{s}(i)$ is a vector of the normalized tangent space. Each of the Riemannian mean Σ_i is mapped into the tangent space resulting in a set of vectors $\mathbf{s}(i)$ that are used as the tangent space features in the method. After that Least Absolute Shrinkage and Selection Operator (Lasso) is applied for feature selection. Linear Discriminant Analysis (LDA) is then used for dimensionality reduction of features. For classification support vector machine (SVM) is used.

Despite a good classification rate, all four methods above use feature selection with computationally heavy methods. It is the main disadvantage of these methods.

3.2.9 Temporally constrained sparse group spatial pattern

Temporally constrained sparse group spatial pattern (TSGSP) [102] uses frequency bands and time windows to extract CSP. It is an extension of SFBCSP methods which additionally uses a sliding time window. Figure 3.8 demonstrates the architecture of the method.

Figure 3.8: The architecture of temporally constrained sparse group spatial pattern [102].

After extracting CSP features for each time and frequency bands, a joint sparse optimization with temporal smoothness constraint is used to select features:

$$u = \arg \min_u \frac{1}{2} \sum_{t=1}^T \|\mathbf{G}^{(t)} \mathbf{u}_t - \mathbf{y}\|_2^2 + \lambda_1 \|\mathbf{U}\|_1 + \lambda_2 \|\mathbf{U}\|_{2,1} + \lambda_3 \sum_{t=1}^T \|\mathbf{u}_t - \mathbf{u}_{t+1}\|_1, \quad (3.20)$$

where $\mathbf{u}$ is sparse vector from the Lasso penalty which is learned for feature selection, $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_T] \in \mathbb{R}^{D \times T}$, sliding window $t = 1, 2, \dots, T$, D is the dimensionality of features, $\mathbf{y} \in \mathbb{R}^N$ is a vector with class labels 1,2, λ is a hyperparameter for regulating the sparsity of $\mathbf{u}$. Column vectors in feature set $\mathbf{G}$ related to zero entries in $\mathbf{u}$ are excluded. For classification, a linear support vector machine classifier is trained.

3.2.10 Local region frequency-optimized common spatial pattern

Local region CSP (LRCSP) and Local region frequency-optimized CSP (LRFCS) [91] are methods that construct a subset of the local regions of EEG channels to extract CSP features. It is done that way because different local regions of the brain are activated for different imagery movements. Figure 3.9 and figure 3.10 demonstrate the architecture of the methods.

Figure 3.9: The architecture of local region common spatial pattern [91].

Figure 3.10: The architecture of local region frequency optimized common spatial pattern [91].

LRCSP defines a set of channels or local region $\mathbb{R}^h \in (1, 2, \dots, ch)$ as main channel

ch and its immediate neighbor channels. Then it finds CSP filters $\mathbf{w}_{ch}$ for each region. The local regions with significant disparity among variance ratios are considered as highly relevant to the given MI task and used for feature extraction:

$$\mathbf{f} = \left[\log(\text{var}(\mathbf{w}_{max}^{(ch)T} \mathbf{y}^{(ch)})) \quad \log(\text{var}(\mathbf{w}_{min}^{(c)T} \mathbf{y}^{(ch)})) \right], \quad (3.21)$$

where $\mathbf{w}_{max}^{(ch)}$ and $\mathbf{w}_{min}^{(c)}$ are CSP filters that maximizing or minimizing the averaged variance from local region $\mathbb{R}^{ch}$, $\mathbf{y}^{(ch)}$ is query EEG from local region $\mathbb{R}^{ch}$. To select good local regions, the method uses two criteria: the distance between the maximum and the minimum variance ratios, and the distance between the means of within-class features.

LRFCSP divides EEG signal into eight frequency bands evenly with 4Hz within the range from 4 to 36Hz. For each local region, the best two frequency bands, which may vary with the local region, are selected based on the mutual information criterion of the filter-bank CSP algorithm:

$$\begin{aligned} I(\mathbf{f}_{i,max}^{(ch,b)}; j) &= H(j) - \sum_{j=1}^2 \sum_{i=1}^N p(j|\mathbf{f}_{i,max}^{(ch,b)}) \log_2(p(j|\mathbf{f}_{i,max}^{(ch,b)})), \\ I(\mathbf{f}_{i,min}^{(ch,b)}; j) &= H(j) - \sum_{j=1}^2 \sum_{i=1}^N p(j|\mathbf{f}_{i,min}^{(ch,b)}) \log_2(p(j|\mathbf{f}_{i,min}^{(ch,b)})), \end{aligned} \quad (3.22)$$

where j is classlabel, b is frequency band, i is number of trial, N is all trials for class j , and:

$$\begin{aligned} \mathbf{f}_{i,max}^{(ch,b)} &= \log(\text{var}(\mathbf{w}_{i,max}^{(ch,b)T} \mathbf{y}^{(ch,b)})), \\ \mathbf{f}_{i,min}^{(ch,b)} &= \log(\text{var}(\mathbf{w}_{i,min}^{(ch,b)T} \mathbf{y}^{(ch,b)})), \end{aligned} \quad (3.23)$$

Mutual information $I(\mathbf{f}_{i,max}^{(ch,b)}; j)$ are computed based on estimated $p(j|\mathbf{f}_{i,max}^{(ch,b)})$ and $p(j|\mathbf{f}_{i,min}^{(ch,b)})$ using Parzen window. The feature vector representing a frequency-optimized local region is defined as a four dimensional vector by concatenating

the local CSP feature vectors for two frequency bands. For LRFCSP, the good local regions are selected similarly to the LRCSP. For classification, support vector machine (SVM) is used.

3.3 Classification

The goal of the classification step is to identify features obtained in the previous step as a class. In this section, we present different classification algorithms used in BCI.

3.3.1 Linear discriminant analysis

The linear discriminant analysis (LDA) classifier is the most widely used classifier for MI-BCI [40, 70, 74, 75]. It is calculated as follows:

$$\begin{aligned} \mathbf{a} * \mathbf{y}^T - z &= l; \\ \text{with } \mathbf{a} &= \frac{(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)}{\mathbf{S}_{F_1} + \mathbf{S}_{F_2}} \\ z &= (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) * \mathbf{a}^T / 2 \end{aligned} \tag{3.24}$$

where $\mathbf{y} \in \mathbb{R}^{1 \times d}$ is band-pass and spatially filtered trial for classification, $\boldsymbol{\mu}_j \in \mathbb{R}^{1 \times d}$ is mean of $\mathbf{F}_j$, $\mathbf{F}_j$ consists of feature vectors $\mathbf{f}$ from each trial of training set N_j of class Ω_j and $\mathbf{S}_{F_j} \in \mathbb{R}^{d \times d}$ is covariance matrix of $\mathbf{F}_j$. The decision rule: $\mathbf{y} \rightarrow j = 1$ if $l < 0$ or, $\mathbf{y} \rightarrow j = 2$ if $l > 0$

3.3.2 Naive Bayesian Parzen window

Naive Bayesian Parzen window (NBPW) is used with FBCSP [73]. The classification rule of NBPW classifier is:

$$k = \operatorname{argmax} p(j|\mathbf{y}) \tag{3.25}$$

Using Bayes rule

$$p(j|\mathbf{y}) = \frac{p(\mathbf{y}|j)P(j)}{p(\mathbf{y})} \quad (3.26)$$

where $p(j|\mathbf{y})$ denotes posterior probability of class j given the random trial $\mathbf{y}$; $p(\mathbf{y}|j)$ is the conditional probability of $\mathbf{y}$ given class j and $P(j)$ is the prior probability of class j estimated from training data; and $p(\mathbf{y})$ is:

$$p(\mathbf{y}) = \sum_{j=1,2} p(\mathbf{y}|j)P(j) \quad (3.27)$$

and an estimate of $p(\mathbf{y}|j)$ using Parzen Window:

$$p(\mathbf{y}|j) = \frac{1}{N_j} \sum_{i=1}^{N_j} \varphi(\mathbf{y} - \mathbf{y}_i, h) \quad (3.28)$$

where φ is a window function and h is a window width parameter determining the smoothness of the window function:

$$\begin{aligned} \varphi(\mathbf{y}, h) &= \frac{1}{(2\pi)} e^{-\left(\frac{\mathbf{y}^2}{2h^2}\right)} \\ h &= \left(\frac{4}{3N_j}\right)^{1/5} \sigma \end{aligned} \quad (3.29)$$

where σ denotes the standard deviation of the distribution.

3.3.3 Support vector machine

Support vector machine (SVM) is a very popular classification method in BCI [90, 91, 109, 115] that uses discriminant hyperplane. Figure 3.11 demonstrates the hyperplane and support vectors (SV).

Figure 3.11: Illustration of support vector machine classifier [116].

The goal of SVM classifier is to find hyperplane with a large edge or distance between data points of different classes. To define hyperplane, weight vector $\mathbf{w} \in \mathbb{R}^{d \times 1}$ and offset $b \in \mathbb{R}$ are used. Training set is assumed as $(x_i, j_i), i = 1, 2, \dots, N, x \in \mathbb{R}^{1 \times d}$, where $j_i \in -1, 1$ same as $j_i \in 1, 2$. To find hyperplane, the discriminant function for both classes is:

$$\begin{aligned} \mathbf{x}_i * \mathbf{w} + b &\geq 1, \text{ for } j_i = 1 \\ \mathbf{x}_i * \mathbf{w} + b &\leq -1, \text{ for } j_i = -1 \end{aligned} \quad (3.30)$$

If it is combined:

$$j_i(\mathbf{x}_i * \mathbf{w} + b) - 1 \geq 0 \quad (3.31)$$

The optimization problem of finding the $\mathbf{w}$ vector with minimum norm that will lead to maximum hyperplane edge can be solved by using the multipliers method of Lagrange:

$$L_p = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i j_i (\mathbf{x}_i * \mathbf{w} + b) + \sum_{i=1}^N \alpha_i, \quad (3.32)$$

where α_i are the positive coefficients for each vector of the training set. This formula should be minimized for parameters $\mathbf{w}$ and b . Vectors for which $i > 0$ will become support vector. After SVM classifier has been trained, during the testing a vector $\mathbf{y}$ as input will be considered as class 1 or class -1 by the $\text{sgn}(\mathbf{w}\mathbf{y} + b)$.

3.3.4 Minimum Mahalanobis distance

Minimum Mahalanobis distance classifier is calculated as follows:

$$c_j = (\mathbf{y} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_{F_j}^{-1} (\mathbf{y} - \boldsymbol{\mu}_j) \quad (3.33)$$

where $\mathbf{y} \in \mathbb{R}^{1 \times d}$ is band-pass and spatially filtered trial for classification, $\boldsymbol{\mu}_j \in \mathbb{R}^{1 \times d}$ is mean of $\mathbf{F}_j$, $\mathbf{F}_j$ consists of feature vectors $\mathbf{f}$ from each trial of training set N_j of class Ω_j , and $\boldsymbol{\Sigma}_{F_j} \in \mathbb{R}^{d \times d}$ is covariance matrix of $\mathbf{F}_j$.

For classification, distance c is estimated as:

$$c = c_1 - c_2 \quad (3.34)$$

The decision rule: $\mathbf{y} \rightarrow j = 1$ if $c < 0$ or, $\mathbf{y} \rightarrow j = 2$ if $c > 0$

It is not difficult to prove that the LDA classifier is identical to the minimum Mahalanobis distance classifier using pooled class-conditional covariance matrix, also called within-class scatter matrix of the feature vector. If instead of one covariance matrix, the summation of both is used, Mahalanobis distance changes

to:

$$c_j = (\mathbf{y} - \boldsymbol{\mu}_j)^T (\boldsymbol{\Sigma}_{F_1} + \boldsymbol{\Sigma}_{F_2})^{-1} (\mathbf{y} - \boldsymbol{\mu}_j) \quad (3.35)$$

The equation above is completely identical to LDA method described in section 3.3.1. It can be seen after certain transforms as follows:

$$c = c_1 - c_2 = (\mathbf{y} - \boldsymbol{\mu}_1)^T (\boldsymbol{\Sigma}_{F_1} + \boldsymbol{\Sigma}_{F_2})^{-1} (\mathbf{y} - \boldsymbol{\mu}_1) - \quad (3.36)$$

$$-(\mathbf{y} - \boldsymbol{\mu}_2)^T (\boldsymbol{\Sigma}_{F_1} + \boldsymbol{\Sigma}_{F_2})^{-1} (\mathbf{y} - \boldsymbol{\mu}_2) = \quad (3.37)$$

$$\frac{2\mathbf{y}\boldsymbol{\mu}_2 - 2\mathbf{y}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_1^2 - \boldsymbol{\mu}_2^2}{\boldsymbol{\Sigma}_{F_1} + \boldsymbol{\Sigma}_{F_2}} = \quad (3.38)$$

$$\frac{\mathbf{y}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)}{\boldsymbol{\Sigma}_{F_1} + \boldsymbol{\Sigma}_{F_2}} - \frac{(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 + \boldsymbol{\mu}_1)}{2(\boldsymbol{\Sigma}_{F_1} + \boldsymbol{\Sigma}_{F_2})} \quad (3.39)$$

where c becomes equal to l as in 3.24.

Despite the number of different methods and enhancements, there is still no method that works best for different subjects, datasets and setups. Therefore, the main goal of this research is to enhance classification and find the best methods for motor imagery BCI.

Chapter 4

Feature Weighting and Regularization

EEG signals are known to have very low spatial resolution. Moreover, electrodes capture signals that are overlapping with each other. To extract the discriminative features and alleviate the overfitting problem for motor imagery BCI, spatial filtering is widely applied. However, in most cases only a few common spatial patterns (CSP) are selected as features and the others are ignored. Using only a few CSP features though alleviates overfitting problem, loses the discriminating information. The loss of discriminating information by discarding most CSP features may adversely affect the classification accuracy.

Motivated by preserving discriminating information, we propose a novel Feature Weighting and Regularization (FWR) method that utilizes all CSP features to avoid information loss. The proposed method can be applied in all CSP-based approaches. The experiments of this chapter show the effect of the proposed method applied in the standard CSP and its two extensions, CSSP and RCSP. Results on BCI Competition III Dataset IIIa and IV Dataset IIa demonstrate that the proposed FWR method enhances the classification accuracy compared to the

conventional Feature Selection (FS) approaches.

4.1 Proposed method

All CSP based approaches in MI-BCI select the features obtained by the spatial filters. The total number of possible features is the number of EEG channels, ch . However, literatures show that the best classification accuracy is achieved at very few number of features, such as 2 to 8 in most cases compared to, for example, $ch = 60$. The loss of discriminating information by discarding most CSP features may adversely affect the classification accuracy [117–120]. To alleviate this problem, we propose a novel feature weighting and regularization (FWR) method that utilizes all CSP features for classification. As shown in [119, 120], overfitting problem can be alleviated not only by removing the unreliable features, but also by regularizing the unreliable features.

