

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.



Doctoral Thesis

Amardeep Singh

Supervisors

Prof. Hans W. Guesgen and Dr. Sunil Lal

School of Fundamental Sciences

Massey University, Manawatu, New Zealand

A thesis presented in partial fulfilment of the requirements for the degree of Doctor of
Philosophy in Computer Science

2022

Recommended Citation

Amardeep Singh (2022), Reducing Calibration Time in Motor Imagery Based Brain-Computer Interface. School of Fundamental Sciences, Massey University, Palmerston North, New Zealand.

Declaration

The thesis complies with the ‘Guidelines for Doctoral Thesis by Publications’ and with the requirements from the Handbook for Doctoral Study by the Doctoral Research Committee (DRC), Massey University. January 2011. Version 7.

Disclaimer

The opinions, figures and conclusions in the thesis are solely those of the author(s). Under no circumstances will the author(s) be responsible for any loss or damage of any kind resulted from the use of these methods and techniques presented in the thesis.

REDUCING CALIBRATION TIME IN
MOTOR IMAGERY BASED-BRAIN
COMPUTER INTERFACE

A THESIS PRESENTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
SCHOOL OF FUNDAMENTAL SCIENCES
AT MASSEY UNIVERSITY, PALMERSTON NORTH,
NEW ZEALAND.

Amardeep Singh

2022

Contents

Abstract	ix
Acknowledgements	xii
1 Introduction	1
1.1 Problem Statement & Research Contribution	4
1.2 Structure of the Thesis	7
1.3 Research outcomes	10
2 A Review on Critical Issues and Possible Solutions of Motor Imagery based BCI	12
2.1 Introduction	13
2.2 Architecture of MI based BCI	16
2.2.1 Data Acquisition	17
2.2.2 MI training	18
2.2.3 Signal pre-processing and artifacts removal	19
2.2.4 Feature Extraction	21
2.2.5 Channel and Feature Selection	28
2.2.6 Dimensionality Reduction	31
2.2.7 Classification	33
2.2.8 Performance Evaluation	36
2.3 Key issues in MI-BCI	38
2.3.1 Enhancement of MI-BCI performance	38
2.3.2 Reduce or Zero calibration time	43
2.3.3 BCI illiteracy	46
2.3.4 Asynchronised MI-BCI	48
2.3.5 Increase number of commands	49

2.3.6	Adaptive BCI	50
2.3.7	Online MI-BCI	52
2.3.8	Training Protocol	53
2.4	Conclusion	54
3	Reduce Calibration Time Using Spatially Regularized Symmetric Positive-Definite Matrices	57
3.1	Introduction	58
3.2	Geometry of SPD Matrices	61
3.2.1	Riemannian Manifold	62
3.2.2	Riemannian Distance	63
3.2.3	The Choice of a Reference SPD matrix	64
3.2.4	Minimum Distance to Riemannian Mean (MDRM)	64
3.3	Methodology	65
3.4	Data and Experiment	68
3.4.1	Experimental Setup	69
3.4.2	Evaluation Metrics	70
3.5	Results and Discussion	71
3.5.1	Dataset IVa, BCI Competition III	71
3.5.2	Dataset IIIa, BCI Competition III	74
3.5.3	Dataset IIa, BCI Competition IV	76
3.6	Conclusion	77
4	Small Sample Motor Imagery Classification Using Regularized Riemannian geometry Features	79
4.1	Introduction	80
4.2	Geometry of SCM matrices	84
4.2.1	Data Model	84
4.2.2	Riemannian Manifold	85
4.2.3	Minimum Distance to Riemannian Mean (MDRM)	87
4.3	Methodology	88
4.4	Data and Experiments	92
4.4.1	Experimental Setup	93
4.4.2	Evaluation Metrics	94
4.5	Results and Discussion	94
4.5.1	Dataset IVa, BCI competition III	95

4.5.2	Dataset IIIa, BCI competition III	101
4.5.3	Dataset IIa, BCI competition IV	103
4.6	Conclusion	104
5	Statistical Testing and Discussion	105
5.1	Introduction	106
5.2	Methodology	106
5.3	Results and Discussion	107
5.3.1	Data set IVa, BCI competition III	109
5.3.2	Dataset IIIa, BCI competition III	112
5.3.3	Dataset IIa, BCI competition IV	115
5.4	Conclusion	117
6	Conclusion and Future Work	119
6.1	Conclusion	119
6.2	Limitations and directions of future Work	121
	References	124

List of Tables

2.1	This table provides a summary of the feature extraction methods	28
2.2	This table provides a summary of the classification methods describe in the section 2.2.7	36
2.3	Multi class Confusion matrix	37
2.4	Summary of all the Metrics	38
3.1	Summary of Dataset IVa, Dataset IIIa and Dataset IIa from BCI competitions (BCIC).	69
3.2	Hyper-parameters α, r of SR-MDRM for all subjects belonging to different datasets.	71
3.3	Classification accuracy (Mean and Standard deviation in percent) of the proposed approach and other MI classification approaches on Dataset IVa, BCI Competition III.	72
3.4	The performance of proposed approach and existing Riemannian geometry based approaches on Dataset IVa of BCI Competition III in terms of kappa values.	74
3.5	Classification accuracy (Mean and Standard deviation in percent)of the proposed approach and other MI classification approaches on Dataset IIIa, BCI Competition III.	75
3.6	Classification accuracy (mean and standard deviation in percent) of the proposed approach and other approaches on Dataset IIa, BCIC IV.	77
4.1	Summary of dataset IVa, dataset IIIa and dataset IIa from BCI competitions (BCIC)	92
4.2	Classification accuracy of the proposed approach and other conventional approaches on dataset IVa, BCI Competition III . . .	95

4.3	Regularization hyper-parameters β, γ of R-MDRM for all subjects belonging to different datasets	95
4.4	Classification accuracy of the proposed approach and existing Euclidean approaches on dataset IVa, BCIC III	100
4.5	The performance of proposed approach and existing Riemannian geometry based approaches on Dataset IVa of BCI Competition III in terms of kappa values	101
4.6	Classification accuracy of the proposed approach and other conventional approaches on dataset IIIa, BCIC III	101
4.7	Classification accuracy of the proposed approach and existing approaches on dataset IIIa, BCIC III	102
4.8	Classification accuracy of the proposed approach and other conventional approaches on dataset IIa, BCIC IV	103
4.9	Classification accuracy of the proposed approach and existing methods in literature on dataset IIa, BCIC IV	103

List of Figures

1.1	Block diagram of general online MI-BCI	3
1.2	The aim and challenges addressed in this thesis	6
2.1	Breakdown of the chapter	15
2.2	Block diagram showing the typical structure of MI-based BCI . . .	17
2.3	The international 10–20 standard electrode position system. The left image presents a left side view of the head with electrode positions, and the right image shows a top view of the head. . . .	18
2.4	An illustration of one trial’s timing in the Graz protocol	20
2.5	An illustration of the Riemannian Manifold displaying tangent space at point C	28
3.1	Illustration of Tangent space of manifold \mathcal{M} at point C . The Logarithmic map projects the matrix $C_i \in \mathcal{M}$ into the tangent space. The Exponential map projects the element of the tangent space S_i back to the manifold. Geodesic curve illustrate unique and shortest curve connecting C_i and C	62
3.2	Framework for proposed approach.	65
3.3	Classification accuracy of all subjects from Dataset IVa with respect to hyper-parameter r and best value of α	73
3.4	Classification accuracy of the proposed approach and other MI classification approaches on Dataset IIIa, BCI Competition III. . .	75
3.5	Classification accuracy according to hyper-parameter r and $10 \log_{10}(\alpha)$ of the proposed approach on subject $K6B$ from Dataset IIIa, BCI Competition III.	76
4.1	Framework for MDRM approach	85

4.2	Illustration of Tangent space of manifold \mathcal{M} at point C . Tangent vector S_i is the projection of C_i . Geodesic curve illustrate unique and shortest curve connecting C_i and C . The length of geodesic curve is Riemannian distance between C_i and C	87
4.3	A block diagram of proposed approach	89
4.4	Accuracy according to number of trials (M) from other subjects .	96
4.5	Accuracy of AA with respect to hyper-parameters β, γ values . . .	97
4.6	Covariance matrices of subject “ay” test trials (a)without spatial filtering,(b) After spatial filtering done through CSP filters and (c) After spatial filtering done through filters obtained by transfer learning, visualize using t-SNE	98
4.7	Electrode weights for filters obtained for different subjects, (a) and (c) shows electrode weights of filters obtained though CSP algorithm for subjects “av” and “ay” respectively. Similarly (b) and (d) shows electrode weights for filters obtained through transfer learning based CSP algorithm for subjects “av” and “ay” respectively	98
4.8	Electrode weights for filters obtained for subjects “k6b”, (a) shows electrode weights of filter obtained though CSP algorithm and (b) shows electrode weights for filter obtained through transfer learning based CSP algorithm	102
5.1	Average classification accuracies according to the number of training data available for the target subject.	108
5.2	Box-plot of all the subjects at 20 training trials in dataset IVa. . .	110
5.3	Average classification accuracies according to the number of training data available for the target subject.	111
5.4	Box-plot of all the subjects at 20 training trials in dataset IIIa. . .	112
5.5	Average classification accuracies according to the number of training data available for the target subject.	113
5.6	Average accuracies of all subjects in Dataset IIIa and Dataset IIa, respectively.	114
5.7	Box-plot of all the subjects at 20 training trials in Dataset IIa. . .	115
5.8	Average time in training algorithm on Single Subject in different datasets	117

Abstract

Motor imagery (MI) based Electroencephalogram (EEG) Brain-computer interface (BCI) detects neural activity generated due to kinesthetic imagination of limbs from brain scalp and translate it into control commands for external devices. MI-BCIs are indeed very promising for people suffering from neuromuscular disorder, but still lack adoption as access modalities outside laboratories. The main reason that prevents EEG based MI-BCIs from being widely used is there long calibration time. Due to considerable inter-subject/inter-session and intra-session variations, a large number of training trials are collected to calibrate systems at the beginning of each MI-BCI session. This time consuming calibration is required to achieve good performance with the BCI system but causes fatigue to user and leaves less time for online BCI interactions. This thesis focuses on developing reliable signal processing and classification pipeline that reduce MI-BCI calibration time while keeping accuracy in an acceptable range.

In the first part of the study, we have provided an extensive review of current state of art in designing a EEG based MI-BCI system. In doing so, I have created an architectural framework which brings together interdisciplinary concepts under a unified umbrella. We used this framework to identify key signal processing, features extraction and learning algorithms and their limitation that must be taken into consideration while designing novel pipeline for reducing calibration in MI-BCI. This architecture is also useful to understand current issues in BCI and to visualize the gaps to be filled by future studies in order to further improve BCI usability. In the second part of the study, we address long calibration issue in MI-BCI under two scenarios. First, when there is only few training trials from new subject (user) is available and no training data from previous sessions or other users is available. Second, reducing (inter-subjects/sessions non-satationarity) calibration time of new

subject when there is previous sessions or other subjects data is available along with few trials from new subject.

In order to contribute to the progress of reducing calibration in MI-BCI, we proposed novel signal processing and classification pipeline that uses spatial, spectral, temporal and geometrical properties of subject's trial from EEG signals and achieve acceptable performance under reduce calibration setting.

This page is intentionally left blank.

Acknowledgements

First and foremost, praises and thanks to the God. During the tough but worthwhile journey of my PhD, there are countless people for me to say thanks. First of all, I would like to express my deep and sincere gratitude to my primary supervisor Professor Hans W. Guesgen for giving me the opportunity to do research and providing his patient guidance and valuable support throughout my study. It was a great privilege and honor to work and study under his guidance. I am very grateful to my co-supervisors Dr. Sunil Lal for encouraging, and giving me freedom in my research. I would also like to thank him for his friendship, empathy, and great sense of humor. I must admit that I have learned a lot about research and life in his valuable company. I am extending my heartfelt thanks to his wife, family for being so nice and helpful to me throughout this PhD Journey. I am sure that the end of my doctoral journey will not be the end of the relationship with my supervisors.