4.1.1 Feature extraction by common spatial patterns

Let a matrix $\mathbf{X} \in \mathbb{R}^{ch \times T}$ capture a trial of band-pass filtered EEG signals that characterizes one imaginary movement, where ch is the number of channels of EEG signals and T is the signal length of a trial. The covariance matrix of a trial is computed by

$$\mathbf{\Sigma} = \mathbf{X}\mathbf{X}^T. \quad (4.1)$$

If the training set contains N_j trials for class Ω_j , the average normalized covariance matrix of class Ω_j is computed by

$$\mathbf{C}_j = \frac{1}{N_j} \sum_{\mathbf{\Sigma} \in \Omega_j} \frac{\mathbf{\Sigma}}{tr(\mathbf{\Sigma})}. \quad (4.2)$$

We can find a spatial filter $\mathbf{w} \in \mathbb{R}^{ch \times 1}$ that maximizes the variance of Ω_1 and minimizes it of Ω_2 by solving

$$\max J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}}{\mathbf{w}^T (\mathbf{C}_1 + \mathbf{C}_2) \mathbf{w}}. \quad (4.3)$$

Solution (4.3) is obtained by solving the generalized eigenvalue problem:

$$\mathbf{C}_1 \mathbf{w} = \lambda (\mathbf{C}_1 + \mathbf{C}_2) \mathbf{w}, \quad (4.4)$$

where λ and $\mathbf{w}$ are the generalized eigenvalue and eigenvector. Equation (4.4) yields ch eigenvectors $\mathbf{w}_k$ and eigenvalues λ_k , $k = 1, \dots, ch$.

In fact, equation (4.4) is different from equation (3.4) that was explained in previous chapter. Using equation (3.4) that was originally proposed for CSP will lead to eigenvalues of range: $0 \leq \lambda \leq \infty$, where λ is the generalized eigenvalues of 3.4. So that largest eigenvalues up to ∞ will correspond to one class and smallest near 0 to another, while middle range is 1. This makes it difficult to understand which features are more discriminative for one or another class. Another disadvantage of this approach is that the objective function is very sensitive to the covariance matrix because it is multiplied by its inverse. If covariance matrix of one class is small due to unreliable information such as small energy of the trial, then multiplying by its inverse will cause large misclassification rate.

While there is no difference in outcome results between using equation (3.4) or equation (4.4) we prefer to use the second one. It solves two disadvantages mentioned earlier. The largest and the smallest eigenvalues of 4.4 correspond to the largest differences of variances of the two classes and $0 \leq \lambda \leq 1$, where λ is the generalized eigenvalues of 4.4. So that largest eigenvalues close to 1 will correspond to one class and smallest near 0 to another, while middle range is 0.5. Therefore in order to find most discriminative features from 4.4, features are sorted in descending order of $|\lambda - 0.5|$. Secondly, if covariance matrix of one class is small due

to unreliable information it will not affect the objective function.

The output of a spatial filter $\mathbf{w}_k$ to a trial of EEG signals is

$$\mathbf{y}_k = \mathbf{w}_k^T \mathbf{X}. \quad (4.5)$$

The logarithm of the filtered signal variance serves as a feature f_k for classification,

$$f_k = \log(\text{var}(\mathbf{y}_k)) = \log(\mathbf{y}_k \mathbf{y}_k^T) = \log(\mathbf{w}_k^T \mathbf{\Sigma} \mathbf{w}_k). \quad (4.6)$$

Hence, a feature vector $\mathbf{f} = [f_1, \dots, f_k, \dots, f_{ch}]$ produced by all spatial filters can be computed by

$$\mathbf{f} = \log(\text{diag}(\mathbf{W}^T \mathbf{\Sigma} \mathbf{W})), \quad (4.7)$$

where $\mathbf{f} \in \mathbb{R}^{ch \times 1}$ and $\mathbf{W}$ is the eigenvector matrix whose columns are eigenvectors from (4.4).

4.1.2 Feature weighting

All existing CSP based approaches in BCI select d features and remove the other $ch - d$ features. This is equivalent to weighting the features by a step function that has sharp transition from 1 to 0:

$$f_k^\gamma = \gamma_k f_k \quad (4.8)$$

where γ_k :

$$\gamma_k = \begin{cases} 1, & k \leq d \\ 0, & k > d \end{cases} \quad (4.9)$$

The weighted feature vector is $\mathbf{f}_\gamma = [f_1^\gamma, \dots, f_k^\gamma, \dots, f_d^\gamma]^T$.

Although it is known that using step function and removing CSP features may help problems such as the curse of dimensionality, small sample size problem, noise removal effect, and better generalization in a lower dimensional space, it is not straightforward which features should be extracted and which should be removed for a more robust classification. A common problem of all CSP approaches is that they all lose some discriminative information by removing features. In fact, the discriminative information can reside in all features.

On the other hand, noise disturbance and poor estimates of small eigenvalues due to the finite number of training samples are the culprits. The limited number of training trials results in very small eigenvalues that may not represent the true variance well enough, and corresponding to it CSP features are unreliable. Therefore, we propose to weight and regularize features instead of removing it.

We propose to weight the features by a soft function. As the inverse function $1/k$ has long tail and the sigmoid function needs two parameters to control its shift and smoothness, this work uses Gaussian function that has short tail to weight the features:

$$\gamma_k = e^{-\frac{(k-1)^2}{2\sigma^2}}, \quad (4.10)$$

$$f_k^\gamma = \gamma_k f_k, \quad (4.11)$$

where σ is a parameter, which can be chosen so that the Gaussian function decays the fastest between m and $m + 1$. This leads to $\sigma = m - 0.5$, where m is the Eigenfeature Regularization and Extraction (ERE) regularization parameter shown later. The weighted feature vector is $\mathbf{f}_\gamma = [f_1^\gamma, \dots, f_k^\gamma, \dots, f_{ch}^\gamma]^T$. Figure 4.1 shows difference between soft and step function.

As can be seen from the figure 4.1, instead of removing features we are weighting them according to their position. Since initially we have sorted features so

Figure 4.1: Soft and step functions.

that most discriminative features are located first, these features will have highest ranking. Since Gaussian function decays rapidly after $\sigma = m - 0.5$ and first features will have weight 1, it becomes similar to the step function approach. The main difference is that instead of removing all other features, we weight them with smallest rank and include them.

4.1.3 Feature regularization and classification

The linear discriminant analysis (LDA) classifier is the most widely used classifier for MI-BCI [40, 70, 74, 75]. More information about this method can be found in section 3.3.1. It is not difficult to prove that the LDA classifier is identical to the minimum Mahalanobis distance classifier using pooled class-conditional covariance matrix $\mathbf{S}_w$, also called within-class scatter matrix of the feature vector. The proof can be found in 3.3.4. The Mahalanobis distance d_j of a feature vector $\mathbf{f}_\gamma$ to the class mean $\boldsymbol{\mu}_j$ of Ω_j is evaluated by

$$d_j = (\mathbf{f}_\gamma - \boldsymbol{\mu}_j)^T \mathbf{S}_w^{-1} (\mathbf{f}_\gamma - \boldsymbol{\mu}_j) \quad (4.12)$$

As LDA classifier or the minimum Mahalanobis distance classifier whitens the feature vector by its within-class scatter matrix, feature weighting will not affect the classification results, i.e. feature vector $\mathbf{f}$ and $\mathbf{f}_\gamma$ will lead to the same classification results. However, feature weighting will greatly affect the classification results if we regularize the classifier because the regularization techniques modify the feature variance differently based on its value.

Eigen-decomposition provides an effective tool to simplify the problem. It provides a powerful tools to study, analyze and find problems of the high dimensional matrix. As shown in [119, 121], the Mahalanobis distance can be evaluated in the eigenspace in scalar form as:

$$d_j = (\mathbf{f}_\gamma - \boldsymbol{\mu}_j)^T \mathbf{S}_w^{-1} (\mathbf{f}_\gamma - \boldsymbol{\mu}_j) = \sum_{i=1}^{ch} \frac{(z_i - \bar{z}_{ji})^2}{\lambda_i}, \quad (4.13)$$

where

$$z_i = \boldsymbol{\phi}_i^T \mathbf{f}_\gamma, \quad (4.14)$$

$$\bar{z}_{ji} = \boldsymbol{\phi}_i^T \boldsymbol{\mu}_j, \quad (4.15)$$

$\boldsymbol{\phi}_i$ and λ_i are the i th eigenvector and eigenvalue of $\mathbf{S}_w$:

$$\mathbf{S}_w = \sum_{\mathbf{f}_\gamma \in \Omega_1} (\mathbf{f}_\gamma - \boldsymbol{\mu}_1)(\mathbf{f}_\gamma - \boldsymbol{\mu}_1)^T + \sum_{\mathbf{f}_\gamma \in \Omega_2} (\mathbf{f}_\gamma - \boldsymbol{\mu}_2)(\mathbf{f}_\gamma - \boldsymbol{\mu}_2)^T. \quad (4.16)$$

where $\boldsymbol{\phi}_i$ and λ_i are found by solving the generalized eigenvalue problem:

$$\mathbf{S}_w \boldsymbol{\phi} = \lambda \mathbf{S}_w \boldsymbol{\phi}. \quad (4.17)$$

If we define a whitened eigenfeature as $z_i/\sqrt{\lambda_i}$, the classifier becomes a simple minimum Euclidian distance classifier.

As all CSP features are used in the classification, we need to regularize the classi-

fier, or equivalently, regularize the features to alleviate the problems of instability, overfitting or poor generalization. From 4.13 it is evident that if some eigenvalues are unreliable, especially if some eigenvalues are small and approach to zero it will cause problems. So that small noise disturbances that have little effect on the initial portion of the eigenspectrum cause large vibrations of the inverse eigenspectrum in the region of small eigenvalues. Therefore, we propose to modify or regularize the unreliable eigenvalues for better generalization.

There are a few regularization techniques in the literature [119], such as “adding a constant to all eigenvalue” [122], the probabilistic subspace learning and its enhanced version [123], and ERE eigenspectrum model [120].

The curve of cyan color in Fig 4.2 shows the eigenspectrum before regularization of a sample BCI training task. To clearly show the differences between various regularized eigenspectrum, the very small eigenvalues are out of the plot range in Fig 4.2. The very small eigenvalues cause great trouble for classification, which need be alleviated by various regularization techniques [119, 120, 122, 123]. Both the probabilistic subspace learning and ERE keep the eigen-values from 1 to m unchanged [119, 120, 123]. The probabilistic subspace learning uses a constant to replace the small eigen-values from $m + 1$ to ch [119, 123]. It takes the average of the eigenvalues from $m + 1$ to ch as the constant, as shown by the black curve in Fig 4.2. As this is undesired, we propose to enhance it by using the maximum eigenvalues from $m + 1$ to ch as the constant [119, 123] shown by the blue curve in Fig 4.2, which greatly improves the classification performance [119, 123].

ERE further enhances the classification performance by using a model that fits the eigen-values from 1 to m to predict the eigenvalues from $m + 1$ to ch [120]. The predicted eigen-values by ERE from $m + 1$ to ch are not a constant but decrease as the original ones but at a much slower rate, as shown by the red curve in Fig 4.2. ERE is a further development from the probabilistic subspace learning and many papers show that ERE is better than the probabilistic subspace

Figure 4.2: Raw eigen-values (cyan), adding a constant (green), probabilistic subspace learning (black), enhanced probabilistic subspace learning (blue), and ERE (red) and ERE+constant (yellow) of a sample BCI training task. Note that the yellow curve is invisible as it is visually overlapped by the red curve.

learning [120,124–126]. Therefore, we choose ERE and combine it with the “adding a constant to all eigenvalues” to strengthen the regularization effect in this work.

Often population variances first decay very rapidly and then stabilize so that the smallest eigenvalues are biased much more severely than the largest ones. The large deviations of the small eigenvalues from the population variances result in a severe overfitting problem of the classifier that adversely affects the classification accuracy. The decay of the eigenvalues should be slowed down to compensate the effect of noise and the finite number of training samples. This can be done by replacing the eigenspectrum with the proposed model:

$$\tilde{\lambda}_i = \begin{cases} \lambda_i + c, & i < m \\ \frac{\alpha}{i+\beta} + c, & m \leq i \leq ch \end{cases} \quad (4.18)$$

where

$$\alpha = \frac{\lambda_1 \lambda_m (m-1)}{\lambda_1 - \lambda_m}, \quad (4.19)$$

$$\beta = \frac{m \lambda_m - \lambda_1}{\lambda_1 - \lambda_m}. \quad (4.20)$$

where α and β are two constants determined by the reliable eigenvalues. The above eigenvalues λ_i of $\mathbf{S}_w$ are sorted in descending order before the regularization. c and m are two parameters to choose for regularization. This work chooses $c = \lambda_{ch/2}$ to reduce the free parameter. It is used to ensure a minimum amount of regularization [119, 122] in case ERE fades caused by a large value of m .

To further strength the regularization effect we combine ERE with adding a constant because ERE and the probabilistic subspace learning are exclusive, not combinable. Fig 4.2 shows the raw Eigen-values (cyan), adding constant (green), probabilistic subspace learning (black), enhanced probabilistic subspace learning (blue), ERE (red) and ERE+constant (yellow) of a sample BCI training task. As the constant is chosen as the eigenvalue of $ch/2$, we cannot see the difference between ERE and ERE+constant in Fig 4.2. Therefore, the regularization is mainly contributed by ERE but +constant ensures a minimum amount of regularization for whatever value of m chosen for ERE.

The above procedure of feature weighting and regularization can be applied in all CSP based approaches of MI-BCI to replace the conventional feature selection. The proposed feature weighting and regularization (FWR) approach is summarized below:

At the training stage:

1. Compute and sort the spatial filters $\mathbf{W}$ by CSP algorithm or one of its extensions.
2. Apply all ch spatial filters to EEG signals to extract feature vector $\mathbf{f}$ as Eq.

(4.6) or Eq. (4.7).

3. Compute the weighted feature vector $\mathbf{f}_\gamma$ by Gaussian function as Eq. (4.10) and Eq. (4.11).
4. Compute the within-class scatter matrix $\mathbf{S}$ of feature vector $\mathbf{f}_\gamma$ and regularize the descending order sorted eigenvalues λ_i by Eq. (4.18), Eq. (4.19) and Eq. (4.20) to get the regularized eigenvalues $\tilde{\lambda}_i$.

At the recognition stage:

1. Apply the steps (2) and (3) of the training stage to compute the weighted feature vector $\mathbf{f}_\gamma$ for the query trial of EEG signals.
2. Classify the weighted feature vector $\mathbf{f}_\gamma$ by minimum Mahalanobis distance classifier Eq. (4.13) but using the regularized eigenvalues $\tilde{\lambda}_i$ instead of λ_i

4.2 Experiments

4.2.1 EEG datasets

EEG data from two publicly most widely used motor imagery BCI datasets are used:

- 1) BCI competition IV, Dataset IIa [127] contains EEG data of 22 channels recorded from 9 healthy subjects(A01-A09). The location of the electrodes can be seen in Figure 4.3. The cue-based BCI paradigm consisted of four different motor imagery tasks, namely the imagination of movement of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). Two sessions on different days were recorded for each subject. Each session is comprised of 6 runs separated by short breaks. One run consists of 48 trials (12 for each of the four possible classes), yielding a total of 288 trials per session or 72 trials per class for each subject for training and testing. The EEG signal was sampled with 250

Hz and later bandpass filtered between 0.5Hz and 100 Hz (with 50 Hz notch filter enabled).

Figure 4.3: Position of EEG electrodes in Dataset Ila [127].