I am extremely grateful to my parents (Jagtar Singh and Lakhbir Kaur) for their love, prayers, caring and sacrifices for educating and preparing me for my future. Also I express my thanks to my wife Dr. Arminder Kaur, brother Jagraj Singh, sister Rupinderjit Kaur and brother-in-law Mandeep Singh for their strong support and source of inspiration for me. While I was busy doing my doctoral research, they looked after our parents. My Special thanks goes to my friends Dr. Van Nguyen, Dr. Debashree Roy, Dr. Rachel Blagojevic and Sharnjit Singh for the keen interest shown to complete this thesis successfully. I would like to say thanks to my friends and research colleagues, Achinthya Perera, Akashdeep Singh, Saad Ghafoor, Chanjief Chandrakumar, Ali Abdul Hussain, Parmavir Singh, Manveet Singh, Manroop Kaur, Ramanjot Singh, Dr. Athar Imtiaz, Dr. Jaspreet, Dr. Preet and his family for their constant encouragement and support to make my life in New Zealand nicer and more enjoyable. I acknowledge the editors, anonymous referees and examiners

(Professor Stephen Marsland, Professor Marco Congedo and Dr. Kuda Dube)
for carefully reading, giving feedback and suggestions on five presented works
here to help me improve the thesis substantially.

Chapter 1

Introduction

People suffering from neuromuscular disorders lose control of their voluntary muscles. This leads to a condition known as locked-in syndrome, in which the affected person is unable to communicate or control ordinary things in daily life. For such people, communication with small motion-based systems such as eye-tracking systems is not suitable. Brain-computer interface (BCI) is an alternative way to communicate and control objects without using natural muscle channels. BCI allows direct communication between neural activity and external devices [1].

BCI systems fall into invasive and non-invasive categories based on the type of brain imaging technique applied to acquire brain activity. Invasive BCI acquired neural activity is recorded under the skull through surgery, where as a non-invasive approach acquires brain activity from the scalp [2]. Electroencephalography (EEG) is one such non-invasive technique that measures the projection of the cortical source (usually pyramidal cells) in the scalp using an EEG sensor [3]. EEG sensors quantify neural activity as a multivariate time series of potential difference between electrodes [4]. BCIs broadly use two types of paradigms namely exogenous or endogenous, to generate neural activity that is measured by EEG sensors. An exogenous paradigm correspond to brain activity generated due to external stimuli. Examples of EEG-based exogenous paradigms include control signals like visual-evoked potentials (VEP), steady-state visual evoked potentials (SSVEP) and P300 [2]. Essentially, exogenous BCI paradigms requires very little training to generate a control signal although the user needs to be focused on the presentation (for external stimulus). This causes fatigue to the users and is not

suitable for people who cannot control eye movements. Furthermore, the performance of the BCI system decreases with infrequent stimulus presentation [4]. In contrast, endogenous paradigms corresponds to voluntarily generated brain activity. Examples of EEG-based endogenous BCI paradigms include control signals like slow cortical potentials (SCPs), sensorimotor rhythms (SMRs) and other brain rhythms. Based on the application and user requirements, different paradigms are employed to control external devices and applications. One of the most widely used endogenous paradigms is based on SMRs control signals and is also referred to as motor imagery (MI)-BCI.

Motor imagery (MI) is a cognitive technique which involves mentally rehearsing the motor activity without performing it when a visual cue is displayed on a screen. This mental imitation generates a similar neural activity pattern at the sensorimotor area of the brain as if the subject is actually performing the motor action. More specifically, imagination of the movements corresponds to the increase or decrease in EEG signal power in specific frequency bands and locations [5]. Different MI tasks (such as imagining left/right-hand movement) produce different temporal, spectral, and spatial patterns. A common method for detecting MI tasks from a multivariate time series is to obtain the covariance matrix from preprocessed data in a specific frequency band (usually between 8 and 30 Hz) and the time period (usually after the cue). A covariance matrix of the MI task contains the variance of the signal at each EEG sensor (or electrode) at the diagonal position and the covariance between all pairs of EEG electrodes at the position of the non-diagonal elements [6]. As EEG sensors measure the projection of cortical sources at the scalp, therefore, with a large number of electrodes with a high sampling rate we can notice cognitive changes in real time but at the cost of high dimensionality. Subsequently, BCI pipeline comprised of signal processing and classification algorithms extract the spatiotemporal information contained in the covariance matrices for feature extraction and classification of the MI task. MI-BCI operates in two phases: (a) calibration phase and (b) online phase. In the calibration phase, the user learns to generate neural activity for different MI tasks, and the BCI learns to assign the neural activity for the control signal. To do this, the user must undertake MI training in which they performs MI tasks, such as left- or right-hand imagination, etc., and the BCI trains the signal processing and classification pipeline to discriminate between the neural activity generated by the user in

different MI tasks. Once the system is calibrated, it works in the online phase in which BCI recognizes MI task-generated neural activity in real time and translates it into control commands for external devices (or applications). This whole MI-BCI architecture is summarized in Figure 1.1.

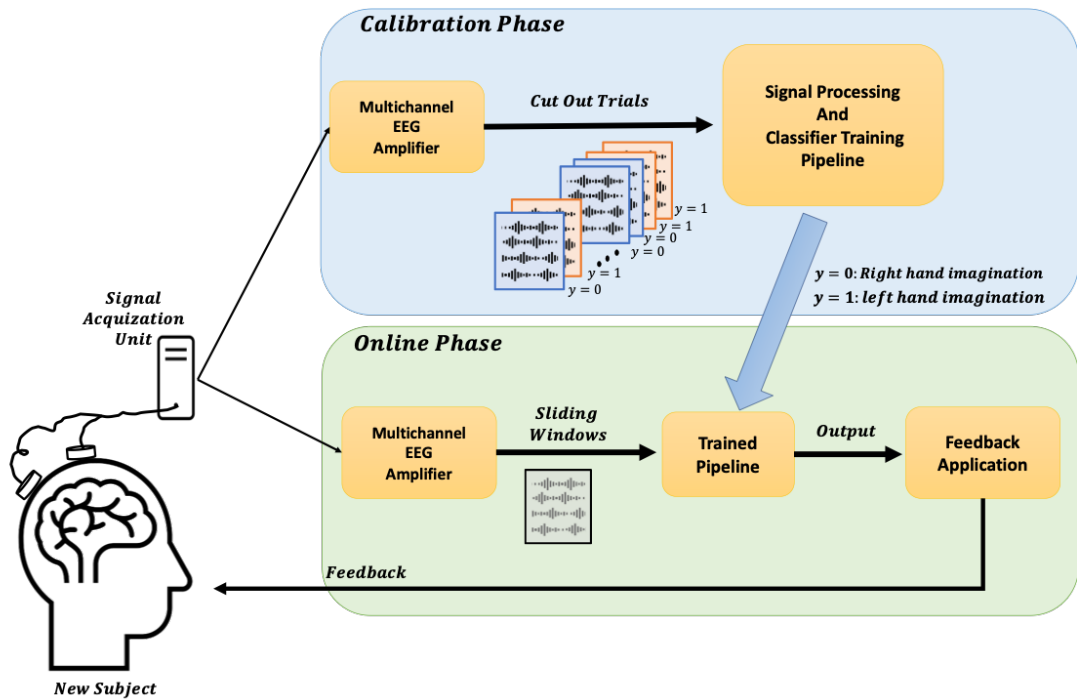


Figure 1.1: Block diagram of general online MI-BCI

Because EEG-based MI-BCIs enable computer control without voluntary muscle movements, they promise to revolutionize many application areas, especially to enable severely motor-impaired people to control assistive technologies, e.g. wheelchairs [7] or smart typing systems [8], as rehabilitation devices for stroke patients [9], as gaming input devices [10], or to develop adaptive human-computer interfaces that can respond to the user's mental states [11]. All these applications show the importance of developing a BCI system that can be used in the real world. However, even after more than thirty years of active BCI research, most of the EEG-based MI-BCI applications are still limited to laboratories. There are many challenges that need to be addressed to make MI-BCI usable in everyday life outside the laboratory environment.

1.1 Problem Statement & Research Contribution

MI-BCI offers a self-placed way that is close to natural and intuitive control. However, there are still many limitations that need to be addressed in order to develop reliable and accurate MI-BCI applications for daily practise. One of the major limitations of MI-BCI is its long calibration time. At the beginning of each MI-BCI session, a large number of training trials are carried out to calibrate parameters (for signal processing and classification pipeline) of the BCI system depending on the target user. Typically, this calibration phase can take up to 20 to 30 minutes for each new session. Blankertz et al. [12] in their study found that 40 trial per class are necessary to obtain a reasonable performance in BCI system. This is very time consuming, causes fatigue, and leaves less time for online BCI interactions.

There are a few possible reasons for this long calibration time. First, EEG signals are non-stationary. This non-stationarity could be caused by many factors such as variation in users' mental and psychological states variations, fatigue and miss-concentration; also it may also be affected by variance in measurements [13]. Because of this non-stationarity, the characteristics of neural activity vary between sessions/between subjects and within sessions. Therefore, the classifier usually performs poorly in new session data if trained using the features extracted from data of the previous sessions recorded on a different day. Thus, a large number of training trials before every session ensures that the BCI can reliably and accurately decode a new user's neural activity in the context of different MI tasks. Second, EEG signals are high dimensional and very noisy. Because of this, it is hard to estimate probability distributions of the features, especially few trials with high-dimensional signals are available for training [13]. High dimensional, noisy EEG signals get adversely impacted by outliers. Third, is the uniqueness of brain patterns for every person. Thus, it is not straightforward to calibrate a BCI model for a new subject from EEG data collected from previous subjects or sessions.

These challenges could be considered at different levels, e.g. at the human level by developing more advanced and successful user training techniques; or at the signal processing level, to build more robust approaches which could be calibrated with the least possible data [13]. Our main aim is to improve the

usability of BCI as a future technology by reducing such long calibration sessions by building more robust signal processing and classification pipeline. To address the problem of long calibration time, one popular approach is regularization [14] that mitigate the issue of signal processing and classification pipeline that rely on reliable covariance matrices that represent EEG activity during MI. However, if few trials are available then data does not reflect most of the variability occur during BCI use and leads to covariance matrices may be poorly estimated. Another popular approach is user-to-user transfer that uses data from other users to improve calibration for the target user for which few training data is available [15]. Semi-supervised learning another approach that uses labelled and unlabelled data to calibrate signal processing and classification pipeline under small training setting [16]. Another way to reduce calibration time is to use a-priori information about which features or channels are likely to be useful, in order to guide the optimization of the BCI pipeline [16]. All the above approaches fail to notice a very distinctive characteristic of data (covariance matrices): their structure (geometry), or more specific, the manifold space and the interrelation across the manifold space [17]. Data treatment based on the concept of manifolds have been proved to be more effective and adopted in many applications. Recently, geometry-aware algorithms have gained more attention in the field of BCI as they have shown that understanding the intrinsic geometry of the data provides the advantage of using geometric features along with spatial and temporal information in covariance matrices, leading to better classification of MI tasks under small training set [18]. However, performance of popular geometry-aware algorithms such as Riemannian Geometry based classification declines as the size of covariance matrices grows. As high-dimensional covariance matrices are ill conditioned with respect to inversion, jeopardizing the numerical stability of all Riemannian geometry manipulations, which are based on spectral functions of the eigenvalues such as the logarithm, inverse, etc. [6, 19]. This curse of dimensionality inevitably limits Riemannian Geometry based classification performance.

The main motivation of this thesis is providing novel BCI signal processing pipeline that combine best of both Euclidean and geometry-aware approaches to such that they mitigate issues of dimensionality as well as lead to better BCI system with less calibration time and improved accuracy. As there is a trade-off between calibration time and performance of the system, my goal is to reduce

this calibration time as much as possible without losing performance (classification accuracy) and even with improving it. Moreover, while reducing calibration, we aim to achieve high classification accuracy which is comparable to traditional state of the art methods. To achieve this aim, two main challenges needed to be addressed based on the available training data from the current user, and previous sessions or users. First, reducing intra-session non stationarity, i.e. when there is no training data available from previous sessions or other users is available. Second, reducing (inter-subjects/sessions non-satationarity) the calibration time of a new subject when there is previous sessions or other subject's data are available. Figure 1.2 summarizes the aim and challenges that have been addressed in this research.

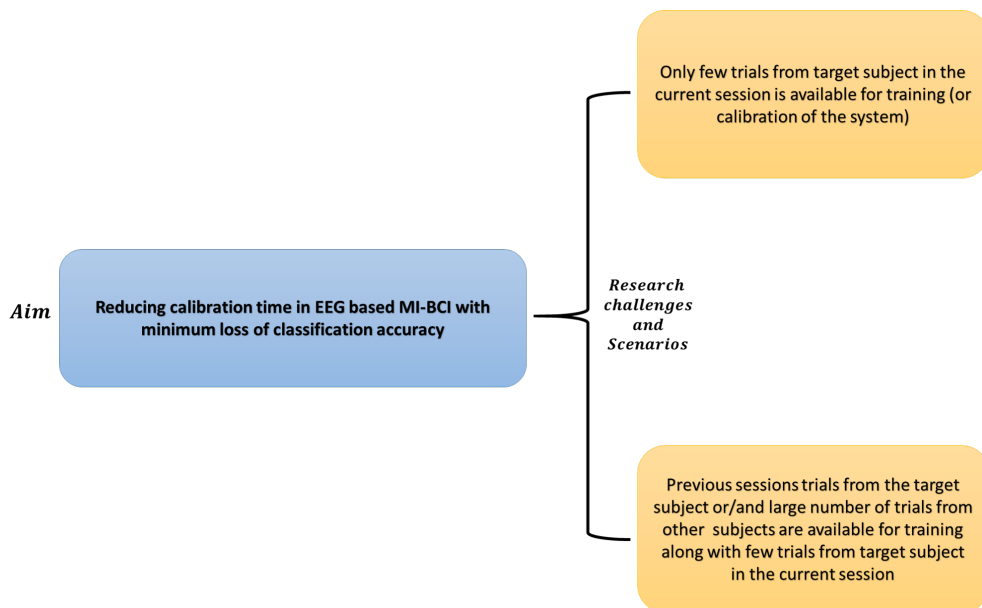


Figure 1.2: The aim and challenges addressed in this thesis

In order to contribute to the progress of reducing calibration in MI-BCI, we will be setting off with the following objectives:

- Developing an architectural framework for MI-BCI which assists in clear and methodological visualization of all the BCI system components and their possible interactions. This architecture will help in identifying key learning algorithms and any limitations that must be taken into consideration while designing pipeline for reducing calibration in MI-BCI.

- Developing a novel signal processing and classification pipeline that utilize most discriminative characteristics of a new subject's training trials and is able to improve MI-BCI performance when only a few training samples are available for the new subject.
- Developing a novel signal processing and classification pipeline that explore and utilize common information across subjects and sessions, to improve feature extraction and classification of MI tasks for new subjects with limited training trials .

The developed pipelines are evaluated using a classical offline training-test framework where a bootstrap technique systematically reduces the number of trials to simulate reduction of calibration time. We check our proposed pipelines performance with respect to the standard BCI pipeline. Moreover, their performance is evaluated across different groups of users trained with different number of electrodes to analyze performance under different dimension of covariance matrices. The advantages and disadvantages of these pipelines are discussed in terms of accuracy under different training set and computational time. As per Blankertz et al. [12] study 40 trials per class are necessary to obtain a reasonable BCI performance. Thus, performance of pipelines will be compared with BCI pipelines calibrated with 40 trials per class to measure success of achieving the aim.

1.2 Structure of the Thesis

The rest of this thesis is organised as follows.

- **Chapter 2** provides a comprehensive review of the electroencephalogram-based MI-BCI. This review presents an architectural framework that describes the current state of the art in different stages of MI-BCI. Moreover, this framework helps in finding answers to following research questions: (a) what are the most useful methods (e.g., spectral filtering , temporal filtering, spatial filtering etc.) for signal processing components in the pipeline? (b) what are the best feature extraction and classification techniques and their possible interaction? This architectural framework is also useful for mapping the literature onto each of the MI-BCI components in order to understand

current limitations that must be taken into consideration while designing novel pipeline for reducing calibration in MI-BCI. In addition, literature review shows recent studies focus on reducing calibration time. Apart from the calibration issue, we visualized the gaps to be filled by future studies in order to further improve BCI usability.

- **Chapter 3** introduces the subject-specific learning framework that uses covariance matrices as a feature and Riemannian distance as a pattern recognition metric for classification. The performance of a Riemannian metric based classifier degrades when the number of training samples are low and the size of the covariance matrices is large. To address this, the approach described in Chapter 3 uses a priori information about channels that are likely to be useful, in order to guide the optimization of the spatial filters. Optimized spatial filters leads to transforming covariance matrices into lower dimensions as well as maximizing the ratio of variance between MI classes. A Riemannian geometry based classifier is employed to classify these transformed covariance matrices.
- **Chapter 4** presents a novel transfer learning approach that reduces calibration time using spatio-temporal features and Riemannian metric based classification. The classification performance of Riemannian metric based classifiers declines under high-dimensional covariance matrices under the small training trials setting. To address this issue, Chapter 4 approaches the same problem by optimizing the spatial filter using MI trials from other subjects or previous sessions data from the target subject. In the same vein as Chapter 3, optimized spatial filters transform covariance matrices into lower dimension and classify them using Riemannian based classifiers. This improved the performance of MI-BCI system under small sample setting.
- **Chapter 5** introduces a classical offline training-test framework that uses a bootstrap technique to systematically reduce the number of training trials and check the performance of our proposed pipelines with respect to the standard Riemannian geometry-based BCI pipeline. Moreover, we statistically verify our results and discuss the reasoning behind the performance of the proposed methods.

- **Chapter 6** This chapter concludes the thesis along with the discussion on some of the future directions that may originate from this research.

1.3 Research outcomes

This research has produced the following peer-reviewed publications.

Journal publications


- **Amardeep Singh**, Sunil Lal and Hans W. Guesgen. “Reduce calibration time in motor imagery using spatially regularized symmetric positives-definite matrices based classification”, *Sensors* 2019, 19(379), 2879.
- **Amardeep Singh**, Sunil Lal and Hans W. Guesgen. “Small Sample Motor Imagery Classification Using Regularized Riemannian Features” in *IEEE Access*, vol. 7, pp. 46858-46869, 2019.
- **Amardeep Singh**, Ali Abdul Hussain, Sunil Lal and Hans W. Guesgen. “A Comprehensive Review on Critical Issues and Possible Solutions of Motor Imagery Based Electroencephalography Brain Computer Interface” in *Sensors* 2021.

Conference publications

- **Amardeep Singh**, Sunil Lal and Hans W. Guesgen. “Architectural review of co-adaptive brain computer interface”, In *2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, 200-207, 2017.
- **Amardeep Singh**, Sunil Lal and Hans W. Guesgen. ”Motor Imagery Classification Based on Subject to Subject Transfer in Riemannian Manifold.” In *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*,1-6, 2019

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Amardeep Singh
Name/title of Primary Supervisor:	Prof. Hans W. Guesgen
In which chapter is the manuscript /published work: Chapter 2	
Please select one of the following three options:	
<input checked="" type="radio"/> The manuscript/published work is published or in press <ul style="list-style-type: none"> • Please provide the full reference of the Research Output: Amardeep Singh, Ali Abdul Hussain, Sunil Lal and Hans W. Guesgen. " A comprehensive Review on Critical Issues and Possible Solutions of Motor Imagery Based Electroencephalography Brain Computer Interface" in Sensors 2021 	
<input type="radio"/> The manuscript is currently under review for publication – please indicate: <ul style="list-style-type: none"> • The name of the journal: • The percentage of the manuscript/published work that was contributed by the candidate: 60.00 • Describe the contribution that the candidate has made to the manuscript/published work: Conceptualization, Methodology, Investigation, Formal analysis, Validation, Writing - Original Draft, Writing - Review & Editing, Visualization. 	
<input type="radio"/> It is intended that the manuscript will be published, but it has not yet been submitted to a journal	
Candidate's Signature:	Amardeep Singh <small>Digitally signed by Amardeep Singh Date: 2021.02.26 13:13:11 +13'00'</small>
Date:	26-Feb-2021
Primary Supervisor's Signature:	
Date:	26-Feb-2021

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

Chapter 2

A Review on Critical Issues and Possible Solutions of Motor Imagery based BCI

Quick Summary

Motor imagery (MI) based brain-computer interface (BCI) aims to provide a means of communication through the utilization of neural activity generated due to kinesthetic imagination of limbs. Every year a significant number of publications related to new improvements, challenges and breakthrough in MI-BCI are made. This chapter provides a comprehensive review of the electroencephalogram (EEG) based MI-BCI system. It describes the current state of the art in different stages of the MI-BCI pipeline. Although MI-BCI research has been going for many years, this technology is confined mostly to controlled lab environments. We discuss recent developments and critical algorithmic issues in MI based BCI for commercial deployment.

Related Paper

- **Amardeep Singh**, Ali Abdul Hussain, Sunil Lal and Hans W. Guesgen. “A Comprehensive Review on Critical Issues and Possible Solutions of Motor Imagery Based Electroencephalography Brain Computer Interface” in *Sensors*, 2021.

2.1 Introduction

Numerous people with serious motor disorders are unable to communicate properly if at all. This significantly impacts their quality of life and ability to live independently. In this respect, brain-computer interface (BCI) aims to provide a means of communication. BCIs translate the acquired neural activity into control commands for external devices [2]. Primarily, BCI systems can be cast into various categories based on interactions with a user interface and neuroimaging technique applied to capture neural activity. Based on users interaction with brain-computer interface, the EEG-BCI system is categorized into synchronous and asynchronous BCI. In the synchronous BCI system, brain activity is generated by the user, based on some cue or event taking place in the system at a certain time. This cue helps in differentiating between intentional neural activity for a control signal from unintentional neural activity in the brain [20]. On the other hand, asynchronous BCI works independently of a cue. The asynchronous BCI system also needs to differentiate between neural activity intentionally generated by a user from the unintentional neural activity [21].

Based on neuroimaging techniques, BCI systems fall into invasive and non-invasive categories. In an invasive BCI, neural activity is captured under the skull, thus requiring the surgery to plant the sensors in different parts of the brain. This results in a high-quality signal but prone to scar tissue build-up over time resulting in loss of signal [22]. Additionally, once implanted sensors cannot be moved to measure the other parts of the brain [23]. In contrast to this, non-invasive BCI captures brain activity from the surface of the skull. A signal acquired through non-invasive technologies has a low signal to noise ratio. Some examples of invasive neuroimaging techniques are electrocorticography (ECoG) and micro electrodes. Examples of non-invasive neuroimaging techniques are electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and functional near infrared (fNIR) [24]. All these methods work on different principles and provides different levels of portability, spatial and temporal resolution [25]. Among these brain imaging methods, an EEG is widely employed because of its easy of use, safety, high portability, relatively low cost and most importantly high temporal resolution. Electroencephalography (EEG) is one of the non-invasive and portable neuroimaging techniques that records electrical activity generated due to the

synchronized activity of cerebral neurons. Primarily, pyramidal neurons' activity contributes more to EEG recordings because of their very stable orientated electric field to the cortical surface [24]. This is due to the perpendicular orientation of pyramidal cells with respect to the cortical surface. As a result, the electrical field is projected stably towards the scalp in contrast to the other brain cells whose electrical field is very dispersed and cancels out [25]. The measured EEG signal is due to the complex firing pattern of billions of neurons in the brain. Owing to this pattern, the EEG signal is a combination of various rhythms which reflect certain cognitive states of the individual [25]. These rhythms have different properties like frequency, amplitude and shape etcetera. These properties depend upon individual, external stimulus and the internal state of the individual. Broadly, these rhythms are classified into various categories based on their frequency, amplitude, shape and spatial localization [24]. Furthermore, these rhythms are broadly categorized under six frequency bands: delta band (1 – 4 Hz), theta band (4 – 8 Hz), alpha band (8 – 12 Hz), mu band (8 – 12 Hz), beta band (13 – 25 Hz) and gamma band (> 25 Hz). EEG control signals can be categorized as evoked and spontaneous. An evoked signal corresponds to neural activity generated due to external stimuli. Examples of evoked control signals are steady-state visual-evoked potentials (SSVEP), visual-evoked potentials (VEP), and P300 [22]. On the other hand, a spontaneous control signal is due to voluntarily neural activity without any aid of external stimulus. Slow cortical potentials (SCPs) and sensorimotor rhythms (SMRs) are such control signals [22]. As mentioned above, an evoked control signal requires external stimulus, thus the user needs to focus on presentation to generate neural activity. This continuous focus causes fatigue in users. Although much less training is required to generate evoked control signals. Spontaneous control signals offer natural control over neural activity but require long training to master self regulation of brain rhythms. To do so, different cognitive tasks are employed to generate spontaneous control signals.