The subjects were sitting in a comfortable armchair in front of a computer screen. At the beginning of a trial ($t = 0s$), a fixation cross appeared on the black screen. In addition, a short acoustic warning tone was presented. After two seconds ($t = 2s$), a cue in the form of an arrow pointing either to the left, right, down or up (corresponding to one of the four classes - left hand, right hand, foot or tongue) appeared and stayed on the screen for 1.25s. This prompted the subjects to perform the desired motor imagery task. No feedback was provided. The subjects were asked to carry out the motor imagery task until the fixation cross disappeared from the screen at $t = 6s$. A short break followed where the screen was black again. Figure 4.4 illustrates the timing scheme of the dataset. We use time point $t = 2s$ as the beginning of the trial.

2) BCI competition III, Dataset IIIa [128] contains EEG signals of 60 channels recorded from 3 subjects. The location of the electrodes can be seen in Figure 4.5. The data set contains 45, 30 and 30 trials per class with time point markers for

Figure 4.4: Timing scheme of the paradigm in Dataset IIa [127].

training and testing for subjects k3b, k6b and l1b, respectively. The EEG signal was sampled with 250 Hz and filtered between 1Hz and 50 Hz with notch filter enabled.

Figure 4.5: Position of EEG electrodes in Dataset IIIa [128].

The subject sat in a relaxing chair with armrests. The task was to perform imagery left hand, right hand, foot or tongue movements according to a cue. The order of cues was random. The experiment consists of several runs (at least 6) with 40 trials each. After trial begins, the first 2s were quite, at $t = 2s$ an acoustic stimulus indicated the beginning of the trial, and a cross + is displayed; then

from $t = 3s$ an arrow to the left, right, up or down was displayed for 1s; at the same time the subject was asked to imagine a left hand, right hand, tongue or foot movement, respectively, until the cross disappeared at $t = 7s$. Each of the 4 cues was displayed 10 times within each run in a randomized order. Figure 4.6 illustrates the timing scheme of the dataset. We use time point $t = 3s$ as the beginning of the trial.

Figure 4.6: Timing scheme of the paradigm in Dataset IIIa [128].

Both datasets contain four motor imagery actions of the movement: left hand, right hand, foot and tongue. Since CSP based approaches work on discriminating two classes, we pair the 4 MIs to construct 6 binary classification tasks: task 1 - left-hand versus right-hand, task 2 - left-hand versus foot, task 3 - left-hand versus tongue, task 4 - right-hand versus foot, task 5 - right-hand versus tongue and task 6 - foot versus tongue.

4.2.2 Preprocessing and experiment settings

At first, the raw EEG signals are filtered between 8-30 Hz as in [74] with fifth-order Butterworth filter as preprocessing. Data of each trial are captured from the time segment starting from 0.5 to 2.5 s after the beginning of each trial (as done by the winner of BCI competition IV, data set IIa). The first 0.5 s is excluded as reaction time on visual clue. Feature vector extracted from each trial is used as one sample in training or testing. For fair comparison, same setting of filtering and time window is adopted for all methods.

For the feature selection (FS) method, the first d features are selected while the proposed FWR regularizes features controlled by m . Experiments are conducted for values of d and m from 2 to 16. The parameters δ for RCSP, τ for CSSP and σ for FWR are all selected by 5-fold cross-validation on training data. As mentioned earlier, σ should be around $m - 0.5$ so its value is limited between $m - 1.5$ and $m + 0.5$. The minimum Mahalanobis distance classifier is used in all experiments.

Training and testing of 6 methods, CSP, CSSP, RCSP, using FS and FWR, are conducted for 12 subjects and 6 tasks for each subject. A detailed explanation of each method can be found in section 3.2. In total 432 training and testing, experiments are performed. In each of the 432 experiments, classification results using parameter values (d and m) are recorded.

At first, we demonstrate figs. 4.7 to 4.12 which show classification accuracy for different subjects and tasks from Data set IIa. Figures 4.13 to 4.18 show classification accuracy for different subjects and tasks from Data set IIIa. Parameter value was chosen according to the best average accuracy over all subjects of each task and it can be found in Table 4.5 and Table 4.7 for Data set IIa and Data set IIIa, respectively.

Later, to see a whole picture, all results of the 432 experiments are summarized into the average accuracy over all tasks of each subject using the same parameter value, and the average accuracy over all subjects for each task using the same parameter value. The best of such accuracies among all tested parameters are recorded in Table 4.1 for different subjects from Data set IIa, in Table 4.3 for different subjects from Data set IIIa, in Table 4.5 for different tasks in Data set IIa, and in Table 4.7 for different tasks in Data set IIa.

Additionally to compare the overall classification performance at all different decision thresholds, we also present the average areas under ROC-curve (AUC) in Table 4.2 for different subjects from Data set IIa, in Table 4.4 for different subjects

from Data set IIIa, in Table 4.6 for different tasks in Data set IIa, and in Table 4.8 for different tasks in Data set IIa. Tables with AUC correspond to the tables with classification accuracy.

4.2.3 Results and discussion

For traditional feature selection (FS) methods, figs. 4.7 to 4.12 from Data set IIa show that CSSP outperform CSP in 31 out of 54 cases (9 subjects x 6 tasks) and RCSP in 26 out of 54 cases although they are extensions of CSP. Comparing the two extensions of CSP, CSSP is better than RCSP in 32 out of 54 cases while RCSP is better than CSSP in 19 out of 54 cases. Among the three feature selection approaches, CSSP performs the best in 24 out of 54 cases, RCSP performs the best in 13 out of 54 cases and the basic CSP performs the best in 9 out of 54 cases. Therefore, there is no approach that can consistently perform better than others in all cases.

To assess the performance of the proposed FWR algorithm, we apply it to the standard CSP and its two typical extensions in two different directions, CSSP utilizing the temporal information and RCSP regularizing the CSP. In total of 162 cases (9 subjects x 6 tasks x 3 FWR) from Data set IIa, the proposed FWR method outperforms FS in 85 cases and shows the same performance in 28 cases. It underperforms FS in 49 out of 162 cases. The average classification results below show that FWR method enhance overall classification.

Figures 4.13 to 4.18 from Data set IIIa show that CSSP outperform CSP in 9 out of 18 cases (3 subjects x 6 tasks) and RCSP in 6 out of 18 cases. Comparing the two extensions of CSP, CSSP is better than RCSP in 7 out of 18 cases while RCSP is better than CSSP in 6 out of 18 cases. Among the three feature selection approaches, CSSP performs the best in 5 out of 18 cases, RCSP performs the best in 4 out of 18 cases and the basic CSP performs the best in 2 out of 18 cases. Many results are the same for Database IIIa.

Figure 4.7: Comparison of classification rate for task 1 from BCI competition III, Dataset IIa

Figure 4.8: Comparison of classification rate for task 2 from BCI competition III, Dataset IIa

Figure 4.9: Comparison of classification rate for task 3 from BCI competition III, Dataset IIa

Figure 4.10: Comparison of classification rate for task 4 from BCI competition III, Dataset IIa

Figure 4.11: Comparison of classification rate for task 5 from BCI competition III, Dataset IIa

Figure 4.12: Comparison of classification rate for task 6 from BCI competition III, Dataset IIa

In total of 54 cases (3 subjects x 6 tasks x 3 FWR) from Data set IIIa, the proposed FWR method outperforms FS in 22 cases and shows the same performance in 14 cases. It underperforms FS in 18 out of 54 cases.

Figure 4.13: Comparison of classification rate for task 1 from BCI competition III, Dataset IIIa

Figure 4.14: Comparison of classification rate for task 2 from BCI competition III, Dataset IIIa

Figure 4.15: Comparison of classification rate for task 3 from BCI competition III, Dataset IIIa

Figure 4.16: Comparison of classification rate for task 4 from BCI competition III, Dataset IIIa

Figure 4.17: Comparison of classification rate for task 5 from BCI competition III, Dataset IIIa

Figure 4.18: Comparison of classification rate for task 6 from BCI competition III, Dataset IIIa

For FS methods, tables 4.1, 4.3, 4.5 and 4.7 show that CSSP outperform CSP in 22 out of 24 cases (9 subjects + 3 subjects + 6 tasks + 6 tasks) and RCSP in 15 out of 24 cases. Comparing the two extensions of CSP, CSSP is better than RCSP in 21

out of 24 cases while RCSP is better than CSSP in 2 out of 24 cases. Among the three feature selection approaches, CSSP performs the best in 20 out of 24 cases, RCSP performs the best in 2 out of 24 cases and the basic CSP performs the best in 1 out of 24 cases. This shows the difficulty of EEG-based motor imagery BCI. It is very difficult to develop an approach that can consistently perform better than others in all cases.

Table 4.1: Comparison of the best average performance for BCI competition IV, Data set IIa

	A01	A02	A03	A04	A05	A06	A07	A08	A09
CSP-FS	92.94	68.06	90.63	81.25	68.87	67.82	92.36	88.43	90.86
(d)	(10)	(10)	(4)	(2)	(4)	(2)	(8)	(4)	(4)
CSP-FWR	93.75	70.6	90.97	82.18	69.1	68.29	91.78	90.28	92.59
(m)	(13)	(7)	(4)	(2)	(3)	(2)	(7)	(3)	(15)
CSSP-FS	95.25	69.44	93.98	84.26	72.34	70.95	92.82	89	90.63
(d)	(10)	(14)	(6)	(4)	(6)	(6)	(16)	(6)	(16)
CSSP-FWR	95.49	72.11	93.98	84.84	72.69	71.18	92.82	91.2	91.2
(m)	(7)	(2)	(3)	(2)	(3)	(5)	(16)	(3)	(2)
RCSP-FS	93.52	68.63	91.67	80.56	66.09	67.71	92.36	89.7	92.13
(d)	(10)	(10)	(6)	(2)	(4)	(8)	(10)	(4)	(2)
RCSP-FWR	93.63	69.91	91.2	81.37	69.21	68.52	92.36	90.28	92.71
(m)	(9)	(4)	(6)	(2)	(4)	(2)	(6)	(5)	(14)

Table 4.2: Comparison of the average areas under ROC-curve (AUC) for BCI competition IV, Data set IIa

	A01	A02	A03	A04	A05	A06	A07	A08	A09
CSP-FS	0.035	0.213	0.032	0.103	0.242	0.263	0.035	0.025	0.020
CSP-FWR	0.031	0.198	0.031	0.100	0.237	0.246	0.037	0.020	0.014
CSSP-FS	0.026	0.192	0.013	0.077	0.201	0.218	0.030	0.021	0.023
CSSP-FWR	0.024	0.195	0.013	0.079	0.204	0.207	0.026	0.021	0.022
RCSP-FS	0.033	0.233	0.027	0.106	0.254	0.256	0.035	0.027	0.023
RCSP-FWR	0.030	0.190	0.028	0.104	0.240	0.239	0.038	0.023	0.016

Table 4.3: Comparison of the best average performance for BCI competition III, Dataset IIIa

	k3b	k6b	l1b
CSP-FS	96.85	76.11	91.39
(d)	(6)	(10)	(4)
CSP-FWR	96.85	76.94	91.94
(m)	(6)	(6)	(2)
CSSP-FS	97.41	81.94	92.78
(d)	(4)	(6)	(4)
CSSP-FWR	97.41	83.89	94.44
(m)	(5)	(4)	(3)
RCSP-FS	95.93	79.72	91.39
(d)	(6)	(12)	(4)
RCSP-FWR	96.85	81.39	91.94
(m)	(6)	(2)	(3)

Table 4.4: Comparison of the average areas under ROC-curve (AUC) for BCI competition III, Data set IIIa

	k3b	k6b	l1b
CSP-FS	0.005	0.161	0.029
CSP-FWR	0.003	0.147	0.021
CSSP-FS	0.003	0.079	0.032
CSSP-FWR	0.005	0.083	0.020
RCSP-FS	0.006	0.109	0.033
RCSP-FWR	0.003	0.106	0.030

Table 4.5: Comparison of the best average performance for BCI competition IV, Data set IIa

	task 1	task 2	task 3	task 4	task 5	task 6
CSP-FS	79.71	84.03	85.96	83.64	84.18	74.77
(d)	(4)	(4)	(8)	(2)	(6)	(4)
CSP-FWR	80.71	85.26	87.81	84.65	85.26	75.31
(m)	(3)	(2)	(13)	(3)	(11)	(3)
CSSP-FS	80.48	85.26	87.27	85.8	86.03	77.7
(d)	(8)	(10)	(6)	(6)	(4)	(6)
CSSP-FWR	81.94	87.19	87.35	86.34	86.34	79.32
(m)	(4)	(2)	(13)	(3)	(3)	(3)
RCSP-FS	78.86	84.1	86.19	85.19	85.03	74.38
(d)	(6)	(4)	(10)	(6)	(4)	(4)
RCSP-FWR	79.01	85.03	87.11	84.26	84.72	76.16
(m)	(4)	(2)	(13)	(5)	(4)	(3)

Table 4.6: Comparison of the average areas under ROC-curve (AUC) for BCI competition IV, Data set IIa

	task 1	task 2	task 3	task 4	task 5	task 6
CSP-FS	0.138	0.074	0.068	0.091	0.095	0.158
CSP-FWR	0.125	0.067	0.068	0.084	0.088	0.159
CSSP-FS	0.129	0.059	0.066	0.079	0.074	0.126
CSSP-FWR	0.123	0.063	0.063	0.074	0.076	0.124
RCSP-FS	0.149	0.094	0.083	0.092	0.091	0.169
RCSP-FWR	0.135	0.068	0.079	0.090	0.100	0.154

Table 4.7: Comparison of the best average performance for BCI competition III, Dataset IIIa

	task 1	task 2	task 3	task 4	task 5	task 6
CSP-FS	87.96	91.85	94.44	83.89	91.85	86.85
(d)	(10)	(6)	(2)	(4)	(4)	(6)
CSP-FWR	86.67	90.56	93.89	84.63	92.04	86.85
(m)	(6)	(5)	(2)	(2)	(7)	(2)
CSSP-FS	90.93	88.7	95.56	88.89	94.07	89.44
(d)	(12)	(6)	(4)	(6)	(10)	(2)
CSSP-FWR	91.67	90.19	95	91.85	95	90.19
(m)	(5)	(4)	(2)	(4)	(3)	(3)
RCSP-FS	89.07	88.52	95.37	88.89	92.41	87.96
(d)	(12)	(14)	(4)	(10)	(2)	(8)
RCSP-FWR	89.26	88.7	95.56	89.63	93.52	89.63
(m)	(8)	(2)	(2)	(10)	(12)	(6)

Table 4.8: Comparison of the average areas under ROC-curve (AUC) for BCI competition III, Data set IIIa

	task 1	task 2	task 3	task 4	task 5	task 6
CSP-FS	0.084	0.043	0.012	0.101	0.042	0.067
CSP-FWR	0.081	0.046	0.016	0.097	0.047	0.045
CSSP-FS	0.065	0.046	0.014	0.027	0.027	0.040
CSSP-FWR	0.061	0.044	0.012	0.032	0.020	0.040
RCSP-FS	0.047	0.064	0.007	0.034	0.045	0.066
RCSP-FWR	0.057	0.049	0.007	0.040	0.026	0.064

For the comparison between the conventional FS method and the proposed FWR method, the better one is highlighted in bold in tables 4.1, 4.3, 4.5 and 4.7. In total of 72 cases (3 FWR x (9 subjects + 3 subjects + 6 tasks + 6 tasks)), the proposed

FWR method outperforms FS in 58 cases and shows the same performance in 6 cases. It underperforms FS only in 8 out of 72 cases.