One of the most widely used cognitive tasks is motor imagery (MI) which corresponds to sensorimotor rhythms (SMRs) as a control signal. Motor imagery has advantages for brain-computer interface in both synchronous and asynchronous mode. MI can be defined as the user sending a command to a system through the imagination of a kinesthetic movement of his/her limbs. For

example, a user moving a prosthetic arm by imagining his/her left/right hand moving. Imagination of movement creates a similar brain activity to that of an actual movement which decreases the percentage of power relative to a reference baseline in both the mu and beta frequencies over the sensorimotor cortex; this is known as event related desynchronization (ERD) [26]. Immediately after the user's imagination task, the user's brain activity can experience an event-related synchronization (ERS) which is the increase to the percentage of power relative to the reference baseline [26]. Since ERD/ERS are mixed with other brain activity created unintentionally by the user, such as involuntarily muscle movements and eye blinks, the signal to noise ratio (SNR) is low. The algorithm designed for MI-BCI must be able to differentiate between MI activity for control signal from other involuntary activity. In doing so, the MI-BCI pipeline consists of many stages like data acquisition, preprocessing, feature extraction and classification. Therefore, the objective of this manuscript is to review the MI based BCI system with regards to algorithms utilized at different stages of MI-BCI pipeline. This brief survey is structured under an architectural framework which helps in mapping literature to each component of the MI-BCI pipeline. In doing so, this article identifies critical research gaps which warrant further exploration along with current developments to mitigate these issues.

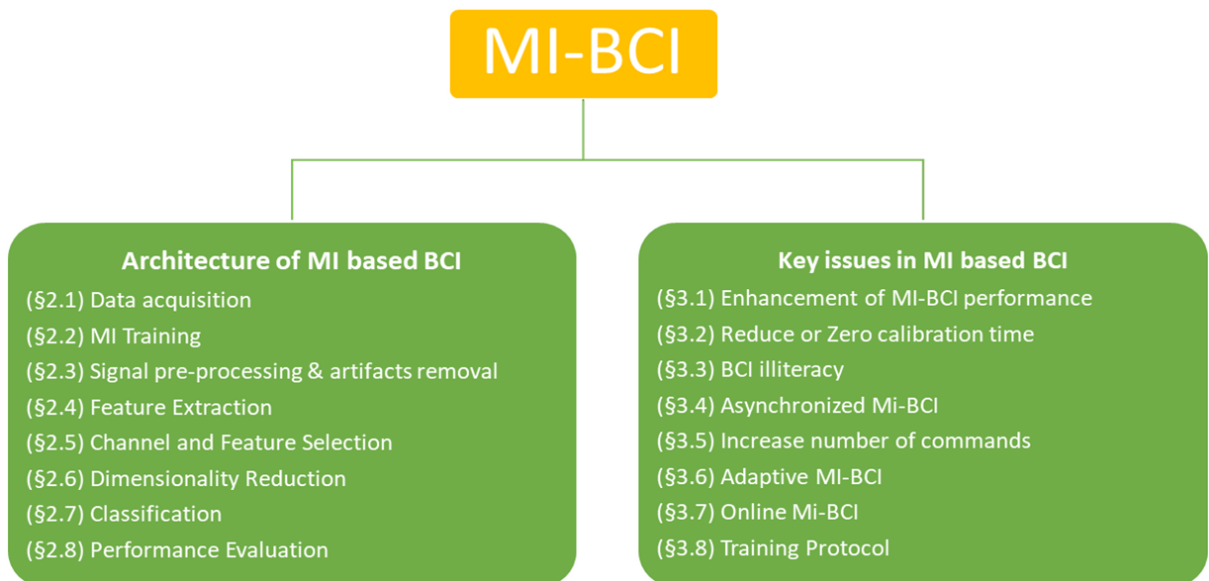


Figure 2.1: Breakdown of the chapter

Figure 2.1 breaks down the contents of the entire chapter. This review article is divided into two parts. The first part of this article introduces the Architecture of MI based BCI. More specifically, how the EEG signal is captured from the brain is described under subsection “data acquisition”. In subsection “MI training” we discuss how during the calibration phase the user acquires skills to modulate brain waves into control commands. The signal pre-processing subsection explains how unwanted artifacts are removed from the EEG signal to improve signal to noise ratio. Subsection on “feature extraction” discusses different approaches to extract information related to a motor imagery event in terms of features that are finally classified into control commands. Subsection on “channel and feature selection” and “dimensionality reduction” deals with issues related to finding optimal channels or features and reducing dimensionality of feature space in order to improve BCI performance. Subsection “classification” provides details of how features are classified into control commands. Lastly, subsection “performance evaluation” covers how to evaluate the performance of BCI. The last part of this article discusses key issues that need further exploration along with the current state of the art that address these research challenges.

2.2 Architecture of MI based BCI

We present a framework of MI-BCI pipeline encompassing all the components responsible for its working in Figure 2.2. In short, MI-BCI works in calibration and online mode respectively. During calibration mode, the user learns voluntary ERD/ERS regulation in the EEG signal and BCI learns ERD/ERS mapping through temporal, spectral and spatial characteristics of the user’s EEG signal. In online mode, the user’s characteristics are translated into a control signal for external application and feedback is given to the user. In framework, optional steps enclosed in yellow box such as channel selection, feature selection and dimensionality reduction. This framework is also helpful in mapping the literature to different components of the MI-BCI pipeline in order to understand the current research gaps.

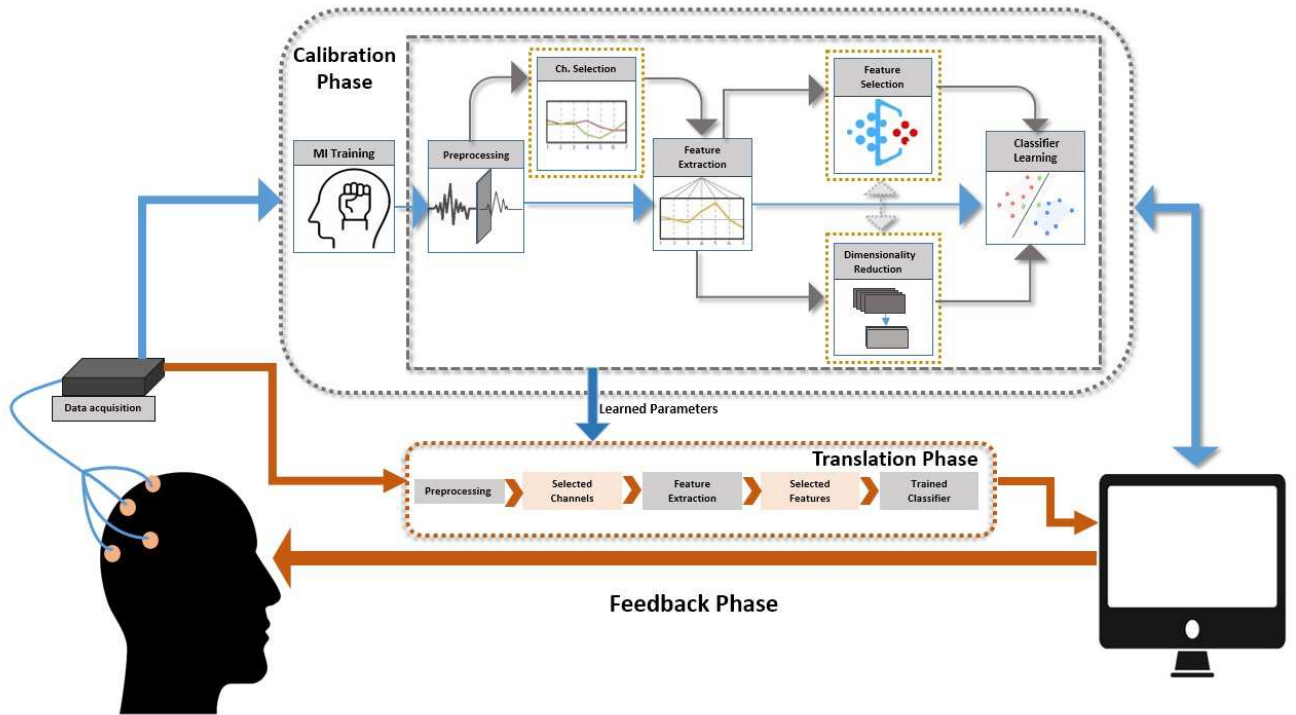


Figure 2.2: Block diagram showing the typical structure of MI-based BCI

2.2.1 Data Acquisition

The signal acquisition unit is represented by electrodes whether they are invasive or non-invasive. In non-invasive approach, usually electrodes are connected with the skin via conductive gel to create a stable electrical connection for a good signal. The combination of conductive gel and electrode attenuate the transmission of low frequencies but take very long time to setup. Another alternative is dry electrodes which make direct contact with skin without conductive gel. Dry electrodes are easy and faster to apply but more prone to motion artifacts [23]. EEG signals are usually acquired under unipolar and bipolar modes. In unipolar mode, a potential difference between all the electrodes with respect to one reference are acquired. Each electrode-reference pair form one EEG channel. On the contrary, in bipolar mode, the potential difference between two specified electrodes are acquired and each pair make a EEG channel [27]. To standardize positions and naming, electrodes are placed on scalp under international 10–20 standard. This help in reliable data

collection and consistency among different BCI sessions [28]. Figure 2.3 shows the international 10–20 electrodes' placement scheme from the side and top view of the head. Once the potential difference has been identified by the EEG electrodes, it is amplified and digitized in order to store it in a computer. This process can be expressed as taking one sample (discrete snapshots) of the continuous cognitive activity. This discrete snapshot (sample) depends upon the sampling rate of the acquisition device. For example, an EEG acquisition device with a sampling rate of 256 Hz can take 256 samples per second. High sampling rates and more EEG channels are used to increase temporal and spatial resolutions of an EEG acquisition device.

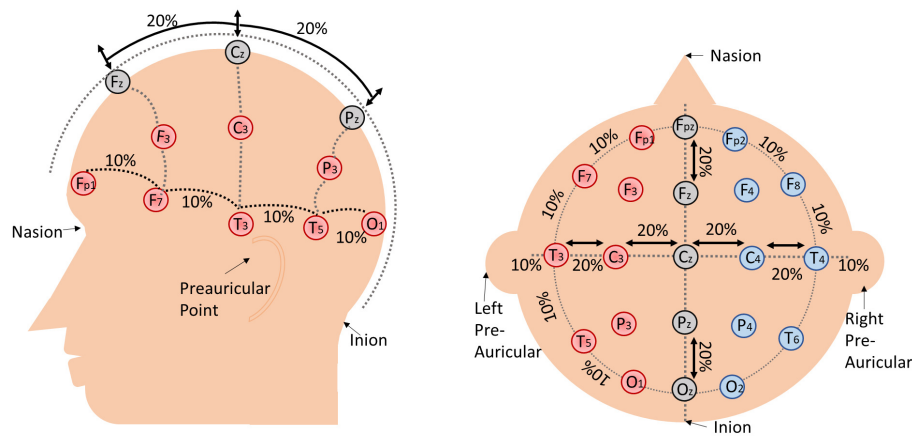


Figure 2.3: The international 10–20 standard electrode position system. The left image presents a left side view of the head with electrode positions, and the right image shows a top view of the head.