The results on the two datasets clearly suggest that the proposed FWR visibly enhances the performance of CSP based approaches in motor imagery BCI. The benefit of applying the proposed FWR method in the CSP based approaches is evident in tables 4.1, 4.3, 4.5 and 4.7. Tables 4.2, 4.4, 4.6 and 4.8 show that AUCs reflect the relative performances of different methods very similar to those of classification accuracies.

4.3 Summary

This chapter addresses problems of extracting discriminant features from the EEG signal for motor imagery BCI based on a set of training samples. To alleviate the problems of overfitting or poor generalization and hence boost the classification accuracy, the CSP based approaches in motor imagery BCI select some CSP features and remove the others. However, this will result in the loss of discriminative information.

This chapter proposes a novel approach that performs feature weighting and regularization (FWR) to replace the feature selection in CSP based approaches. FWR weights the features extracted by all spatial filters and then regularizes the eigen-spectrum of the within-scatter matrix of the weighted feature vector for reliable classification. The main advantage of the proposed method is that it can be applied in all CSP based approaches in MI-BCI.

This chapter evaluates the proposed method applying to CSP and its two extensions in two different directions, CSSP and RCSP on EEG data from 12 subjects with 6 tasks from BCI competition. Experimental results clearly demonstrate that the proposed feature weighting and regularization method yields superior classification accuracy compared to the existing feature selection methods.

Chapter 5

Time-Frequency Common Spatial Patterns

In this chapter, we propose a novel approach Time-Frequency Common Spatial Pattern (TFCSP) that extracts the complex common spatial patterns on time and frequency decomposed EEG signals to find the robust and discriminative features for classification in motor imagery brain-computer interface. In the proposed method, EEG time signal is decomposed into reaction, action, and completion stages of motor imagery and the EEG time signal of each stage is further decomposed into frequency components. We show that frequency decomposition not only leads to sharper spatial filters but also generates rich informative features. The proposed algorithm yields higher classification accuracy than other state-of-the-art methods on multiple BCI competition datasets.

5.1 Proposed method

5.1.1 Time decomposition of EEG signal

Most of the existing methods use time window of 0.5-2.5 seconds out of the full signal length of 0-4 seconds of BCI competition trials. This time window was used by the winner in BCI competition IV, Dataset IIa and then applied in many other methods. It is clear from the CSP presented in section 3.2.1 that the CSP features extract the spatial information of the EEG signal that carries the information of the different motor imagery. The CSP features do not carry any temporal information as the EEG signals are averaged over the time to compute the variance and covariance.

Figure 5.1: Time stages of the proposed TFCSP method (below) - reaction, action and completion stages by 3 overlapped time windows, 0-2 seconds, 1-3 seconds, and 2-4 seconds compared to usual time windows (above) 0.5-2.5 seconds.

We believe that different motor imagery may show different temporal pattern, though it may appear roughly or coarsely. To capture the coarse temporal pattern of different motor imagery, we propose to decompose the whole motor imagery

movement into reaction, action, and completion stages by 3 overlapped time windows, for example, the full length of 0-4 seconds signal into 3 signals of 0-2 seconds, 1-3 seconds, and 2-4 seconds, respectively. Let a matrix $\mathbf{X}$, $\mathbf{X} \in \mathbb{R}^{ch \times T}$ captures a trial of band-pass filtered EEG signals that characterizes one imaginary movement, where ch is the number of channels of EEG signals and T the signal length of a trial. We decompose the whole motor imagery movement into reaction, action, and completion stages by:

$$\mathbf{X}_p = \mathbf{X}(:, pT/4 + 1 : pT/4 + T/2), p = 0, 1, 2. \quad (5.1)$$

So we have $\mathbf{X}_p \in \mathbb{R}^{ch \times t}$, $t = T/2$ is the signal length of reaction, action or completion stages.

This is similar to the decomposition of signal into imagination preparation stage, imagination stage, and postimagination stage [129], but using longer and overlapped windows. The longer and overlapped windows make each of them capture part of the imagination stage since we are in fact not focusing on preparation and post imagination, but rather on beginning, middle, and ending of the imagery action. Every subject has its own way and speed of motor imagery so that it is a challenging task to check when imagination has started, how long the subject is imagining the action itself, as well as when it will end. Nevertheless, it is essential to capture some temporal information of the motor imagery to discriminate different motor imagery actions. Therefore, we use longer window length of 2 seconds for each stage to consider them as reaction, action, and completion stages of the motor imagery.

5.1.2 Frequency decomposition of EEG signal

It is clear from the CSP presented in section 3.2.1 that the CSP features extract the spatial information of the EEG signal that carries the information of the different

motor imagery. The CSP features do not carry any frequency information. We believe that the different motor imagery may show different frequency pattern. This is evidenced by several approaches [83–86, 89] that employ a filter bank to separate EEG signal into multiple narrow band signals and apply CSP to each of the filtered narrow band signals. In this work, instead of generating multiple narrow band EEG time signals, we propose to transform the EEG time signal into frequency domain and extract CSP features in various groups of frequency components.

After dividing the whole trial $\mathbf{X}$ into reaction, action, and completion stages $\mathbf{X}_p$, $p = 0, 1, 2$, using 3 overlapped windows, we apply discrete Fourier transform (DFT) to each stage to transform the time signal into frequency components. While most of the methods use bandwidth filter bank to generate n narrow-band EEG time-signals and hence increase the data size n times, we decompose the signal into frequency components without the increase of the data size.

$$\hat{\mathbf{X}}_p = \text{DFT}(\mathbf{X}_p) = (\mathbf{W}_F \mathbf{X}_p^T)^T, \quad (5.2)$$

where $\mathbf{W}_F \in \mathbb{C}^{t \times t}$ is the DFT coefficient matrix given by::

$$\mathbf{W}_F = \{w_{km}\} = e^{\frac{-j2\pi km}{t}}, \quad (5.3)$$

where $k, m = 0, 1, 2, \dots, t - 1$. To capture the frequency pattern of different motor imagery, we propose to partition the frequency components in range of 2-40Hz into 10 cells using 50% overlapped windows. Data in each cell are captured in a matrix denoted by $\hat{\mathbf{X}}_p^q$, $p = 0, 1, 2$, $q = 1, \dots, 10$, $\hat{\mathbf{X}}_p^q \in \mathbb{C}^{ch \times b}$. Note that $b \ll T$, e.g. $b = 14$ where $T = 1000$ for a 4 seconds EEG signal with 250Hz sampling frequency.

5.1.3 Complex common spatial patterns on decomposed frequency components

Common spatial pattern (CSP) [40] is one of the most popular algorithms in MI-BCI to extract features. It finds spatial filters that maximize the variance of one class and minimize the variance of the other class. There are different versions of CSP. Our proposed approach performs the following described complex CSP on each of our decomposed cells from EEG signal. The complex covariance matrix of a cell $\hat{\mathbf{X}}_p^q \in \mathbb{C}^{ch \times b}$ is computed by

$$\Sigma_p^q = \hat{\mathbf{X}}_p^q \hat{\mathbf{X}}_p^{qH}, \quad (5.4)$$

where H denotes conjugate transpose of a complex matrix. If the training set contains N_j trials for class Ω_j , the average normalized covariance matrix of class Ω_j is computed by

$$\mathbf{C}_{pj}^q = \frac{1}{N_j} \sum_{\Sigma_p^q \in \Omega_j} \frac{\Sigma_p^q}{tr(\Sigma_p^q)}. \quad (5.5)$$

We can find a spatial filter $\mathbf{w}_p^q \in \mathbb{C}^{ch \times 1}$ that maximizes the variance of Ω_1 and minimizes it of Ω_2 by solving

$$\max J(\mathbf{w}_p^q) = \frac{\mathbf{w}_p^{qH} \mathbf{C}_{p1}^q \mathbf{w}_p^q}{\mathbf{w}_p^{qH} (\mathbf{C}_{p1}^q + \mathbf{C}_{p2}^q) \mathbf{w}_p^q}. \quad (5.6)$$

Solution of eq. (5.6) is obtained by solving the generalized eigenvalue problem:

$$\mathbf{C}_{p1}^q \mathbf{w}_p^q = \lambda_p^q (\mathbf{C}_{p1}^q + \mathbf{C}_{p2}^q) \mathbf{w}_p^q, \quad (5.7)$$

where λ_p^q and $\mathbf{w}_p^q$ are the generalized eigenvalue and eigenvector. Applying the above complex CSP in each of the 30 cells separately, we extract CSP features from each of the 30 cells. A normalized feature vector of each cell is computed as:

$$\mathbf{f}_p^q = \log(\text{diag}(\mathbf{W}_p^{qH} \mathbf{\Sigma}_p^q \mathbf{W}_p^q / \text{tr}(\mathbf{\Sigma}_p^q))), \quad (5.8)$$

5.1.4 Removing indiscriminative and unreliable features

Now we have a large number of features, i.e., 30 groups of *ch* CSP features. This work does not try to select the most discriminative features because the feature selection based on the training data may lead to the problems of overfitting the training data and losing some discriminative information. Instead, this work first removes some extremely unreliable and indiscriminative cells (no more than half) and then using feature weighting and regularization approach circumvents the problem of feature selection.

Extremely low energy of the frequency components in a cell shows unreliable information in the cell. To identify unreliable cell, we compute the total energy of a cell over all training data Ω as:

$$En_p^q = \sum_{\mathbf{\Sigma}_p^q \in \Omega} \text{tr}(\mathbf{\Sigma}_p^q). \quad (5.9)$$

Then we remove the 5 cells that have the lowest total energy that are deemed as containing unreliable information. EEG signals contain rich information, not just the information of the motor imagery. Eigenvalue λ of eq. (3.4) shows the power of a spatial filter $\mathbf{w}$ to discriminate different motor imagery actions. Since both the largest and smallest eigenvalues of eq. (3.4) correspond to the largest differences of variances of the two classes and $0 \leq \lambda \leq 1$, the value of $|\lambda - 0.5|$ is proportional to the discriminative power of a spatial filter. The sum of the value $|\lambda - 0.5|$ over

all spatial filters of a cell:

$$Ev_p^q = \sum (|\lambda_p^q - 0.5|) \quad (5.10)$$

shows the discriminative power of the cell. We remove the 10 cells that have the lowest Ev_p^q that are deemed as containing ineffective information to discriminate different motor imagery actions.

To avoid using additional cross-validation during training phase to find optimal number of cells to exclude, we predefined the number as no more than half. After removal of 5 unreliable cells and 10 non-informative cells, we have remaining nc cells, $15 \leq nc \leq 20$ (some cells could be both unreliable and non-informative). After that in first round of experiments we apply classification in low dimensionality subspace using Principal Component Analysis (PCA) approach to remove unreliable dimensions and therefore enhance classification.

In the second round of experiments, we try not to select the most effective or discriminative features because the feature selection based on the training data may lead to the problems of overfitting the training data and loss of some discriminative information. Instead, after removing some extremely indiscriminative and unreliable features in terms of cells (no more than half of the cells), we then use feature weighting and regularization in order to solve the problem of feature selection. Both methods and experimental results are described below.

5.1.5 Classification in low dimensionality subspace

Dimensionality reduction has two objectives. One is to reduce the computational complexity of the subsequent classification with the minimum loss of information needed for classification and the other is to circumvent the prediction problem of the subsequent classification and hence enhance its accuracy and robustness.

Although the ultimate objective of a classification is to extract the most discriminative information, it is the most discriminative information about the whole data population, not on a specific training set. A classifier is trained to capture the most discriminative information on the training samples. If some statistics estimated on the training data deviate from those of the data population, misclassification rate on the novel data increases.

Therefore, to boost the classification accuracy, the dimensionality reduction by machine learning should be targeted at removing the dimensions unreliable (harmful) for the classification or regularizing (correcting) the unreliable statistics in these dimensions. We apply classification in low dimensionality subspace using PCA approach to remove unreliable dimensions and therefore enhance classification [119].

We first concatenate d CSP features from remaining cells together and then sort them in descending order of their discriminative values $|\lambda_p^q - 0.5|$ to generate our feature vector $\mathbf{f} = [f_1, \dots, f_k, \dots, f_{nd}]^T$, $nd = nc \times d$.

Then we find covariance matrix of the total training data or total scatter matrix:

$$\mathbf{S}_p = \sum_{\mathbf{f} \in \Omega_{1,2}} (\mathbf{f} - \boldsymbol{\mu}^{1,2})(\mathbf{f} - \boldsymbol{\mu}^{1,2})^T, \quad (5.11)$$

where $\Omega_{1,2} = \Omega_1 \cup \Omega_2$, and $\boldsymbol{\mu}^{1,2}$ is mean of both classes from training set.

PCA applies eigendecomposition on $\mathbf{S}_p$:

$$\mathbf{S}_p \mathbf{w} = \lambda \mathbf{S}_p \mathbf{w}, \quad (5.12)$$

and keeps p eigenvectors $\mathbf{w} \in \mathbb{C}^{nd \times 1}$, corresponding to p largest PCA eigenvalues. Total matrix of PCA eigenvectors becomes $\mathbf{W} \in \mathbb{C}^{nd \times p}$. An nd -dimensional feature vector $\mathbf{f}$ is transformed to an p dimensional feature vector $\bar{\mathbf{f}} \in \mathbb{C}^{p \times 1}$:

$$\bar{\mathbf{f}} = \mathbf{W}^T \mathbf{f}, \quad (5.13)$$

p is values to chose.

For classification method, we use minimum Mahalanobis distance classifier. The Mahalanobis distance d_j of a feature vector $\bar{\mathbf{f}}$ to the class mean $\boldsymbol{\mu}_j$ of Ω_j is evaluated by

$$d_j = (\bar{\mathbf{f}} - \boldsymbol{\mu}_j)^T \mathbf{S}_w^{-1} (\bar{\mathbf{f}} - \boldsymbol{\mu}_j) \quad (5.14)$$

where

$$\mathbf{S}_w = \sum_{\bar{\mathbf{f}} \in \Omega_1} (\bar{\mathbf{f}} - \boldsymbol{\mu}_1)(\bar{\mathbf{f}} - \boldsymbol{\mu}_1)^T + \sum_{\bar{\mathbf{f}} \in \Omega_2} (\bar{\mathbf{f}} - \boldsymbol{\mu}_2)(\bar{\mathbf{f}} - \boldsymbol{\mu}_2)^T. \quad (5.15)$$

Figure 5.2 illustrates the overall framework of the proposed TFCSP method with PCA.

Figure 5.2: The overall framework of the proposed TFCSP method. EEG samples are first decomposed into 3 stages: reaction, action and completion stages of motor imagery action. Each of the time stages is transformed into the frequency domain and further segmented into 10 groups of frequency components. Complex CSP is then applied to extract d features from each of the time-frequency cells. After that, cells with the lowest energy and discriminative power are excluded. PCA is performed on all remaining CSP features for dimensionality reduction followed by the minimum Mahalanobis distance classifier.

5.1.6 Classification with feature weighting and regularization

All CSP based approaches select the features produced by the spatial filters. The total number of possible features is the number of EEG channels, ch . However, literatures show that the best classification accuracy is achieved at very few number of features, such as 2 to 8 in most cases compared to, for example, $ch = 60$.

The loss of discriminating information by discarding most CSP features may adversely affect the classification accuracy. To alleviate this problem, we use feature weighting and regularization (FWR) [120, 130] method that utilizes all CSP features for classification.

We first concatenate all CSP features of the remaining cells together and then sort them in descending order of their discriminative values $|\lambda_p^q - 0.5|$ to generate our feature vector $\mathbf{f} = [f_1, \dots, f_k, \dots, f_{nf}]^T, nf = nc \times ch$. Then we weight the features by Gaussian function as:

$$\gamma_k = e^{-\frac{(k-1)^2}{2\sigma^2}}, \quad (5.16)$$

$$f_k^\gamma = \gamma_k f_k, \quad (5.17)$$

where σ is a parameter, which can be chosen so that the Gaussian function decays the fastest between m and $m + 1$. This leads to $\sigma = m - 0.5$, where m is the Eigenfeature Regularization and Extraction (ERE) regularization parameter shown later. The weighted feature vector is $\mathbf{f}_\gamma = [f_1^\gamma, \dots, f_k^\gamma, \dots, f_{ch}^\gamma]^T$.

For classification method, we use minimum Mahalanobis distance classifier. The Mahalanobis distance d_j of a feature vector $\mathbf{f}_\gamma$ to the class mean $\boldsymbol{\mu}_j$ of Ω_j is evaluated by:

$$d_j = (\mathbf{f}_\gamma - \boldsymbol{\mu}_j)^T \mathbf{S}_w^{-1} (\mathbf{f}_\gamma - \boldsymbol{\mu}_j). \quad (5.18)$$

Mahalanobis distance can be evaluated in the eigenspace in scalar form as:

$$d_j = \sum_{i=1}^{ch} \frac{(z_i - \bar{z}_{ji})^2}{\lambda_i}, \quad (5.19)$$

where

$$z_i = \boldsymbol{\phi}_i^T \mathbf{f}_\gamma, \quad (5.20)$$

$$\bar{z}_{ji} = \phi_i^T \mu_j, \quad (5.21)$$

ϕ_i and λ_i are the i th eigenvector and eigenvalue of $\mathbf{S}_w$:

$$\mathbf{S}_w = \sum_{\mathbf{f}_\gamma \in \Omega_1} (\mathbf{f}_\gamma - \mu_1)(\mathbf{f}_\gamma - \mu_1)^T + \sum_{\mathbf{f}_\gamma \in \Omega_2} (\mathbf{f}_\gamma - \mu_2)(\mathbf{f}_\gamma - \mu_2)^T. \quad (5.22)$$

where ϕ_i and λ_i are found by solving the generalized eigenvalue problem:

$$\mathbf{S}_w \phi = \lambda \mathbf{S}_w \phi. \quad (5.23)$$

The decay of the eigenvalues should be slowed down to compensate the effect of noise and the finite number of training samples. This can be done by replacing the eigenspectrum with the proposed model:

$$\tilde{\lambda}_i = \begin{cases} \lambda_i + c, & i < m \\ \frac{\alpha}{i+\beta} + c, & m \leq i \leq ch \end{cases} \quad (5.24)$$

where

$$\alpha = \frac{\lambda_1 \lambda_m (m-1)}{\lambda_1 - \lambda_m}, \quad (5.25)$$

$$\beta = \frac{m \lambda_m - \lambda_1}{\lambda_1 - \lambda_m}. \quad (5.26)$$

where α and β are two constants determined by the reliable eigenvalues. The above eigenvalues λ_i of $\mathbf{S}_w$ are sorted in descending order before the regularization. c and m are two parameters to choose for regularization. This work chooses $c = \lambda_{ch/2}$ to reduce the free parameter.

The main steps of the proposed TFCSP with FWR is listed as following:

1. Utilize time paradigm to decompose EEG time signal into 3 stages: reaction, action and completion stages.

2. Apply DFT to decompose EEG time signal of each time stage into 10 groups of frequency components.
3. Perform complex CSP on frequency components of each of the 30 time-frequency cells.
4. Remove 5 cells of the lowest total energy and 10 cells of the lowest discriminative power from the 30 cells.
5. Sort all CSP features and their spatial filters of all remaining cells in descending order of the CSP eigenvalues (5.7), concatenate them into feature vector and weight it by the Gaussian function (5.16), (5.17).
6. Compute the within-class scatter matrix of the feature vector, its eigenvector and eigenvalues. Regularize the eigenvalues of the within-class scatter matrix by (5.24), (5.25), (5.26).
7. Apply the minimum Mahalanobis distance classifier using the regularized eigenvalues (5.18).

Figure 5.3 illustrates the overall framework of the proposed TFCSP method with FWR. At the recognition phase, a query EEG signal is first processed in the same way as the first 2 steps of the above training procedure. Then CSP features of the query EEG signal are extracted by the spatial filters of the reliable cells obtained in the training stage and weighted by Gaussian function (5.16), (5.17). Finally, feature vector of the query EEG signal is projected to the eigenvectors of the within-class scatter matrix and normalized by the regularized eigenvalues. The minimum distance classifier is applied to the weighted and regularized eigenfeatures to generate the class label of the query EEG signal.

Figure 5.3: The overall framework of the proposed TFCSP method. EEG samples are first decomposed into 3 stages: reaction, action and completion stages of motor imagery action. Each of the time stages is transformed into the frequency domain and further segmented into 10 groups of frequency components. Complex CSP is then applied to extract features from each of the time-frequency cells. After that, cells with the lowest energy and discriminative power are excluded. Feature weighting and regularization is performed on all remaining CSP features followed by the minimum Mahalanobis distance classifier.

5.2 Experiments

5.2.1 EEG datasets

In the experiments, EEG data from three widely used motor imagery BCI datasets are used:

1) BCI competition IV, Dataset IIa [127] contains EEG data of 22 channels recorded from 9 healthy subjects(A01-A09). The data set contains 144 trials per class with time point markers provided. The EEG signal was sampled with 250Hz and later bandpass filtered between 0.5Hz and 100Hz (with 50Hz notch filter enabled). Each task was performed for 4s. More detailed explanation is provided in subsection 4.2.1

2) BCI Competition IV, Dataset IIb [131] contains EEG signals recorded from 9 subjects at electrodes C3, Cz, and C4. The location of the electrodes can be seen in Figure 5.4. The data set contains 80 trials for each class with time point markers provided. The sampling rate is 250Hz. We only use the third training sessions of the dataset, i.e., "B0103T," "B0203T," . . . , "B0903T.". In each trial, the subject was indicated by a visual cue to perform MI task for 4.5s.

Figure 5.4: Position of EEG electrodes (blue) in Dataset IIb [131].

At the beginning of each trial ($t = 0s$) in third training sessions which we use in our experiments, the smiley was shown on the screen. At $t = 2s$, a short warning

sound was given. The cue was demonstrated from 3s to 7.5s. Depending on the cue, the subjects were asked to move the smiley towards the left or right side by left or right hand motor imagery. If it was successful and motor imagery was recognized correctly smiley turned green, otherwise, it was red.. At $t = 7.5s$ the screen went blank and a random interval between 1s and 2s was added to the trial. Figure 5.5 illustrates the timing scheme of the dataset. We use time point $t = 3s$ as the beginning of the trial.

Figure 5.5: Timing scheme of the paradigm in Dataset IIb [131].

3) BCI competition III, Dataset IIIa [128] contains EEG signals of 60 channels recorded from 3 subjects. The data set contains 90, 60 and 60 trials per class with time point markers for subjects k3b, k6b, and l1b, respectively. The EEG signal was sampled with 250Hz and filtered between 1Hz and 50Hz with notch filter enabled. Each task was performed for 4s. More detailed explanation is given in subsection 4.2.1

The first and the third datasets contain four motor imagery actions: movements of left hand, right hand, foot, and tongue. Since CSP based approaches work on discriminating two classes, we pair the 4 MIs to construct 2 binary classification tasks: task 1 - left-hand versus right-hand and task 2 - foot versus tongue.

5.2.2 Preprocessing, parameter settings and visualization of proposed TFCSP

At first, the raw EEG signals are filtered between 1.5-40Hz with fifth-order Butterworth filter as preprocessing to attenuate noise and outliers such as blinking. Then the signal is divided into 3 equal overlapped time windows that represent reaction, action and completion stages of a motor imagery action.

Table 5.1: Decomposition parameters on different datasets

Dataset	Sampling frequency, Hz	Duration of 1 trial, seconds	Time stages, samples	Frequency components, Hz*(time stages samples/sampling frequency)
Dataset IIa	250	4	[1-500], [250-749], [500-999].	[4-17, 11-24, 18-31, 25-38, 32-45, 39-52, 46-59, 53-66, 60-73, 67-80]
Dataset IIb	250	4.5	[1-563], [282-844], [563-1125].	[4-19, 12-27, 20-35, 28-43, 36-51, 44-58, 52-66, 60-74, 67-82, 75-90]
Dataset IIIa	250	4	[1-500], [250-749], [500-999].	[4-17, 11-24, 18-31, 25-38, 32-45, 39-52, 46-59, 53-66, 60-73, 67-80]

Dataset IIa and Dataset IIIa have motor imagery actions conducted for 4s and sampling frequency of datasets is 250Hz which makes 1000 samples. Thus, the reaction, action and completion stages are composed of the discrete samples of [1-500], [250-749], [500-999], respectively. BCI competition IV, Dataset IIb have motor imagery actions conducted for 4.5s and the sampling frequency is 250Hz which makes 1125 samples. Thus, the reaction, action and completion stages are composed of the discrete samples of [1-563] [282-844] [563-1125], respectively. After FFT is applied to each time stage, we take 10 equal overlapped frequency components from FFT within 2-40Hz, which is [2-8.5, 5.5-12, 9-15.5, 12.5-19, 16-22.5, 19.5-26, 23-29.5, 26.5-33, 30-36.5, 33.5-40] Hz. For Dataset IIa and Dataset IIIa, the overlapped frequency components in cells consist of [4-17, 11-24, 18-31, 25-38, 32-45, 39-52, 46-59, 53-66, 60-73, 67-80] DFT samples, respectively. Due to

the different time duration of Dataset IIb, the overlapped frequency components in cells consist of [4-19, 12-27, 20-35, 28-43, 36-51, 44-58, 52-66, 60-74, 67-82, 75-90] DFT samples. All parameters for time and frequency decomposition of the EEG signal are recorded in Table 5.1.

After extracting CSP features from each cell, we exclude 10 cells with the smallest CSP eigenvalues and 5 cells with the smallest energies. Parameters m , d , $npca$ and σ are selected during 5-fold cross-validation on training data, where m and d are chosen within values from 2 to 40, σ is selected within values from $m - 1.75$ to $m + 0.5$ with 0.25 step and $npca$ is chosen within values from $2 \times d$ to $5 \times d$ with 0.5 step. The minimum Mahalanobis distance classifier is used in all experiments.

Figure 5.6: Visualization in topography map of the first 2 spatial filters of CSP without time and frequency decomposition of the data. Results are generated from training data of task 1 - left-hand versus right-hand, of subject A06 from Dataset IIa.

Figure 5.7, Figure 5.8, and Figure 5.10 demonstrate in topography map the first 2 or 3 spatial filters of CSP without time and frequency decomposition of the data. Figure 5.7, Figure 5.9, and Figure 5.11 show the first spatial filters of some time-frequency cells obtained by the proposed TFCSP method. They are generated from the training dataset of the task left-hand versus right-hand performed by of

subjects A03, A06, and A07, respectively. Unshaded boxes in the top of figures show cells excluded and the topography maps of cells in the colored rectangles are illustrated, respectively, in the bottom of the figure. Topography maps represent spatial filters where red color corresponds to the highest weights while blue to the smallest weights. It provides neurophysiological interpretations of the spatial filters. For each part of the human body, respective region in the motor and somatosensory area of the neurocortex is presented [59]. For example, brain activity for left hand motor imagery arises from the right part in the motor cortex area, while the right hand arises from the left part.

Figure 5.7, Figure 5.9, and Figure 5.11 show that most cells of high frequency have low energy and hence are excluded. The topographies of the cells show that most of the excluded time-frequency cells appear with large spatial filter weights in several unexpected locations or unsmoothing weight distribution from a neurophysiological point of view. The included cells by the proposed TFCSP method show in general smoother, spatial sharper and physiologically more relevant topography maps, with strong weights over the motor cortex areas.

Figure 5.7: Visualization in topography map of the first spatial filters of some time-frequency cells from the proposed TFCSP. Results are generated from training data of task 1 - left-hand versus right-hand, of subject *A06* from Dataset IIa. Unshaded boxes in the top of Figure show cells excluded and the topography maps of cells in the colored contours are illustrated, respectively, in the bottom of the figure. Most cells of high frequency have low energy and hence are excluded. Note that the first spatial filters of some cells represent left hand while those of some other cells may represent right hand.

Figure 5.8: Visualization in topography map of the first 3 spatial filters of CSP without time and frequency decomposition of the data. Results are generated from training data of task 1 - left-hand versus right-hand, of subject *A03* from Dataset IIa.

For example, in Figure 5.7 the third column of the topography maps at $16-22.5Hz$ illustrates that cells at $0-2s$ and $2-4s$ have no clear activation and hence are excluded while cell at $1-3s$ has higher weights in the motor cortex area and hence is included. In Figure 5.9 the third column of the topography maps at $16-22.5Hz$ illustrates that cells at $0-2s$ have activation in different parts and hence are excluded while cell at $1-3s$ and $2-4s$ has higher weights in the motor cortex area from left side and hence is included. Figure 5.11 shows differences in activation in second column of the topography maps at $16-22.5Hz$ topography maps are smooth and neurophysiologically conceivable and hence is included while in third column at $23-29Hz$ activation is hardly interpretable, therefore is excluded.

Figure 5.9: Visualization in topography map of the first spatial filters of some time-frequency cells from the proposed TFCSP. Results are generated from training data of task 1 - left-hand versus right-hand, of subject A03 from Dataset IIa. Unshaded boxes in the top of Figure show cells excluded and the topography maps of cells in the colored contours are illustrated, respectively, in the bottom of the figure. Most cells of high frequency have low energy and hence are excluded. Note that the first spatial filters of some cells represent left hand while those of some other cells may represent right hand.

Figure 5.10: Visualization in topography map of the first 2 spatial filters of CSP without time and frequency decomposition of the data. Results are generated from training data of task 1 - left-hand versus right-hand, of subject *A07* from Dataset IIa.

Since spatial filters are sorted by $|\lambda_p^q - 0.5|$, the first spatial filters of some cells represent left hand while those of some other cells may represent right hand. This can be seen in the figures as activation in right part of the motor cortex and in left part of the motor cortex, respectively. This supports that the proposed TFCSP method chooses cells and filters are neurophysiologically conceivable and thus interpretable. The smoother and physiologically more relevant topography maps of the included cells than those of CSP without time-frequency decomposition of the data (Figure 5.6, Figure 5.8, and Figure 5.10) show that CSP features in the narrow time-frequency cells are more discriminative than those on the whole EEG data. Moreover, the diverse patterns of the topography maps over different cells indicate that the proposed time-frequency decomposition extracts more relevant information than CSP on the whole EEG data.