2.2.2 MI training

During calibration phase, the user learns how to modulate EEG signals with MI task pattern. Just as with any skill, MI training helps in acquiring the ability to produce a distinct and stable EEG pattern while performing the different MI tasks [29]. The Graz training paradigm is the standard training approach for motor imagery [26] [29]. The Graz approach is based on machine learning where the system adapts with the user's EEG pattern. During this training paradigm, the user is instructed through a cue to perform a motor imagery task such as left and right-hand imagination. EEG signals collected during different imagination

tasks are used to train the system differentiate between the MI-tasks from the EEG pattern. Once the system is trained, users are instructed to perform MI tasks, but this time feedback is provided to the user. This process is repeated multiple times over different sessions. Each session has further multiple runs of the Graz training protocol.

The trial time vary depending on scenario. Typically, one trial of graz training protocol lasts eight seconds as illustrated in figure 2.4. At the outset of each MI trail, that is $t = 0s$, a fixation cross is displayed to instruct the user that the trial has started. After a two-second break ($t = 2s$), a beep is used to prepare the user for the upcoming MI task. This 2s break acts as a baseline period to see the MI task pattern in the EEG signal in the upcoming MI task at $t = 3s$. After three seconds, an arrow appears on the screen indicating the MI task. For example, the arrow in the right direction means right hand motor imagery. During the initial training phase, no feedback is provided. After the system is calibrated, feedback is provided for four seconds. The direction of the feedback bar shows recognition of the MI pattern by system and the length of the bar represents confidence of the system in its recognition of the MI class pattern. Various other extensions of the Graz paradigm is proposed in the literature, mostly focusing on providing alternative MI instructions and feedback from the system. For example, the bar feedback is replaced by auditory [5] and tactile [30] feedback to reduce the workload on the visual channel. Similarly, virtual reality based games and environments are explored to provide MI instructions and feedback for training [31], [32].

2.2.3 Signal pre-processing and artifacts removal

Artifacts are nothing but unwanted activities during signal acquisition. They are comprised of an incorrect collection of signal or signals acquired from areas other than the cerebral origin of the scalp area. Generally, artifacts are classified into two major categories, termed as endogenous and exogenous artifacts. Endogenous artifacts are generated from the human body excluding the brain, and extra-physiologic artifacts are generated from external sources (i.e. sources from outside the human body) [25]. Some of the common endogenous and exogenous artifacts that accrue during EEG signal acquisition are bad electrode position, poor ground electrode, obstructions to electrode path (e.g. hair), eye blinks, electrode impedance, electromagnetic noise,

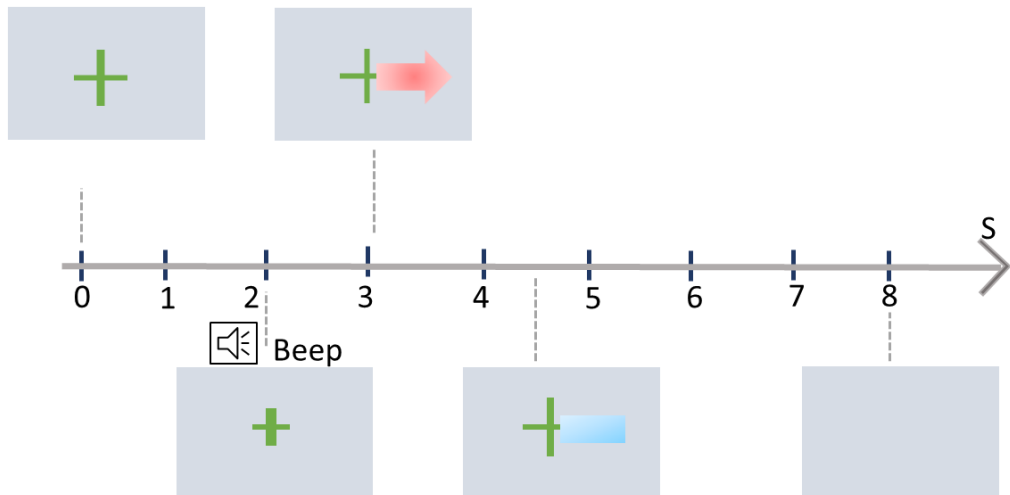


Figure 2.4: An illustration of one trial's timing in the Graz protocol

equipment problem, power line interference, ocular artifacts, cardiac artifacts and muscle disturbances [33]. The signal pre-processing block is responsible for the removal of such exogenous and endogenous artifacts from the EEG signal. MI-BCI systems mainly rely on a temporal and spatial filtering approach.

Temporal filtering is the most commonly used pre-processing approach for EEG signals. Temporal filters are usually low pass or band pass filters used to restrict signals in the frequency band where neurophysiological information relevant to the cognitive task lies. For MI, this usually means a Butterworth or Chebyshev bandpass filter of 8 – 30 Hz frequency. This bandpass filter keeps both the mu and beta frequency bands as they are known to be associated with motor-related tasks [26]. However, MI task-related information is also present in the spatial domain. Similar to temporal filters, spatial filters extract necessary spatial information associated with a motor-related task embedded in EEG signals.

A common average reference (CAR) is a spatial filter which removes the common components from all channels, leaving channels with only channel specific signals [34]. This is done by removing the mean of all k channels from each channel:

$$V_i^{\text{CAR}} = V_i^{\text{ER}} - \frac{1}{k} \sum_{j=1}^k V_j^{\text{ER}}. \quad (2.1)$$

where V_i^{CAR} is the filtered i^{th} channel (electrode) by the CAR method and V_i^{ER} is the potential difference between i^{th} electrode and reference electrode. An updated version of CAR is the Laplacian spatial filter. The Laplacian spatial filter aims to remove the common components of neighboring signals which increases the difference between channels [35]. The Laplacian spatial filter is calculated through the following equation:

$$V_i^{LAP} = V_i^{ER} - \sum_{j \in S_i}^k g_{ij} V_j^{ER} \quad (2.2)$$

$$g_{ij} = \frac{1/d_{ij}}{\sum_{j \in S_i}^k 1/d_{ij}} \quad (2.3)$$

where V_i^{LAP} is the i^{th} channel filtered by the Laplacian method, S_i is the set of neighboring electrodes to the i^{th} electrode, d_{ij} is the Euclidean distance between i^{th} and j^{th} electrode [35].

2.2.4 Feature Extraction

Feature extraction is about converting input data into a set of features. Features are distinctive properties of input patterns that help in differentiating between categories of input patterns. For MI-BCI, the feature extraction methods can be divided into six categories: (a) time domain methods that exploits temporal information embedded in the EEG signal; (b) spectral methods extract information embedded in the frequency domain of EEG signals; (c) time-frequency methods works together on information in the time and frequency domain; (d) spatial methods extract spatial information from EEG signals coming from multiple electrodes; (e) spatio-temporal methods works together with spatial and temporal information to extract features; (f) spatio-Spectral methods use spatial and spectral information of the multivariate EEG signals for feature extraction; (g) Riemannian Manifold methods which are essentially a sub category of spatio-temporal methods that exploits manifold properties of EEG data for feature extraction. Table 2.1 summarizes all the feature extraction methods discussed in the following subsections.

Time Domain Methods

An EEG is a non-stationary signal whose amplitude, phase and frequency changes with SMR modulations. Time domain methods investigate how the SMR modulation changes as a function of time [36]. Time domain methods work on each channel individually and extract temporal information related to the task. Extracted features from different channels are fused together to make a feature set for a single MI trial. In MI-BCI literature, statistical features like mean, root mean square (RMS), integrated EEG (power of signal), standard deviation, variance, skewness and kurtosis are vastly employed to classify MI tasks [37] [38]. Other alternative time domain methods based on variance of signal are Hjorth parameters. A Hjorth parameter measures power (activity), mean frequency (mobility) and change in frequency (complexity) of EEG signal [39]. Similarly, fractal dimension (FD) is non-linear method that measures EEG signals complexity [40]. Another typical time domain approach is auto-regressive (AR) modeling of EEG signal. The AR models signal from each channel as a weighted combination of its previous samples and AR coefficients are used as features. An extension of AR modelling is adaptive auto regressive modelling (AAR) and is also used for MI-BCI studies. Unlike AR, coefficients in AAR are not constant and, in fact, varies with time [39]. Information theory-based features like entropy, are also used in time domain to quantify complexity of the EEG signal [41]. Temporal domain entropy works with amplitude of EEG signal [42].

Another way of extracting temporal information is to represent the signal in terms of peaks (local maximum) and valley (local minimum) [43]. In this peak-valley representation, various features points are extracted between neighbouring peak and valley points. Using the peak-valley model, Yilmaz et al. [44] approximated EEG signal in 2D vector that contains cosine angle between transition points (peak or valley) and normalised the ratio of Euclidean distance in a left/right transition (peak or valley) points. In the same vein, Mendoza et al. [45] proposed a quaternion based signal analysis that represent multi-channel EEG signal in terms of their orientation and rotation then obtained statistical features for classification. Recently, EEG signal analysis based on graph theory and functional connectivity (FC) are employed in MI-BCI [46]. These methods take advantage of the functional communication between the brain regions during cognitive task like MI. In graph based

methods, the EEG data is represented through graph adjacency matrices corresponding to temporal correlations (correlation approaches used like Pearson or Correntropy) between different brain regions (electrodes). Features are extracted from this graph in terms of the graph node's importance such as centrality measure [34].

The advent of data driven approaches like deep learning has largely alleviated the need for hand crafted features. In these approaches, a raw or preprocessed EEG signal is passed through different convolution and pooling layers to extract temporal information [47]. In the same vein, Lawhern et al. [48] proposed EEGNET deep learning architecture that works with raw EEG signals. It starts with a temporal convolution layer to learn frequency filters (equivalent to preprocessing), another depth-wise convolution layer is used to learn frequency-specific spatial filters. Lastly, a combination of a depth-wise convolution along with point wise convolution are used to fuse features coming from previous layers for classification. Instead of using a raw or preprocessed signal, another approach is for the signal to be approximated then passed to a deep neural network model. One way of achieving this is a one dimension-Aggregate approximation (1d-AX) [49]. 1d-AX takes a signal from each channel in a single trial, normalize it and apply linear regression. These regression results are passed as features to the neural network.

Spectral domain methods

Spectral methods extract information from EEG signals in the frequency domain. Similar to the temporal method, statistical methods are also applied in the frequency domain. Samuel et al. [37] used statistical methods in both time and frequency domain to decode motor imagery. Most used spectral method is the power of EEG signals in specific frequency band. Usually, spectral power is calculated in mu (μ), beta (β), theta (θ) and delta (δ) frequency bands. This is done by decomposing the EEG signal into its frequency components at the chosen frequency band using Fast Fourier Transform (FFT) [50] [51]. Other frequency domain based method is Power Spectral Density (PSD). PSD is the measure of how the power of a signal is distributed over frequency. There are multiple methods of estimating it, such as Welch's averaged modified periodogram [52], Yule-Walker equation [53] or Lomb-Scargle periodogram [54]. Spectral entropy is another spectral feature that relies on PSD to quantify

information in the signal [55].

Time-Frequency Methods

Time-frequency (t-f) methods works simultaneously in both temporal and spectral domains to extract information in signal. One of the approaches used in t-f domain is the short Term Fourier Transform (STFT) which segments the signal into overlapping time frames on which FFT is applied by the fixed window function [50]. Another way to generate t-f spectra is through a wavelet transform [56] that decomposes the signal into into a family of wavelets. Unlike sinusoids, which are symmetric, smooth, and regular, wavelets can be either symmetric or asymmetric, sharp or smooth, regular or irregular. This captures the signal characteristics in the joint time-frequency domain. Another similar method in the t-f domain is empirical mode decomposition (EMD) [57]. However, instead of decomposing the signal into wavelets, it decomposes a signal $x(t)$ into simple oscillatory functions called Intrinsic Mode Functions (IMFs) [58]. IMFs are orthogonal representation of signals such that first IMF captures a higher frequency and subsequent IMFs capture lower frequencies in EEG signals. Table 2.1 sum up all the t-f methods.