Figure 5.11: Visualization in topography map of the first spatial filters of some time-frequency cells from the proposed TFCSP. Results are generated from training data of task 1 - left-hand versus right-hand, of subject *A07* from Dataset IIa. Unshaded boxes in the top of Figure show cells excluded and the topography maps of cells in the colored contours are illustrated, respectively, in the bottom of the figure. Most cells of high frequency have low energy and hence are excluded. Note that the first spatial filters of some cells represent left hand while those of some other cells may represent right hand.

5.3 Results and discussion

For most of the state-of-the-art methods presented in literature, their source codes are not available. Reproducing the codes may not be able to optimize the published methods and very often leads to lower performance than those presented in the original papers. For a fair comparison with the state-of-the-art methods, we evaluate our proposed method under the exact same experimental settings as those used in the original papers of the state-of-the-art methods. The performances of our proposed method under such experimental settings are then compared with the results of the the methods reported in the published papers.

TGCSP [102] and LRFCSP [91] are two recently published and high performing methods. They are also the two approaches that are most related to our proposed method. Classification performances of the two methods are better than most existing CSP-based methods over publicly available BCI competition datasets. Therefore, we first compare our proposed method with TGCSP and other 4 methods which are compared with TGCSP published in [102], under the same experimental setting as those in [102]. Then, we further compare our proposed method with LRFCSP and other 4 methods compared with LRFCSP published in [91], under the same experimental settings as those in [91].

Experiments for TSGSP [102] were performed with 5×5 -fold crossvalidation. We applied the same settings and the training data was randomized five times. After each randomization, the data was divided into five folds. Four of these folds were used for training and the remaining fold was used for analyzing the performance. This procedure results in five validation runs for each randomization, which leads to a total of $5 \times 5 = 25$ validation runs of training and testing. Other method-dependent parameter settings for different methods are given in [102].

Experiments for LRFCSP [91] were performed with 10×10 -fold crossvalidation for left vs right hand task from BCI competition IV, Data set IIb. We applied

the same settings and the training data was randomized ten times. After each randomization, the data was divided into ten folds. Nine of these folds were used for training and the remaining fold was used for analyzing the performance. This procedure results in ten validation runs for each randomization, which leads to a total of $10 \times 10 = 100$ validation runs of training and testing. Other method-dependent parameter settings for different methods are given in [91].

First, we demonstrate the results for our proposed method using PCA, and then later the results and discussion for our proposed method TFCSP with FWR.

5.3.1 Experimental results of TFCSP with PCA

Table 5.2, Table 5.3, Table 5.4 and Table 5.5 show the comparison of the average classification accuracy over 25 training and testing of the proposed method with TSGSP and other 4 competing methods in the literature reported in [102]. The last row of each table shows the average accuracy of the same motor imagery task (or for both motor imagery tasks in Table 5.10) over different subjects and the standard deviation. Table 5.6 shows the comparison of the classification rate of the proposed method with LRFCSP.

Table 5.2: Comparison of 5x5 crossvalidation performance for left vs right hand task from BCI competition IV, Data set IIa

	CSP	CSP MT	SFBCSP	SFBCSP MT	TSGSP	TF(PCA)
A01	76.9	80.6	82.8	84.1	87	93.50
A02	57.1	62.3	59.8	62.9	64.7	70.79
A03	91.6	92.3	92	92.9	93.8	98.50
A04	65.5	70.9	68	71.6	74.3	85.07
A05	54.8	55.8	81.3	86.9	90.4	94.93
A06	58.3	59.4	59.3	61.2	63.9	74.00
A07	73.4	74.7	89.1	89.8	91.4	99.00
A08	92.8	93.8	92.8	94.3	95.8	99.14
A09	67.8	80.1	70.2	80.9	81.3	89.79
Average	70.9±14.1	74.4±13.7	77.3±13.3	80.5±12.5	82.5±12.2	89.41±10.7

Table 5.3: Comparison of 5x5 cross-validation performance for foot vs tongue task from BCI competition IV, Data set IIa

	CSP	CSP MT	SFBCSP	SFBCSP MT	TSGSP	TF(PCA)
A01	65.4	68.1	79.3	82.6	85.0	87.6
A02	87.3	88.8	88.5	90.7	92.4	90.6
A03	79.4	81.7	84.9	86.1	89.7	90.8
A04	60.9	65.7	68.1	71.1	75.8	85.6
A05	62.4	64.0	64.8	71.3	73.9	77.9
A06	61.1	63.7	69.9	71.7	75.1	80.7
A07	77.3	79.0	84.5	86.6	88.3	93.4
A08	87.7	90.1	88.9	91.1	93.0	93.4
A09	70.6	76.4	75.4	80.9	82.5	94.1
Average	72.5±10.8	75.3±10.4	78.3±9.1	81.3±8.2	84±7.5	88.2±5.8

Table 5.4: Comparison of 5x5 cross-validation performance for left vs right hand task from BCI competition IV, Data set IIb

	CSP	CSP MT	SFBCSP	SFBCSP MT	TSGSP	TF(PCA)
B0103T	76.6	81.8	79.1	81.8	84.00	81.6
B0203T	56.3	58.7	59	60.3	62.60	62.6
B0303T	52.6	50.8	53.1	54	56.30	55.9
B0403T	98.9	98.9	98.9	99.1	99.40	98.5
B0503T	83.6	89.5	91.5	92.6	94.8	95.8
B0603T	67.1	68.5	81.3	82	83.8	83.6
B0703T	83.8	83.6	90.8	91.8	94.1	92.8
B0803T	86.5	88.3	88.9	91.1	93.3	92.0
B0903T	81.9	86.1	85.4	87.3	90.1	87.1
Average	76.4±15	78.5±15.8	80.9±15.3	82.2±15.3	84.3±15	83.3±14.8

Table 5.5: Comparison of 5x5 cross-validation performance from BCI competition III, Data set IIIa

	CSP	CSP MT	SFBCSP	SFBCSP MT	TSGSP	TF(PCA)
left hand vs right hand						
k3b	95.8	96	96.6	97.5	99.2	98.7
k6b	60	65.8	65.8	66.3	67.2	72.2
l1b	93.5	94.2	94.7	94.8	96.5	90.3
foot vs tongue						
k3b	84	86.8	88.9	94	96.4	94.6
k6b	78.4	80.3	83.8	85.2	87.8	96
l1b	73.2	74.7	78	81.5	84	84
Average	80.8±13.4	83±11.7	84.6±11.5	86.6±11.7	88.5±12	89.3±9.8

Table 5.6: Comparison of 10x10 cross-validation performance for left vs right hand task from BCI competition IV, Data set I Ib

	CSP-TSM	SBLFB	I-DFBCSP	LRCSP	LRFCSP	TF(PCA)
B0103T	77.50	78.25	80.75	75.63	82.50	81.1
B0203T	60.50	59.25	58.37	55.72	60.00	65.2
B0303T	50.94	49.32	56.00	54.38	62.78	57.0
B0403T	99.25	99.12	99.37	96.86	99.37	98.6
B0503T	82.82	92.04	90.58	88.75	87.50	95.4
B0603T	76.99	79.49	82.00	65.63	81.88	84.1
B0703T	86.19	92.50	88.87	83.13	90.63	93.7
B0803T	88.56	88.87	89.50	82.50	91.25	91.0
B0903T	80.25	80.62	83.75	85.63	83.13	86.9
Average	78.1±14.6	79.9±16.3	81±14.6	76.5±14.9	82.1±13	83.7±14.1

As can be seen from the results in these 5 tables, our proposed TFCSP method outperforms the approaches reported in [91, 102] except for database I Ib. Moreover, our proposed TFCSP with PCA also results in the smallest standard deviation for all the tables, indicating good stability and robustness of the proposed method. It proves that decomposing time of the signal trial into reaction, action and completion stages and using frequency decomposition lead to sharper spatial filters and generate rich informative features.

Since database I Ib contains only 3 EEG electrodes dimensionality reduction and feature selection which we used with PCA may affect classification. Therefore, we use FWR to avoid feature selection.

5.3.2 Experimental results of TFCSP with FWR

Table 5.7: Comparison of 5x5 crossvalidation performance for left vs right hand task from BCI competition IV, Data set IIa

	CSP	CSP MT	SFBCSP	SFBCSP MT	TSGSP	TFCSP
A01	76.9	80.6	82.8	84.1	87	96.6
A02	57.1	62.3	59.8	62.9	64.7	74.9
A03	91.6	92.3	92	92.9	93.8	98.2
A04	65.5	70.9	68	71.6	74.3	89.6
A05	54.8	55.8	81.3	86.9	90.4	95.5
A06	58.3	59.4	59.3	61.2	63.9	77.9
A07	73.4	74.7	89.1	89.8	91.4	99.3
A08	92.8	93.8	92.8	94.3	95.8	98.8
A09	67.8	80.1	70.2	80.9	81.3	90.9
Average	70.9±14.1	74.4±13.7	77.3±13.3	80.5±12.5	82.5±12.2	91.3±9.1

Table 5.8: Comparison of 5x5 cross-validation performance for foot vs tongue task from BCI competition IV, Data set IIa

	CSP	CSP MT	SFBCSP	SFBCSP MT	TSGSP	TFCSP
A01	65.4	68.1	79.3	82.6	85	88.2
A02	87.3	88.8	88.5	90.7	92.4	92.8
A03	79.4	81.7	84.9	86.1	89.7	95.4
A04	60.9	65.7	68.1	71.1	75.8	85.7
A05	62.4	64	64.8	71.3	73.9	80.9
A06	61.1	63.7	69.9	71.7	75.1	78.2
A07	77.3	79	84.5	86.6	88.3	93.3
A08	87.7	90.1	88.9	91.1	93	92.8
A09	70.6	76.4	75.4	80.9	82.5	95.8
Average	72.5±10.8	75.3±10.4	78.3±9.1	81.3±8.2	84±7.5	89.2±6.4

Table 5.9: Comparison of 5x5 cross-validation performance for left vs right hand task from BCI competition IV, Data set IIb

	CSP	CSP MT	SFBCSP	SFBCSP MT	TSGSP	TFCSP
B0103T	76.6	81.8	79.1	81.8	84	86.9
B0203T	56.3	58.7	59	60.3	62.6	66.9
B0303T	52.6	50.8	53.1	54	56.3	61.9
B0403T	98.9	98.9	98.9	99.1	99.4	98.8
B0503T	83.6	89.5	91.5	92.6	94.8	95.6
B0603T	67.1	68.5	81.3	82	83.8	83.6
B0703T	83.8	83.6	90.8	91.8	94.1	94.1
B0803T	86.5	88.3	88.9	91.1	93.3	92.3
B0903T	81.9	86.1	85.4	87.3	90.1	88.6
Average	76.4±15	78.5±15.8	80.9±15.3	82.2±15.3	84.3±15	85.4±12.8

Table 5.10: Comparison of 5x5 cross-validation performance from BCI competition III, Data set IIIa

	CSP	CSP MT	SFBCSP	SFBCSP MT	TSGSP	TFCSP
left hand vs right hand						
k3b	95.8	96	96.6	97.5	99.2	99.1
k6b	60	65.8	65.8	66.3	67.2	81.3
l1b	93.5	94.2	94.7	94.8	96.5	96.7
foot vs tongue						
k3b	84	86.8	88.9	94	96.4	95.3
k6b	78.4	80.3	83.8	85.2	87.8	95.3
l1b	73.2	74.7	78	81.5	84	79.8
Average	80.8±13.4	83±11.7	84.6±11.5	86.6±11.7	88.5±12	91.3±8.4

Table 5.11: Comparison of 10x10 cross-validation performance for left vs right hand task from BCI competition IV, Data set IIb

	CSP-TSM	SBLFB	I-DFBCSP	LRCSP	LRFCSP	TFCSP
B0103T	77.5	78.3	80.8	75.6	82.5	87.4
B0203T	60.5	59.3	58.4	55.7	60.0	67.3
B0303T	50.9	49.3	56.0	54.4	62.8	61.2
B0403T	99.3	99.1	99.4	96.9	99.4	98.4
B0503T	82.8	92.0	90.6	88.8	87.5	94.9
B0603T	77.0	79.5	82.0	65.6	81.9	82.9
B0703T	86.2	92.5	88.9	83.1	90.6	93.8
B0803T	88.6	88.9	89.5	82.5	91.3	92.6
B0903T	80.3	80.6	83.8	85.6	83.1	87.2
Average	78.1±14.6	79.9±16.3	81±14.6	76.5±14.9	82.1±13.0	85.1±12.8

Comparison with TGCSP and other approaches. Table 5.7, Table 5.8, Table 5.9 and Table 5.10 show the comparison of the average classification accuracy over 25 training and testing of the proposed method with TSGSP and other 4 competing methods in the literature reported in [102]. The last row of each table shows the average accuracy of the same motor imagery task (or for both motor imagery tasks in Table 5.10) over different subjects and the standard deviation. As can be seen from the results in these 4 tables, our proposed TFCSP method outperforms the 5 approaches reported in [102] by 8.8% for left vs right hand motor imagery task and 5.3% for tongue vs foot motor imagery task in Dataset IIa, 1.1% for left vs right hand motor imagery task in Dataset IIb and 2.8% for left vs right hand and tongue vs foot motor imagery tasks in Dataset IIIa. Moreover,

our proposed TFCSP also results in the smallest standard deviation for all these 5 motor imagery tasks, indicating good stability and robustness of the proposed method.

Comparison with LRFCSP and other approaches. Table 5.11 shows the comparison of the average classification accuracy over 100 training and testing of the proposed method with LRFCSP and other 4 competing methods in the literature reported in [91]. FBRCSPP [86] and PTFBCSP [105] are not included in Table 5.11 because LRFCSP has shown better performance than them in [91]. The last row of each table shows the average accuracy of the same motor imagery task over different subjects and the standard deviation.

As can be seen from the results in Table 5.11, our proposed method outperforms the best of the 5 approaches reported in [91] by 3% for left vs right hand motor imagery task in the dataset. For 7 out of 9 subjects our proposed method obtained the highest classification rate. Classification performance of subject *B0203T* and *B0305T* was improved from 60% to 67.31% and from 87.5% to 94.88% respectively. Moreover, our proposed TFCSP also results in the smallest standard deviation, indicating good stability and robustness of the proposed method. Note that the average classification accuracy of SBLFB reported in the original paper [97] (81.7%) is higher than that in Table VI given in [91]. The average classification accuracy of DFBCSP reported in the original paper [87] (81.2%) is higher than the average classification accuracy of its improved version I-DFBCSP given in [91]. Comparing the results given in [91] and [87], our proposed TFCSP still outperforms them by 3.4% and 3.9%, respectively.

5.4 Summary

This chapter proposes a novel Time-Frequency CSP (TFCSP) method that performs complex CSP on a time- and frequency decomposed array of cells. EEG

time signal that carries the information of a motor imagery process is decomposed into reaction, action and completion stages to better discriminate different motor imagery tasks. The proposed frequency decomposition leads to sharper spatial filters and generates rich informative features. The complex CSP which works in the frequency domain greatly reduces the computation complexity compared to the filter bank approaches. Moreover, we perform feature weighting and regularization approach instead of selecting leading spatial filters, which alleviates the information lost in feature selection.