Spatial Domain Methods

Unlike temporal methods which work with only one channel at a time, spatial domain methods work with multiple channels. Spatial methods try to extract features by finding a combination of channels. This can be achieved using blind source separation (BSS) [59]. BSS assumes that every single channel is the sum of clean EEG signals and several artifacts. Mathematically this looks like the following:

$$x(t) = As(t)$$

where $x(t)$ is the channels, $s(t)$ is the sources, and A is mixing matrix. They aim to find a matrix B that reverse the channels back into their original sources:

$$s(t) = Bx(t).$$

An example of a BSS algorithms are Cortical current density (CCD) [60] and independent component analysis (ICA) [61]. In the literature, most employed spatial method is common spatial pattern (CSP). CSP finds spatial filters such that the variance of the transformed data is maximized for one class while it is minimized for the other one [14]. The solution of the following objective function is the desired spatial filters.

$$J(w) = \frac{w^T C_1 w}{w^T C_2 w}$$

where w denotes the spatial filter, T denotes transpose operation and C_1, C_2 represent the estimated co-variance matrix of each MI class. The above equation can be solved using the Lagrange multiplier method. CSP is known to be highly sensitive to noise and performs poorly under small sample settings, thus a regularized version has been developed [14]. There are two ways to regularize the CSP algorithm (also known as regularized CSP), either by penalizing its objective function $J(w)$, or regularizing its inputs (covariance matrices) [14]. One can regularize the objective function by adding a penalty term to the denominator:

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

where $P(\cdot)$ is a penalty function, and α is a constant determined by the user ($\alpha = 0$ for CSP) [14]. While CSP inputs can be regularized by:

$$\begin{aligned} \tilde{C}_c &= (1 - \gamma)\bar{C}_c + \gamma I \\ \bar{C}_c &= (1 - \beta)s_t C_c + \beta G_c \end{aligned}$$

where s_t is a scalar, and G_c is a “generic” covariance matrix [14]. CSP performance becomes limited when the EEG signal is not filtered in the frequency range appropriate to subject. To address this issue, filter bank CSP (FBCSP) algorithm is proposed that passes the signal through multiple temporal filters and CSP energy features are computed from each band [62]. Finally, CSP features from sub-bands are fused together for classification. This results into a large number of features that limits the performance. To address this alternative method, sub-band common spatial pattern (SBCSP) is proposed that employs linear discriminant analysis (LDA) to reduce dimensionality.

Finding multiple sub-bands to compute CSP energy features increases the computational cost. To solve this, discriminant filter bank CSP (DFBCSP) is proposed that utilizes the fisher ratio (FR) to select most discriminant sub-bands from multiple overlapping sub-bands [63].

Spatio-temporal and Spatio-spectral methods

Spatio-temporal methods are algorithms that manipulate both time and space (channels) domains. The main spatio-temporal methods used in past MI-BCI studies are Riemannian Manifold-based methods (discussed in the next section). Other spatio-temporal methods are usually based on deep learning. One such approach is proposed by Echeverri et al. [64] that uses BSS algorithm to separate the input signal $x(t)$ from a single channel into an equal number of estimated source signals $\hat{s}(t)$. These source signals are sorted, based on a correlation between their spectral components. Lastly, continuous wavelet transform is applied to sorted source signals to achieve t-f spectra images that are further subjected to a convolution neural network (CNN) for classification. In the same vein, Li et. al. [65] proposed an end-to-end EEG decoding framework that extracts the spatial and temporal features from raw EEG signals. In a similar manner, Yang et al. [66] propose a combination long short-term memory network (LSTM) and convolutional neural network that concurrently learns temporal and spectral correlations from a raw EEG signal. In addition, they used discrete wavelet transformation decomposition to extract information in the spectral domain for classification of the MI task.

Like spatio-temporal methods, spatio-spectral methods extract information from spectral and spatial domains. Usually, temporal and spatial filters are learned in sequential (linear) order whereas, if they are learned simultaneously, a unified framework will be able to extract information from spatial and spectral domains. For instance, Wu et al. [67] employs a statistical learning theory to learn most discriminating temporal and spectral filters simultaneously. In the same vein, Suk and Lee [68] used a particle-filter algorithm and mutual information between feature vectors and class labels to learn spatio-spectral filters in a unified framework. Similarly, Zhang et al. [69] proposed a deep 3-D CNN network based on AlexNet that learns spatial and spectral EEG representation. Likewise, Bang et al. [70] proposed a method that generates 3D input feature matrix for the 3-D CNN network by stacking multiple-band

spatio-spectral feature maps from multivariate EEG signal.

Riemannian geometry based methods

Covariance matrices (CM) calculated from EEG signals are widely used in BCI algorithms. Let $X_i \in \mathbb{R}^{n \times t}$ be a MI trial, n being the number of electrodes and t the epoch duration expressed in number of samples. A trace normalized sample covariance matrices C_i for trial X is given by:

$$C_i = \frac{X_i X_i^T}{tr(X_i X_i^T)}$$

Covariance matrices lie in the space of symmetric positive definite (SPD) matrices $SPD(n) = \{\mathbf{C} = \mathbf{C}^T, u^T \mathbf{C} u > 0, \forall u \in \mathbb{R}^n, \forall u \neq 0\}$ which forms a Riemannian Manifold of $m = n(n + 1)/2$ dimensional [71]. Given two SPD matrices X and Y there could be several curves passing through them but the unique and shortest curve connecting two points (matrices) is called geodesic curve. The length of such geodesic curve from X and Y gives Riemannian distance (δ_r) [72]. It is given by $\delta_r(X, Y) = \|\text{Log}(X^{-1/2} Y X^{-1/2})\|_{\mathcal{F}}^2$. Riemannian distance poses three fundamental properties of metric space: positivity, symmetry and triangle inequality [6]. One of the most important properties of the Riemannian distance is the invariance of linear transformation. Riemannian geometry a differentiable manifold in which the tangent space at each point is a finite-dimensional Euclidean space. Riemannian distance computations in the manifold can be well approximated by Euclidean distance computations in the tangent space. Furthermore, logarithmic mapping can project points from manifold to tangent plane and inverse operation exponential mapping project an element of the tangent space back to the manifold. Figure 2.5 illustrate Riemannian Manifold and corresponding local tangent space at C . The logarithmic map $Log_C(\cdot)$ projects the data point (matrix) C_i into the tangent space. The Exponential map $Exp_C(\cdot)$ projects the element of the tangent space S_i back to the manifold.

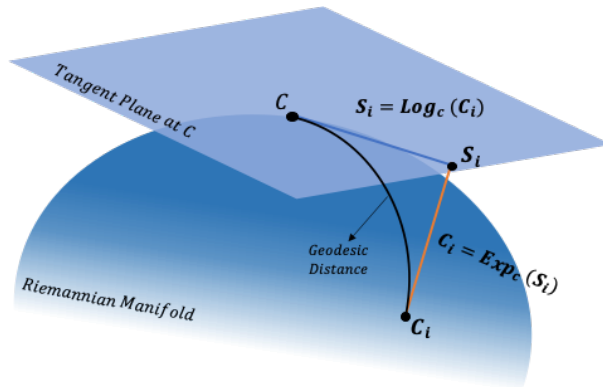


Figure 2.5: An illustration of the Riemannian Manifold displaying tangent space at point C

A summary of feature extraction methods				
Temporal Methods	Spectral Methods	Time-frequency Methods	Spatial Methods	Spatio-temporal Methods
Statistical Features [37] [38]				
Hjorth features [39]				
RMS [38]				
IEEG [38]	Band power [37]	STFT [50]		
Fractal Dimension [40]	Spectral entropy [42]	Wavelet transform [56]	CSP [14]	Covariance matrices [71]
Autoregressive modeling [39]	Spectral statistical features [37]	EMD [57]	BSS [60]	
Peak-Valley modeling [43] [44]				
Entropy [41] [42]				
Quaternion modeling [45]				

Table 2.1: This table provides a summary of the feature extraction methods

2.2.5 Channel and Feature Selection

EEG data is usually recorded through a large number of locations across the scalp. This provides a higher spatial resolution and benefits in identifying optimal locations (channels) relevant to BCI application or task. Here channel selections techniques significantly contribute to identify optimal channels for particular BCI application. Finding optimal channels not only reduces the computational cost of the system, but also reduces the subject's inconvenience due to the large number of channels. Thus, the main objective of channel selection methods is to identify optimal channels for the BCI task for improving classification accuracy and reducing computation time in BCI. The channels' selection problem is similar to that of feature selection where a subset of important features are selected from a vast number of features. Therefore,

channel selection techniques are derived from feature selection algorithm. Once the channels are selected, still we need to extract features for classification of the BCI task. Sometime we are even required to use the feature selection algorithm on selected channels to improve the performance of the system. Feature or channel selection algorithms have many stages. Firstly a candidate subset of features or channels are generated from an original set for evaluation purposes. This candidate subset is evaluated with respect to some selection criterion. This process is repeated for each candidate subset until a stopping criterion is reached. The selection criteria are what differentiates feature selection approaches. There are two stand-alone feature selection approaches filter approach and wrapper approach [73]. Sometimes combination of both are used to make hybrid approaches also known as embedded approach. Embedded method exploits strengths of both filter and wrapper approaches by combining them in feature selection process.

Filter Approach

Filter methods starts with all the features and selects the best subset of features based on some selection criteria. Usually this selection criteria is based on characteristics such as information gain, consistency, dependency, correlation, and distance measures [74]. The advantage of filter methods are their low computational cost and selection of features is independent of the learning algorithm (classifier). Some of the most widely employed filter methods are correlation criteria, and mutual information. Correlation detects linear dependence between variables x_i (features) and target Y (MI task classes). It is defined as:

$$R(i) = \frac{cov(x_i, Y)}{\sqrt{var(x_i)var(Y)}}$$

where $cov()$ is the covariance and $var()$ the variance. Mutual information (I) and its variant are widely used feature selection filter approaches in the MI-BCI literature. Mutual Information [75] $I(c_i; f)$ is the measure of the mutual dependence and uncertainty between two random variables: the features f and the classes c_i . This is measured by subtracting the uncertainty of the class $H(c_i)$ (also called initial uncertainty) from the uncertainty of the class given the

features $H(c_i|f)$:

$$I(c_i; f) = H(c_i) - H(c_i|f)$$

Both class uncertainty $H(c_i)$ and class uncertainty given the features $H(c_i|f)$ can be measured using Shannon's information theory entropy:

$$H(c_i) = - \sum_{i=1}^2 P(c_i) \log P(c_i)$$

$$H(c_i|f) = - \sum_{f=1}^{N_f} P(f) \left(\sum_{i=1}^2 P(c_i|f) \log P(c_i|f) \right)$$

where $P(c_i)$ is the probability density function of class c_i , and $P(c_i|f)$ is the conditional probability density function. When mutual Information is equal to zero $I(c_i|f) = 0$, the class c_i and the feature f are independent and as MI gets higher the more relevant feature f to class c_i . Thus, MI can be used to select the features by relevance.

Similarly, t-test [76] measures relevance of a feature to a class. It achieves this by examining mean $\mu_{i,j}$ and $\sigma_{i,j}$ variance of a feature f_j between class $i = \{1, 2\}$ through the following equation:

$$T(f_j) = \frac{|\mu_{1,j} - \mu_{2,j}|}{\sqrt{\frac{\sigma_{1,j}^2}{n_1} + \frac{\sigma_{2,j}^2}{n_2}}}$$

where n_i (n_1 and n_2) is the number of trials in class $i = \{1, 2\}$. This is then used to select a subset using the highest scoring features. Correlation based feature selection (CFS) [77] evaluate subsets of features based on the hypothesis that a good subset is the one that contains features highly correlated with the output classes and not correlated between them. This is computed using heuristic metric $Metric_S$ that divides productiveness of k feature subset S by the redundancy that exists in the k features that compose the subset S :

$$Metric_S = \frac{k \overline{r_{cf}}}{\sqrt{k + k(k-1) \overline{r_{ff}}}}$$

where $\overline{r_{cf}}$ is the mean of the class-feature correlation, $\overline{r_{ff}}$ is the mean of the inter-feature correlation.

F-score [78] is another feature selection approach that quantify the discriminative ability of variables (features) based on the following equation:

$$F - score_i = \frac{\sum_{k=1}^c (x_i^k - \bar{x}_i)^2}{\sum_{k=1}^c \left[\frac{1}{N_i^k} - 1 \sum_{j=1}^{N_i^k} (x_{ij}^k - \bar{x}_i^k) \right]} (i = 1, 2, \dots, n)$$

where c is the number of classes, n is the number of features, N_i^k number of samples of feature i in class k , x_{ij}^k is the j th training sample for feature i in class k . Features are ranked based on F-score such that higher F-score value corresponds to most discriminative feature.

Wrapper Approach

Wrapper approaches select a subset of features, present them as input to a classifier for training, observe the resulting performance and stop the search according to a stopping criterion or propose a new subset if the criterion is not satisfied [74]. Algorithms that fall under the wrapper approach are mainly searching and evolutionary algorithms. Searching algorithms start with an empty set and add features (remove features) until a maximum possible performance from the learning algorithm is reached. Usually, a searching algorithm's stopping criteria is until the number of features reaches a maximum size specified for the subset. On the other hand, evolutionary algorithms such as particle swarm optimization (PSO) [79], differential evolution (DE) [80, 81] and artificial bee colony (ABC) [82, 83] find an optimal feature subset by maximizing fitness function's performance. Wrapper methods find a more optimal feature subset compared to filter methods but the computational cost is very high, thus not suitable for very large data-sets.

2.2.6 Dimensionality Reduction

In contrast to feature selection techniques, dimensionality reduction methods tends to reduce the number of features in data, but they do so by creating new combinations (transformation) of features, whereas, feature selection methods, achieve this by including and excluding features from the original feature set. Mathematically, dimensionality reduction can be defined as the transformation of high dimensional data ($\mathbf{X} \in \mathbf{R}^D$) into a lower dimensional data ($\mathbf{Z} \in \mathbf{R}^d$),

where $\mathbf{d} \ll \mathbf{D}$. The dimensional reduction techniques can be categorized based on their objective function [84]. Those based on optimizing an convex (no local optima) objective function are convex techniques whereas techniques whose optimization function may have local optima are non-convex techniques. Furthermore, these techniques can be linear or non-linear based on the transform function used to map high dimensional to low dimension. The most used linear-convex technique is the Principal Component Analysis (PCA) which transforms data in a direction that maximizes the variance in the data set [85] [86]. Linear Discriminant Analysis (LDA) [87] is another linear dimensional reduction technique that finds a subspace that maximizes the distance between multiple classes. To do so, it uses class labels whereas PCA is an unsupervised technique. Independent Component Analysis (ICA) is another linear method found in EEG-BCI literature for dimensionality reduction that works on the principle that the EEG signal is a linear mixture of various sources and all sources are independent of each other [88]. To address the non-linearity in a data-points structure, PCA can be extended by embedding it with a kernel function (KPCA) [88]. KPCA first transforms data from the original space into kernel space using non-linear kernel transformation function, then PCA is applied in kernel space. Likewise, Multilayer Autoencoders (AE) is an unsupervised, non-convex and non-linear technique to reduce dimensionality of data [84]. AE [89] takes the original data and reconstructs into lower dimensional output using neural network. The drawback of the above discussed methods is that they do not consider geometry of data prior to transformation. Thus, manifold learning for dimensionality reduction has recently gained more attention in MI-BCI research.

Manifold learning-based methods recovers the original domain structure in reduced dimensional structure of data. Generally, these methods are non-linear and divided into global and local categories based on data matrix used for mapping high-dimensional to low-dimensional. Global methods used full EEG data covariance matrix and aims to retain global structure and do not take into account the distribution of neighbouring points [90]. Isometric feature mapping (Isomap) [91] [92] and diffusion maps [93], [91] are some of these global methods. In order to preserve global structure of manifold, isomap and diffusion maps aims to preserve pairwise geodesic distance and diffusion distance between data-points respectively. In contrast, local methods use a sparse matrix to solve

eigenproblems and their goal is to retain the local structure of the data. Locally, Linear Embedding [94] [95], Laplacian eigenmaps [92], local tangent space alignment (LTSA) [96] are some of these local methods. LLE assume manifold is linear locally and thus reconstruct data point from linear combination of its neighbouring points. Similar to LLE, Laplacian Eigenmaps [92] preserve the local structure by computing low-dimensional subspace in which the pairwise distance between a datapoint and its neighbours is minimal. Similarly, LTSA [96] map datapoints in high dimensional manifold to its local tangent space and there reconstruct the low dimensional representation of the manifold. All the above methods are designed for a general manifold thus approximate the geodesic distance without information of the specific manifold. The EEG covariance matrix lies in Riemannian manifold, therefore, more methods focused on dimensionality reduction are developed.

Considering the space of EEG matrices in Riemannian manifold, Xie et al. [96] proposed bilinear sub-manifold learning (BSML) that preserve the pairwise Riemannian geodesic distance between data points instead of approximating the geodesic distance. Likewise, Horev et al. [72] extended PCA in Riemannian manifold by finding a matrix $W \in \mathbb{R}^{n \times t}$ that maps the data from the current Riemannian space to a lower dimension Riemannian space while maximising variance. Along the same context, Davoudi et al. [97] proposed a nonlinear dimensionality reduction methods that preserves the distances to the local mean (DPLM) and takes into account geometry of the symmetric positive definite manifold. Tanaka et al. [98] proposed creating a graph that contains the location electrodes and their respective signals, and later applies the graph Fourier transform (GFT) to reduce the dimensions.

2.2.7 Classification

Classification is the mapping of the feature space ($\mathbf{Z} \in \mathbb{R}^d$) into the target space ($y \in TargetSpace$). This mapping is usually created by three things: a mapping function $f \in FunctionSpace$, an objective function $J(\mathbf{w})$, and a minimisation/maximisation algorithm (iterative or by direct calculation). Each of these has a role in the classifications process. The mapping function f determines both the space at that is being worked on and the approximation abilities of the classifier, whereas, the objective function $J(\mathbf{w})$ describes the problem that the classifier aims to solve. Finally, the

minimisation/maximisation algorithm aims at finding the best (optimal) mapping function $f : FeatureSpace \rightarrow TargetSpace$ that maps the data to its targets based on the objective function $J(\mathbf{w})$. Classification algorithms falls into Euclidean and Riemannian manifold based on how they interpret EEG feature space.

Euclidean space methods

Euclidean space \mathbb{R}^n is the space of all n -dimensional real number vectors. Most classification algorithms work in this space. One of such algorithms is linear discriminant analysis (LDA) [99]. LDA is an algorithm that creates a projection vector \mathbf{w} that maximises the distance between classes S_B and minimises the variance within a class S_W ($J_{LDA}(\mathbf{w}) = \max_{\mathbf{w} \in \mathbb{R}^n} \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}$). This is done by finding a generalized eigenvector of $S_B \mathbf{w} = \lambda S_W \mathbf{w}$. The classification is achieved by finding a threshold c that separates both classes, such as, if the dot product is below the threshold c , it belongs to class 1, otherwise it belongs to class 2. Duda et al. [100] described extension of LDA for multi-class problem.

Another classification algorithm that works in the Euclidean space is the support vector machine (SVM) [99]. we discuss later the extension of this algorithm into the Riemannian manifold. The objective of SVM is to find a hyperplane that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence. SVM works by projecting the data points $\{x_i\}_{i=1}^M$ onto a hyperplane \mathcal{H} $\{\phi(x_i)\}_{i=1}^M$. A plane in the hyperplane \mathcal{H} is then created by using quadratic programming $J_{SVM}(\alpha) = \max_{\alpha \in \mathbb{R}^m} \left(\sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}} \right)$ subject to $\alpha_i \geq 0$ and $\sum_i \alpha_i y_i = 0$ where $\langle, \rangle_{\mathcal{H}}$ is the dot product in hyperplane \mathcal{H} . This plane is then used to distinguish between classes $f_{SVM}(x) = \text{sgn}(b + \sum_i y_i \alpha_i k(x, x_i))$ where $k(x, x_i) = \langle \phi(x), \phi(x_i) \rangle_{\mathcal{H}}$ is the hyperplane kernel. Different kernels exist for hyperplanes, such as, the linear kernel $k(x, x_i) = x^T x_i$, the polynomial kernel $k(x, x_i) = (x^T x_i + c)^d$ where c is a constant, and the exponential kernel $k(x, x_i) = \exp(-\gamma \|x - x_i\|^2)$.

Another popular classifier is multilayer perceptron (MLP). MLP as the name suggests, $f(x) = \sum_i w_i^{(2)} \psi_i^{(1)}(\sum_j w_j^{(1)} x_j + b)$, is a multilayer algorithms with each layer containing perceptrons that can fire $\psi(\cdot)$. The layers are connected by weights w that are trained using a minimisation algorithm, such as, stochastic

gradient descent (SGD) or Adam algorithm. An extension to MLP is a convolutional neural network (CNN). It extends the MLP algorithm by adding a convolution and pooling layers. In the convolution layer, the high-level information is extracted by using a matrix kernel that is applied to each part of the data matrix. While in the pooling layer, it extracts dominant features and decreases the computational power required to process the data by finding the maximum or average of the sub-matrices.

Riemannian space methods

When the EEG data is taken and converted into sample covariance matrices (SCM), a Riemannian manifold is created. This Riemannian manifold differs for the Euclidean space. For example, a metric for measuring distances between two points in the Riemannian manifold is not equivalent to its Euclidean counterpart. The minimum distance to Riemannian mean (MDRM) is the most popular classification algorithm in the Riemannian manifold [71]. MDRM is the extension of the Euclidean classification algorithm in the Riemannian manifold. This algorithm take in the data in the form of sample covariance matrices (SCM) and calculates the Riemannian mean $\Pi(\mathbf{C}_1, \dots, \mathbf{C}_m) = \arg \min_{\mathbf{C} \in SPD(n)} \sum_{i=1}^m \delta_R^2(\mathbf{C}, \mathbf{C}_i)$ for each class using it to label data where $\delta_R(\mathbf{C}_1, \mathbf{C}_2) = \|\text{Log}(\mathbf{C}_1^{-1}\mathbf{C}_2)\|_F = \left[\sum_{i=1}^n \log^2 \lambda_i\right]^{1/2}$ is the Riemannian distance. The Riemannian mean equation could be thought of as its objective functions $J(\mathbf{P})$, while the algorithm used to find it could be conceptualised as a minimisation algorithm. MDRM has the following mapping function:

$$f_{MDRM}(\mathbf{C}_{m+1}) = \arg \min_{j \in \{1, 2, \dots\}} \delta_R(\mathbf{C}_{m+1}, \mathbf{C}_{\Omega_j})$$

where \mathbf{C}_{Ω_j} is the mean of class j . Similarly, Riemannian SVM (R-SVM) [71], is the natural extension of SVM algorithm into the Riemannian manifold. It uses the tangent space of a reference matrix C_{ref} as its hyper plane. This results in the following kernel:

$$k_R(\text{vect}(\mathbf{C}_i), \text{vect}(\mathbf{C}_j)); \mathbf{C}_{\text{ref}}) = \langle \phi(\mathbf{C}_i), \phi(\mathbf{C}_j) \rangle_{\mathbf{C}_{\text{ref}}}$$

where $\text{vect}(\mathbf{C}) = [C_{1,1}; \sqrt{2}C_{1,2}; C_{2,2}; \sqrt{2}C_{1,3}; \sqrt{2}C_{2,3}; C_{3,3}; \dots; C_{E,E}]$ is the vectorized form of a symmetric matrix, $\phi(\mathbf{C}) = \text{Log}_{\mathbf{C}_{\text{ref}}}(\mathbf{C})$ is the map from the

Riemannian manifold to the tangent space of C_{ref} , and $\langle A, B \rangle_C = \text{tr}(AC^{-1}BC^{-1})$ is the scalar product in the tangent space of C_{ref} .