The proposed method is evaluated on EEG data from 3 publicly available datasets. The proposed TFCSP in this chapter shows the best performance compared to other variants of CSP.

Chapter 6

Conclusion

This research aims to improve feature extraction using machine learning methods and achieve better results during classification of motor imagery tasks. Such improvements will eventually lead BCI technology to real life applications. In this thesis, new approaches of feature extraction and classification in motor imagery BCI are proposed.

Initially, this work addresses problems of extracting discriminant features from the EEG signal for motor imagery BCI based on a set of training samples. To alleviate the problems of overfitting or poor generalization and hence boost the classification accuracy, the CSP based approaches in motor imagery BCI select some CSP features and remove the others. However, this will result in the loss of discriminative information. This thesis proposes a novel approach that performs feature weighting and regularization (FWR) to replace the feature selection in CSP based approaches. It weights the features extracted by all spatial filters and then regularizes the eigenspectrum of the within-scatter matrix of the weighted feature vector for reliable classification. The proposed method can be applied in all CSP based approaches in MI-BCI.

The thesis demonstrates experiments for FWR on two publicly available datasets:

BCI competition IV Dataset IIa and BCI competition III Dataset IIIa. The proposed FWR method enhances the classification accuracy as compared to the conventional feature selection approaches.

Secondly, this research proposes a novel time-frequency CSP (TFCSP) method that performs complex CSP on a time and frequency decomposed array of cells. EEG time signal that carries the information of a motor imagery process is decomposed into reaction, action and completion stages to discriminate different motor imagery tasks better. The proposed frequency decomposition leads to sharper spatial filters and generates rich informative features. The complex CSP which works in the frequency domain greatly reduces the computational complexity in comparison to the filter bank approaches. Moreover, we perform feature weighting and regularization approach instead of selecting leading spatial filters, which alleviates the information lost in feature selection.

The thesis demonstrates experiments for TFCSP on three publicly available datasets: BCI competition IV Dataset IIa, Dataset IIb and BCI competition III Dataset IIIa. The proposed algorithm TFCSP yields higher classification, for the same BCI competition datasets, than existing methods.

6.1 Future work

In this research experiments were performed for binary task classification. For future works it can be extended from binary task to multiclass classification, e.g. using 4 classes such as left hand, right hand, tongue, and foot.

The experiments were also conducted in offline mode with recorded datasets. While for real-life applications, online mode is necessary. In future works, experiments can be conducted in online mode with EEG headsets during motor imagery tasks with different subjects, including subjects with various neuromuscular disorders.

Finally, the methods proposed in this work use subject-specific classification, where the classifier is trained on only one subject. Future work can extend the methods to cross-subject classification, where different subjects are used for training.

Author's Publications

Journal Papers

1. V. Mishuhina and X. Jiang. Feature weighting and regularization of common spatial patterns in eeg-based motor imagery bci. *IEEE Signal Processing Letters*, 25(6):783–787, June 2018.
2. V. Mishuhina and X. Jiang. Complex common spatial patterns on time-frequency decomposed eeg for motor imagery brain-computer interface. *submitted*, 2019.

Bibliography

- [1] J. J. Vidal. Toward direct brain-computer communication. *Annual Review of Biophysics and Bioengineering*, 2:157–180, June 1973.
- [2] Jonathan R. Wolpaw and Elizabeth Winter Wolpaw. *Brain-computer interfaces. [electronic resource] : principles and practice*. New York ; Oxford : Oxford University Press, 2012., 2012.
- [3] VV. Pravdich-Neminsky. Ein versuch der registrierung der elektrischen gehirnerscheinungen. *Zentralblatt fur Physiologie*, 27:951–960, 1913.
- [4] H. Berger. Uber das elektroenkephalogramm des menschen. *Archiv fur Psychatrie*, 87:527–570, 1929.
- [5] E. Yin, Z. Zhou, J. Jiang, F. Chen, Y. Liu, and D. Hu. A speedy hybrid bci spelling approach combining p300 and ssvep. *IEEE Transactions on Biomedical Engineering*, 61(2):473–483, Feb 2014.
- [6] W. Speier, C. Arnold, J. Lu, A. Deshpande, and N. Pouratian. Integrating language information with a hidden markov model to improve communication rate in the p300 speller. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(3):678–684, May 2014.
- [7] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan. Brain-computer interface research at the wadsworth center. *IEEE Transactions on Rehabilitation Engineering*, 8(2):222–226, Jun 2000.
- [8] M. Akcakaya, B. Peters, M. Moghadamfalahi, A. R. Mooney, U. Orhan, B. Oken, D. Erdogmus, and M. Fried-Oken. Noninvasive brain-computer interfaces for augmentative and alternative communication. *IEEE Reviews in Biomedical Engineering*, 7:31–49, 2014.
- [9] A. Frisoli, C. Loconsole, D. Leonardis, F. Banno, M. Barsotti, C. Chisari, and M. Bergamasco. A new gaze-bci-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6):1169–1179, Nov 2012.
- [10] K. K. Ang and C. Guan. Brain-computer interface in stroke rehabilitation. *Journal of Computing Science and Engineering*, 7(2):139–146, 2013.

- [11] T. P. Chou, W. R. Wang, and T. S. Chang. Low complexity real time bci for stroke rehabilitation. In *2015 IEEE International Conference on Digital Signal Processing (DSP)*, pages 809–812, July 2015.
- [12] M. Barsotti, D. Leonardis, C. Loconsole, M. Solazzi, E. Sotgiu, C. Procopio, C. Chisari, M. Bergamasco, and A. Frisoli. A full upper limb robotic exoskeleton for reaching and grasping rehabilitation triggered by mi-bci. In *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*, pages 49–54, Aug 2015.
- [13] Vikram Shenoy Handiru. *Neural signal processing techniques to analyse motor functions in EEG-based brain-computer interfaces*. Interdisciplinary Graduate School, Nanyang Technological University, 2017.
- [14] T. Carlson and J. del R. Millan. Brain-controlled wheelchairs: A robotic architecture. *IEEE Robotics Automation Magazine*, 20(1):65–73, March 2013.
- [15] Y. Li, J. Pan, F. Wang, and Z. Yu. A hybrid bci system combining p300 and ssvep and its application to wheelchair control. *IEEE Transactions on Biomedical Engineering*, 60(11):3156–3166, Nov 2013.
- [16] D. Huang, K. Qian, D. Y. Fei, W. Jia, X. Chen, and O. Bai. Electroencephalography (eeg)-based brain-computer interface (bci): A 2-d virtual wheelchair control based on event-related desynchronization/synchronization and state control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(3):379–388, May 2012.
- [17] G. R. Muller-Putz and G. Pfurtscheller. Control of an electrical prosthesis with an ssvep-based bci. *IEEE Transactions on Biomedical Engineering*, 55(1):361–364, Jan 2008.
- [18] Jonathan R. Wolpaw and Dennis J. McFarland. Brain-computer interface operation of robotic and prosthetic devices. *Computer*, 41:52–56, 2008.
- [19] X. Zhang, R. Li, and Y. Li. Research on brain control prosthetic hand. In *Ubiquitous Robots and Ambient Intelligence (URAI), 2014 11th International Conference on*, pages 554–557, Nov 2014.
- [20] F. Wang, C. Zhou, X. Hao, S. Wang, and G. Yang. Bci control system for humanoid robot based on motor imaginary. In *2013 25th Chinese Control and Decision Conference (CCDC)*, pages 5140–5143, May 2013.
- [21] W. Song, X. Wang, S. Zheng, and Y. Lin. Mobile robot control by bci based on motor imagery. In *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2014 Sixth International Conference on*, volume 2, pages 383–387, Aug 2014.
- [22] Y. Chae, J. Jeong, and S. Jo. Toward brain-actuated humanoid robots: Asynchronous direct control using an eeg-based bci. *IEEE Transactions on Robotics*, 28(5):1131–1144, Oct 2012.

- [23] J. Long, Y. Li, T. Yu, and Z. Gu. Target selection with hybrid feature for bci-based 2-d cursor control. *IEEE Transactions on Biomedical Engineering*, 59(1):132–140, Jan 2012.
- [24] Y. Li, J. Long, T. Yu, Z. Yu, C. Wang, H. Zhang, and C. Guan. An eeg-based bci system for 2-d cursor control by combining mu/beta rhythm and p300 potential. *IEEE Transactions on Biomedical Engineering*, 57(10):2495–2505, Oct 2010.
- [25] Michael Tangermann, Fabien Lotte, and Jan B. F. van Erp. Brain-computer interfaces: Beyond medical applications. *Computer*, 45(undefined):26–34, 2012.
- [26] D. Marshall, D. Coyle, S. Wilson, and M. Callaghan. Games, gameplay, and bci: The state of the art. *IEEE Transactions on Computational Intelligence and AI in Games*, 5(2):82–99, June 2013.
- [27] L. Bonnet, F. Lotte, and A. Lécuyer. Two brains, one game: Design and evaluation of a multiuser bci video game based on motor imagery. *IEEE Transactions on Computational Intelligence and AI in Games*, 5(2):185–198, June 2013.
- [28] Kirill A. Sorudeykin. An educative brain-computer interface. *Computing Research Repository*, abs/1003.2660, 2010.
- [29] Arvaneh Mahnaz T. Toshihisa. *Signal Processing and Machine Learning for Brain-Machine Interfaces*. Institution of Engineering and Technology, 2018.
- [30] Bernhard Graimann, Brendan Z. Allison, and Gert Pfurtscheller. *Brain-Computer Interfaces Revolutionizing Human-Computer Interaction*. Springer Berlin, 2013.
- [31] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil. Brain computer interfaces, a review. *Sensors*, 12(2):1211–1279, 2012.
- [32] Dayo O. Adewole, Mijail D. Serruya, James P. Harris, Justin C. Burrell, Dmitriy Petrov, H. Isaac Chen, John A. Wolf, and D. Kacy Cullen. The evolution of neuroprosthetic interfaces. *Critical Reviews in Biomedical Engineering*, 44(1-02):123–152, 2016.
- [33] Seung-Schik Yoo Byoung-Kyong Min, Matthew J. Marzelli. Neuroimaging-based approaches in the brain-computer interface. *Trends in Biotechnology*, 28(11):552 – 560, 2010.
- [34] J. Zhang, G. Sudre, X. Li, W. Wang, D. J. Weber, and A. Bagic. Clustering linear discriminant analysis for meg-based brain computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19(3):221–231, June 2011.

- [35] Yang Wang and Dongrui Wu. Real-time fmri-based brain computer interface: A review. In *Neural Information Processing*, pages 833–842, Cham, 2017. Springer International Publishing.
- [36] N Naseer and K Hong. fnirs-based brain-computer interfaces: a review. *Frontiers in Human Neuroscience*, 9(3), 2015.
- [37] Nidal Kamel and Aamir Saeed Malik. *EEG/ERP analysis. [electronic resource] : methods and applications*. Boca Raton : CRC Press, Taylor and Francis Group, 2014., 2015.
- [38] Saeid Sanei and Jonathon Chambers. *EEG signal processing*. Chichester, England ; Hoboken, NJ : John Wiley and Sons, c2007., 2007.
- [39] Niels Birbaumer. Brain-computer-interface research: Coming of age. *Clinical Neurophysiology*, 117(3):479 – 483, 2006.
- [40] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller. Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2008.
- [41] H. H. Jasper. The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10:371–375, 1958.
- [42] C. Fonseca, J. P. S. Cunha, R. E. Martins, V. M. Ferreira, J. P. M. de Sa, M. A. Barbosa, and A. Martins da Silva. A novel dry active electrode for eeg recording. *IEEE Transactions on Biomedical Engineering*, 54(1):162–165, Jan 2007.
- [43] A.R. Mota, L. Duarte, D. Rodrigues, A.C. Martins, A.V. Machado, F. Vaz, P. Fiedler, J. Haueisen, J.M. Nóbrega, and C. Fonseca. Development of a quasi-dry electrode for eeg recording. *Sensors and Actuators A: Physical*, 199:310 – 317, 2013.
- [44] C Neuper and G Pfurtscheller. Event-related dynamics of cortical rhythms: frequency-specific features and functional correlates. *International Journal of Psychophysiology*, 43(Thalamo-Cortical Relationships):41 – 58, 2001.
- [45] Wikipedia contributors. Electroencephalography — Wikipedia, the free encyclopedia, 2019. [Online].
- [46] Suzanaerculano-Houzel. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, 3:31, 2009.
- [47] Loos Peter, Riedl Ren’e, , Gernot R. Müller-Putz, Jan vom Brocke, Fred D. Davis, Rajiv D. Banker, and Pierre-Majorique Léger. Neurois: Neuroscientific approaches in the investigation and development of information systems. *Business & Information Systems Engineering*, 2(6):395–401, Dec 2010.
- [48] Xiaogang Chen, Yijun Wang, Masaki Nakanishi, Xiaorong Gao, Tzyy-Ping Jung, and Shangkai Gao. High-speed spelling with a noninvasive

- brain-computer interface. *Proceedings of the National Academy of Sciences*, 112(44):E6058–E6067, 2015.
- [49] G. R. Muller-Putz, R. Scherer, C. Neuper, and G. Pfurtscheller. Steady-state somatosensory evoked potentials: suitable brain signals for brain-computer interfaces? *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(1):30–37, March 2006.
 - [50] Y. Punsawad and Y. Wongsawat. On the enhancement of training session performance via attention for single-frequency/multi-commands based steady state auditory evoked potential bci. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1761–1764, Aug 2012.
 - [51] H. Serby, E. Yom-Tov, and G. F. Inbar. An improved p300-based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(1):89–98, March 2005.
 - [52] H. Hou, Q. Meng, M. Zeng, and B. Sun. Improving classification of slow cortical potential signals for bci systems with polynomial fitting and voting support vector machine. *IEEE Signal Processing Letters*, 25(2):283–287, Feb 2018.
 - [53] N. Birbaumer, T. Hinterberger, A. Kubler, and N. Neumann. The thought-translation device (ttd): neurobehavioral mechanisms and clinical outcome. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):120–123, June 2003.
 - [54] G. Pfurtscheller and F. H. Lopes da Silva. Event-related eeg/meg synchronization and desynchronization: Basic principles. *Clinical Neurophysiology*, 110(11):1842 – 1857, 1999.
 - [55] G. Pfurtscheller, C. Brunner, A. Schlögl, and F.H. Lopes da Silva. Mu rhythm (de)synchronization and eeg single-trial classification of different motor imagery tasks. *Neuroimage*, 31:153 – 159, 2006.
 - [56] Bin He. *Neural Engineering. [electronic resource]*. Boston, MA : Springer US : Imprint: Springer, 2013., 2013.
 - [57] H. Yuan and B. He. Brain-computer interfaces using sensorimotor rhythms: Current state and future perspectives. *IEEE Transactions on Biomedical Engineering*, 61(5):1425–1435, May 2014.
 - [58] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger. Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *Rehabilitation Engineering, IEEE Transactions on*, 6(3):316–325, September 1998.
 - [59] Herbert Jasper and Wilder Penfield. Electrocorticograms in man: Effect of voluntary movement upon the electrical activity of the precentral gyrus. *Archiv for Psychiatrie und Nervenkrankheiten*, 183(1):163–174, 1949.

- [60] Qingguo Wei, Yijun Wang, Xiaorong Gao, and Shangkai Gao. Amplitude and phase coupling measures for feature extraction in an eeg-based brain-computer interface. *Journal Of Neural Engineering*, 4(2):120 – 129, 2007.
- [61] Pawel Herman, Girijesh Prasad, Thomas Martin McGinnity, and Damien Coyle. Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification. *IEEE Transactions On Neural Systems And Rehabilitation Engineering: A Publication Of The IEEE Engineering In Medicine And Biology Society*, 16(4):317 – 326, 2008.
- [62] George Townsend, Bernhard Graimann, and Gert Pfurtscheller. A comparison of common spatial patterns with complex band power features in a four-class bci experiment. *IEEE Transactions on Biomedical Engineering*, 53(4):642 – 651, 2006.
- [63] Dennis J McFarland and Jonathan R Wolpaw. Sensorimotor rhythm-based brain-computer interface (bci): model order selection for autoregressive spectral analysis. *Journal Of Neural Engineering*, 5(2):155 – 162, 2008.
- [64] Separability of eeg signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Transactions On Rehabilitation Engineering: A Publication Of The IEEE Engineering In Medicine And Biology Society*, 6(3):316 – 325, 1998.
- [65] Lei Qin and Bin He. A wavelet-based time-frequency analysis approach for classification of motor imagery for brain-computer interface applications. *Journal Of Neural Engineering*, 2(4):65 – 72, 2005.
- [66] Joseph Dien, Kevin M. Spencer, and Emanuel Donchin. Localization of the event-related potential novelty response as defined by principal components analysis. *Cognitive Brain Research*, 17(3):637 – 650, 2003.
- [67] A. Erfanian and A. Erfani. Ica-based classification scheme for eeg-based brain-computer interface: The role of mental practice and concentration skills. In *Proceedings Of The Annual International Conference- IEEE Engineering In Medicine And Biology Society*, volume 1 of *IEEE Engineering in Medicine and Biology Society; EMBC 2004*, pages 235 – 238, 2004.
- [68] Aboul Ella Hassanien and Ahmad Taher Azar. *Brain-computer interfaces : current trends and applications*. Intelligent Systems Reference Library: volume 74. Cham : Springer, [2014], 2014.
- [69] H Ramoser, J Müller-Gerking, and G Pfurtscheller. Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE Transactions On Rehabilitation Engineering: A Publication Of The IEEE Engineering In Medicine And Biology Society*, 8(4):441 – 446, 2000.
- [70] Steven Lemm, Benjamin Blankertz, Gabriel Curio, and Klaus Robert Muller. Spatio-spectral filters for improving the classification of single trial eeg. *IEEE Transactions on Biomedical Engineering*, 52:1541–1548, 2005.

- [71] Guido Dornhege, Benjamin Blankertz, Matthias Krauledat, Florian Losch, Gabriel Curio, and Klaus Robert Muller. Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE Transactions on Biomedical Engineering*, 53:2274–2281, 2006.
- [72] Quadrianto Novi, Cuntai Guan, Tran Huy Dat, and Ping Xue. Sub-band common spatial pattern (sbcsp) for brain-computer interface. *Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering*, pages 204–207, 2007.
- [73] K. K. Ang, Z. Y. Chin, H. H. Zhang, and C. T. Guan. Filter bank common spatial pattern (fbcsp) in brain-computer interface. *Proceeding of IEEE International Joint Conference on Neural Networks*, pages 2390–2397, 2008.
- [74] Fabien Lotte and Cuntai Guan. Regularizing common spatial patterns to improve bci designs: Unified theory and new algorithms. *IEEE Transactions on Biomedical Engineering*, 58:355–362, 2011.
- [75] Wei Wu, Zhe Chen, Senior Member, and Xiaorong Gao. Probabilistic common spatial patterns for multichannel eeg analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37:639–653, 2015.
- [76] Guido Dornhege, Benjamin Blankertz, Gabriel Curio, and Klaus Robert Muller. Boosting bit rates in noninvasive eeg single-trial classifications by feature combination and multiclass paradigms. *IEEE Transactions on Biomedical Engineering*, 51:993–1002, 2004.
- [77] Moritz Grosse-Wentrup and Martin Buss. Multiclass common spatial patterns and information theoretic feature extraction. *IEEE Transactions on Biomedical Engineering*, 55:1991–2000, 2008.
- [78] Wei Wu, Xiaorong Gao, and Shangkai Gao. One-versus-the-rest (ovr) algorithm: An extension of common spatial patterns (csp) algorithm to multiclass case. *Proceedings of IEEE Engineering in Medicine and Biology Society Conference*, 3:2387–2390, 2005.
- [79] Youngjoo Kim, Jiwoo Ryu, Ko Keun Kim, Clive Cheong Took, Danilo P. Mandic, and Cheolsoo Park. Motor imagery classification using mu and beta rhythms of eeg with strong uncorrelating transform based complex common spatial patterns. In *Comp. Int. and Neurosc.*, 2016.
- [80] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek. Optimizing the channel selection and classification accuracy in eeg-based bci. *IEEE Transactions on Biomedical Engineering*, 58(6):1865–1873, June 2011.
- [81] Alejandro A. Torres-García, Carlos A. Reyes-García, Luis Villaseñor-Pineda, and Gregorio García-Aguilar. Implementing a fuzzy inference system in a multi-objective eeg channel selection model for imagined speech classification. *Expert Systems with Applications*, 59:1 – 12, 2016.

- [82] A. Franklin Alex Joseph and C. Govindaraju. Channel selection using glow swarm optimization and its application in line of sight secure communication. *Cluster Computing*, 22(5):10801–10808, Sep 2019.
- [83] Zheng Yang Chin, Kai Keng Ang, Chuanchu Wang, Cuntai Guan, and Haihong Zhang. Multi-class filter bank common spatial pattern for four-class motor imagery bci. *Proceedings of the IEEE Engineering in Medicine and Biology Society Conference*, 138632:571–574, 2009.
- [84] Dong Huang, Cuntai Guan, Kai Keng Ang, and Zheng Yang Chin. Time-Constrained Filter Bank Common Spatial Pattern for Motor Imagery Brain-Computer Interfaces. *203 Proceedings of the Second APSIPA Annual Summit and Conference*, (December):203–206, 2010.
- [85] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan. Composite filter bank common spatial pattern for motor imagery-based brain-computer interface. In *Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2011 IEEE Symposium on*, pages 1–5, 2011.
- [86] S. Park, D. Lee, and S. Lee. Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(2):498–505, Feb 2018.
- [87] H. Higashi and T. Tanaka. Simultaneous design of fir filter banks and spatial patterns for eeg signal classification. *IEEE Transactions on Biomedical Engineering*, 60(4):1100–1110, April 2013.
- [88] K. P. Thomas, C. Guan, L. C. Tong, and A. P. Vinod. Discriminative filter-bank selection and eeg information fusion for brain computer interface. In *2009 IEEE International Symposium on Circuits and Systems*, pages 1469–1472, May 2009.
- [89] G. Sun, J. Hu, and G. Wu. A novel frequency band selection method for common spatial pattern in motor imagery based brain computer interface. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6, July 2010.
- [90] Yu Zhang, Guoxu Zhou, Jing Jin, Xingyu Wang, and Andrzej Cichocki. Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface. *Journal of Neuroscience Methods*, 255:85 – 91, 2015.
- [91] Y. Park and W. Chung. Frequency-optimized local region common spatial pattern approach for motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(7):1378–1388, July 2019.
- [92] Ryota Tomioka, Guido Dornhege, Guido Nolte, Benjamin Blankertz, Kazuyuki Aihara, and Klaus-Robert Muller. Spectrally weighted Common Spatial Pattern algorithm for single trial EEG classification. *Dept Math Eng Univ Tokyo Tokyo Japan Tech Rep*, 40(July):25, 2006.

- [93] W. Wu, X. Gao, B. Hong, and S. Gao. Classifying single-trial eeg during motor imagery by iterative spatio-spectral patterns learning (isspl). *IEEE Transactions on Biomedical Engineering*, 55(6):1733–1743, 2008.
- [94] Haihong Zhang, Zheng Yang Chin, Kai Keng Ang, Cuntai Guan, and Chuanchu Wang. Optimum Spatio-Spectral Filtering Network for Brain – Computer Interface. *IEEE Transactions On Neural Networks*, 22(1):52–63, 2011.
- [95] H. I. Suk and S. W. Lee. A novel bayesian framework for discriminative feature extraction in brain-computer interfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):286–299, 2013.
- [96] J. Meng, L. Yao, X. Sheng, D. Zhang, and X. Zhu. Simultaneously optimizing spatial spectral features based on mutual information for eeg classification. *IEEE Transactions on Biomedical Engineering*, 62(1):227–240, Jan 2015.
- [97] Yu Zhang, Yu Wang, Jing Jin, and Xingyu Wang. Sparse bayesian learning for obtaining sparsity of eeg frequency bands based feature vectors in motor imagery classification. *International Journal of Neural Systems*, 27(02):1650032, 2017. PMID: 27377661.
- [98] Amirhossein S Aghaei, Mohammad Shahin Mahanta, and Student Member. Separable Common Spatio-Spectral Patterns for Motor Imagery BCI Systems. *IEEE Transactions on Biomedical Engineering*, 2016.
- [99] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan. Mutual information-based selection of optimal spatial-temporal patterns for single-trial eeg-based bcis. *Pattern Recognition*, 45(6):2137 – 2144, 2012. Brain Decoding.
- [100] N. Yamawaki, C. Wilke, Zhongming Liu, and Bin He. An enhanced time-frequency-spatial approach for motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):250–254, June 2006.
- [101] Yuan Yang, Sylvain Chevallier, Joe Wiart, and Isabelle Bloch. Subject-specific time-frequency selection for multi-class motor imagery-based bcis using few laplacian eeg channels. *Biomedical Signal Processing and Control*, 38:302 – 311, 2017.
- [102] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki. Temporally constrained sparse group spatial patterns for motor imagery bci. *IEEE Transactions on Cybernetics*, 49(9):3322–3332, Sep. 2019.
- [103] S. Selim, M. M. Tantawi, H. A. Shedeed, and A. Badr. A csp am-ba-svm approach for motor imagery bci system. *IEEE Access*, 6:49192–49208, 2018.
- [104] Jianjun Meng, Gan Huang, Dingguo Zhang, and Xiangyang Zhu. Optimizing spatial spectral patterns jointly with channel configuration for brain-computer interface. *Neurocomputing*, 104:115 – 126, 2013.

- [105] Victoria Peterson, Dominik Wyser, Olivier Lambercy, Ruben Spies, and Roger Gassert. A penalized time-frequency band feature selection and classification procedure for improved motor intention decoding in multichannel EEG. *Journal of Neural Engineering*, 16(1):016019, jan 2019.
- [106] B. Chen, Y. Li, J. Dong, N. Lu, and J. Qin. Common spatial patterns based on the quantized minimum error entropy criterion. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pages 1–12, 2018.
- [107] F. Yger, M. Berar, and F. Lotte. Riemannian approaches in brain-computer interfaces: A review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1753–1762, Oct 2017.
- [108] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten. Common spatial pattern revisited by riemannian geometry. In *2010 IEEE International Workshop on Multimedia Signal Processing*, pages 472–476, Oct 2010.
- [109] Shiu Kumar, Kabir Mamun, and Alok Sharma. Csp-tsm: Optimizing the performance of riemannian tangent space mapping using common spatial pattern for mi-bci. *Computers in Biology and Medicine*, 91:231 – 242, 2017.
- [110] Xiang Zhang, Lina Yao, Xianzhi Wang, Jessica Monaghan, and David McAlpine. A survey on deep learning based brain computer interface: Recent advances and new frontiers. *Computing Research Repository*, abs/1905.04149, 2019.
- [111] Z. Zhang, F. Duan, J. Solé-Casals, J. Dinarès-Ferran, A. Cichocki, Z. Yang, and Z. Sun. A novel deep learning approach with data augmentation to classify motor imagery signals. *IEEE Access*, 7:15945–15954, 2019.
- [112] Mengxi Dai, Dezhi Zheng, Rui Na, Shuai Wang, and Shuailei Zhang. Eeg classification of motor imagery using a novel deep learning framework. *Sensors*, 19(3):551, Jan 2019.
- [113] N. Lu, T. Yin, and X. Jing. A temporal convolution network solution for eeg motor imagery classification. In *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 796–799, 2019.
- [114] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, 16(5):051001, aug 2019.
- [115] Tsunoda T. Kumar S, Sharma A. An improved discriminative filter bank selection approach for motor imagery eeg signal classification using mutual information. *BMC Bioinformatics*, 18:545, December 2017.
- [116] N Jrad, M Congedo, R Phlypo, S Rousseau, R Flamary, F Yger, and A Rakotomamonjy. sw-SVM: sensor weighting support vector machines for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 8(5):056004, aug 2011.

- [117] G. V. Trunk. A problem of dimensionality: A simple example. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(3):306–307, July 1979.
- [118] A. K. Jain, R. P. W. Duin, and Jianchang Mao. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, Jan 2000.
- [119] Jiang Xudong. Linear subspace learning-based dimensionality reduction. *IEEE Signal Processing Magazine, Signal Processing Magazine, IEEE, IEEE Signal Process. Mag*, (2):16, 2011.
- [120] X. Jiang, B. Mandal, and A. Kot. Eigenfeature regularization and extraction in face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):383–394, March 2008.
- [121] Xudong Jiang. Asymmetric principal component and discriminant analyses for pattern classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):931 – 937, 2009.
- [122] Jerome H. Friedman. Regularized discriminant analysis. *Journal of the American Statistical Association*, 84(405):165–175, 1989.
- [123] X. D. Jiang, B. Mandal, and A. Kot. Enhanced maximum likelihood face recognition. *Electronics Letters*, 42(19):1089–1090, Sep. 2006.
- [124] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis. Color local texture features for color face recognition. *IEEE Transactions on Image Processing*, 21(3):1366–1380, March 2012.
- [125] J. Y. Choi, W. De Neve, K. N. Plataniotis, and Y. M. Ro. Collaborative face recognition for improved face annotation in personal photo collections shared on online social networks. *IEEE Transactions on Multimedia*, 13(1):14–28, Feb 2011.
- [126] Z. Lu, X. Jiang, and A. C. Kot. A color channel fusion approach for face recognition. *IEEE Signal Processing Letters*, 22(11):1839–1843, Nov 2015.
- [127] Bci competition 4 data set iia. http://bbci.de/competition/iv/desc_2a.pdf.
- [128] Bci competition 3 data set iiia. http://www.bbci.de/competition/iii/desc_IIIa.pdf.
- [129] Song Le and Julien Epps. Classifying eeg for brain-computer interface: Learning optimal filters for dynamical system features. *Computational Intelligence and Neuroscience*, pages 1 – 11, 2007.
- [130] V. Mishuhina and X. Jiang. Feature weighting and regularization of common spatial patterns in eeg-based motor imagery bci. *IEEE Signal Processing Letters*, 25(6):783–787, June 2018.

- [131] Bci competition 4 dataset iib. http://www.bbc.de/competition/iv/desc_2b.pdf.
- [132] V. Mishuhina and X. Jiang. Complex common spatial patterns on time-frequency decomposed eeg for motor imagery brain-computer interface. *submitted*, 2019.