	Mapping function	Objective function	Min/Max algorithm
LDA	$f(x) = \begin{cases} 1 & w^T x < c \\ -1 & \text{otherwise} \end{cases}$	$J(\mathbf{w}) = \max_{\mathbf{w} \in \mathbb{R}^n} \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}$	Eigen value solver
SVM R-SVM	$f(x) = \text{sgn}(b + \sum_i y_i \alpha_i k(x, x_i))$	$\max_{\alpha \in \mathbb{R}^m} \left(\sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \right)$	Quadratic Programming
MLP CNN	$f(x) = \sum_i w_i \psi_i(\cdot)$ $f(x) = \text{conv} + \text{pool} + \text{MLP}$	MSE, Cross entropy, Hinge	SGD, Adam
MDRM	$f(\mathbf{P}) = \arg \min_{j \in \{1,2,\dots\}} \delta_R(\mathbf{P}, \mathbf{P}_{\Omega_j})$	$J(P_\Omega) = \arg \min_{\mathbf{P}_\Omega \in \mathcal{P}(n)} \sum_i \delta_R^2(\mathbf{P}_\Omega, \mathbf{P}_i)$	Averaging approaches

Table 2.2: This table provides a summary of the classification methods describe in the section 2.2.7

2.2.8 Performance Evaluation

The general architecture of motor-imagery based brain-computer interface is well understood, yet numerous novel MI based interfaces and strategies are proposed to enhance the performance of MI-BCI. Thus, performance evaluation metrics play an important role to quantify diverse MI strategies. The most widely used performance evaluation is accuracy, which measures the performance of algorithm in terms of correctly predicting target class trials. Accuracy metrics is mostly employed where the number of trials for all classes are equal and there is no bias towards a particular target class [101]. In the case of unbalanced (unequal number of trials) classes, Cohen's kappa coefficient is employed [102]. Kappa coefficient equates an observed accuracy with respect to an expected accuracy (random chance). If kappa coefficient is 0, it means there is no correlation with the target class and predicted class where as kappa coefficient 1 denotes perfect classification. If the MI classification is biased towards one class then the confusion matrix (CM) is an important tool to quantify the performance of the system. Table 2.3 illustrates the confusion matrix for a multi-class problem. Metrics like sensitivity and specificity can be obtained from CM to identify percentages of correctly classified trials from each

MI class.

		Prediction			
		<i>Class₁</i>	<i>Class₂</i>	<i>Class_k</i>	<i>Class_n</i>
	<i>Class₁</i>	D_{11}	D_{12}	D_{1k}	D_{1n}
	<i>Class₂</i>	D_{21}	D_{22}	D_{2k}	D_{2n}
Target	<i>Class_k</i>	D_{k1}	D_{k2}	D_{kk}	D_{kn}
	—	—	—	—	—
	<i>Class_n</i>	D_{n1}	D_{n2}	D_{n3}	D_{nn}

Table 2.3: Multi class Confusion matrix

MI-BCI can be interpreted as a communication channel between user and machine, thus the information transfer rate (ITR) of each trial can be calculated to measure the bit-rate of the system. ITR can be obtained through CM (based on Accuracy) according to Wolpaw et al.'s [103] method as well as based on performance and distribution of each MI classes [104]. The metrics discussed above are summarized in Table 2.4 and applicable for both synchronized and self-placed (asynchronized) as well as multi-class MI-BCIs. As a BCI can be defined as an encoder-decoder system where the user encodes information in EEG signals and the BCI decodes it into commands. The above metrics evaluate how well the BCI decode user's MI task into commands but does not quantify how well the user modulates EEG patterns with MI tasks [105]. Therefore, there is room for improving performance metrics that measure user MI skills or a user's encoding capability.

To address some of the limitations mentioned above, Lotte and Jeunet [105] have proposed stability and distinctiveness metrics. Stability metrics measure how a stable MI EEG pattern is produced by a user. It is done by measuring the average distance between each MI task trial covariance matrix and mean covariance matrix for this MI task (left/right etc.). Distinct metrics measures distinctiveness between MI task EEG patterns. Mathematically, distinct metrics is defined as the ratio of the between class variance to the within class variance. Both stability and distinct metrics are defined in the Riemannian manifold as described in Table 2.4.

	Metrics	Two Class	Multi Class (N-Class)
BCI decoding capability	Accuracy	$\frac{D_{11}+D_{22}}{N_{all}}$	$\frac{\sum_{i=1}^N D_{ii}}{N_{all}}$, where $N_{all} = \sum_{i,j=1}^N D_{ij}$
	Kappa	$\frac{Accuracy - expected\ accuracy}{1 - expected\ accuracy}$, $expected\ accuracy(A_e) = \frac{\sum_{i=1}^N \frac{D_{ii}}{D_{i \cdot}}}{D_{\cdot i}}$	
	sensitivity	$\frac{D_{22}}{D_{21}+D_{22}}$	$\frac{D_{kk}}{N_k}$, where $N_k = \sum_{i=1}^N D_{k,i}$
	ITR_{Wolpaw}	$ITR_{Wolpaw} = \log_2 N + Acc. \log_2(Acc) + (1 - Acc) \log_2(\frac{1-Acc}{N-1})$	
User encoding capability	Stability	$\frac{1}{1+\sigma_{CA}}$	
	Distinct	$\frac{\sum_1^N \delta_r(\bar{C}^A_i, C^A)}{\sum_1^N \sigma_{C^A_i}}$	

Table 2.4: Summary of all the Metrics

2.3 Key issues in MI-BCI

MI based BCI still face multiple issues for it to be commercially usable. A usable MI based BCI should be a plug and play, self paced, highly responsive and consistent that everybody can use it. This could be achieved by solving the following challenges:

2.3.1 Enhancement of MI-BCI performance

A high performance MI-based BCI is important as it increases the responsiveness of the device and prevents user frustration hence improves the users experience. Improving the performance could be achieved by improving its pre-processing stage, channel selection stage, feature selection stage, dimensionality reduction stage, or a combination of them.

Enhancement of MI-BCI performance using preprocessing

Recent enhancements in the pre-processing step have revolved around two aspects: enhancing the incoming signal or enhancing the filtering of the signal. The former can be achieved by reconstructing the signal [106, 107], enhancing the spatial resolution [108], or adding artificial noise [109]. In Casals et al. [106],

they reconstructed corrupted EEG channels by using a tensor completion algorithm. The tensor completion algorithm applied a mask to this corrupted data to estimate it from observed EEG data. They found that this reconstructed the data of the corrupted channels and improved the classification performance in MI-BCI, whereas, Gaur et al. [107] used multivariate empirical mode decomposition (MEMD) to decompose the EEG signal into a set of intrinsic mode functions (IMFs). Based on a median frequency measure, a set of IMFs are selected to reconstruct EEG signals. CSP features are extracted from the reconstructed EEG signal for classification. One can enhance the spatial resolution of the EEG signal by using local activities estimation (LAE) method [108]. The LAE method estimates the recorded value of an EEG channel based on the weighted sum of local values of all EEG channels. The weights assigned to each channel for a weighted sum is based on the distance between channels. Similarly, enhancing the filtering of the signal can be achieved by automated filter (subject specific) tuning based on optimization algorithm like particle swarm optimization (PSO), artificial bee colony (ABC) and genetic algorithm (GA) [110]. Both Kim et al. [111] and Sun et al. [112] proposed filters that are aimed to remove artifacts. Kim et al. [111] removed ocular artifacts by using an adaptive filtering algorithm based on ICA. Sun et al. [112] removed EOG artifacts by a contralateral channel normalization model that aims at extracting EOG artifacts from the EEG signal while retaining MI-related neural potential through finding the weights of EOG artifact interference with the EEG recordings. The Hijorth parameter was then extracted from the enhanced EEG signal for classification. In contrast to above methods, Sampanna and Mitaim [109] have used PSO algorithm to search for optimal Gaussian noise intensity to be added in signals. This helps in achieving higher accuracy compared to a conventionally filtered EEG signal. The Signal that is reliable at run time is very important for online evaluation of MI-BCI. To address this, Sagha et al. [113] proposed a method that quantifies electrode reliability at run time. They proposed two metrics based on Mahalanobis distance and information theory to detect anomalous behaviour of EEG electrodes.

Enhancement of MI-BCI performance using channel selection

Channel selection can both remove redundant and non task relevant channels [114] and reduce power consumption of the device [115]. Removing channels can improve performance by reducing the search space [114], while reducing the power consumption can increase the longevity of a battery-based device [115]. Yang et al. [116] selected an optimal number of channels and time segments to extract features based on Fisher's discriminant analysis. They used the F score to measure discrimination power of time domain features obtained from different channels and different time segments. Jing et al. [117] selected high quality trials (free from artifacts) to find an optimal channel for a subject based on the "maximum projection on primary electrodes". These channels are used to calculate ICA filters for MI-BCI classification pipeline. This method has shown good improvement in classification accuracy even in session to session and subject to subject transfer MI-BCI scenarios. Park et al. [118] applied particle swarn optimization algorithm to find subject specific optimal number of electrodes. These electrodes' EEG data is further used for classification. Jin et al. [119] selected electrodes that contain more correlated information. To do this, they applied Z-score normalization to EEG signals from different channels, then computed pearson's coefficients to measure similarity between every pair of electrodes. From selected channels, RCSP features are extracted for SVM model based classification. This significantly improves the accuracy compared to traditional methods. Yu et al. [120] used Fly optimization algorithm(FOA) to select the best channel for subject and extracted CSP features from these channels for the classification. They also compared FOA performance with GA and PSO. Ramakrishnan and Satyanarayana [115] used large (64) and small (19) number of channels in data acquisition for training and testing phase respectively. They calculated inverse Karhunen Loeve Tranform (KLT) matrix from training trials. This inverse KLT matrix is used to reconstruct all the missing channels in the testing phase. Masood et al [121] employed various flavours of CSP algorithm [14] to obtain the spatial filter weights of each electrode. Based on maximal values of spatial pattern coefficients electrodes are selected to compute features for MI-CSP classification.

Enhancement of MI-BCI performance using feature selection

Similar to channel selection, feature selection improve performance by finding most optimal features. Like, Yang et al. [122] in their study decomposed EEG signals from C3,Cz and C4 channels into a series of overlapping time-frequency areas. They achieved this by cutting the filtered signals from filter bank of width 4 Hz and step 1 Hz (e.g. 8-12, 9-13,...,26-30) into multiple overlapping time segments. They used an F-score to select optimal time-frequency areas to extract features for MI-BCI classification. Rajan and Devassy [123] used a boosting approach that improved classification by a combination of feature vectors. Baboukani et al [124] used an Ant Colony Optimization technique to select a subset of features for SVM based classification of MI-BCI. Wang et al [125] divided all the electrodes in several sensor groups. From these sensor groups CSP features are extracted to calculate EDRs. These EDRs are fused together, based on information fusion to obtain discriminate features for ensemble classification. Liu et al. [126] proposed a feature selection method based on firefly algorithm and learning automata. These selected features are used to classify by a spectral regression discriminant analysis (SRDA) classifier. Kumar et al. [127] used mutual information technique to extract suitable features from CSP features from filter banks. Samanta et al. [128] used an auto encoder-based deep feature extraction technique to extract meaningful features from images of a brain connectivity matrix. The brain connectivity matrix is constructed based on mutual correlation between different electrodes.

Enhancement of MI-BCI performance using dimensionality reduction

Xie et al. [129] learned low dimensional embedding on the Riemannian manifold based on prior information of EEG channels. Where as, She et al. [130] extracted IMFs from EEG signals then employed Kernel spectral regression to reduce the dimension of IMFs. In doing so, they constructed a nearest neighbour graph to model the IMFs intrinsic structure. Özdenizci and Erdoğan [131] proposed information theory based linear and non-linear feature transformation approach to select optimal feature for multi-class MI-EEG BCI system. Pei et al. [89] used stacked auto-encoders on spectral features to reduce dimension and achieve high accuracy in a multi class asynchronous MI-BCI system. Razzak et al. [132] applied sparse PCA to reduce the dimensionality of features for SVM based classification. Horev et al. [72] extended the PCA to SPD manifold space

such that it preserved more variance in data while mapping SPD matrices to a lower dimension. Harandi et al. [133] proposed an algorithm that maintains the SPD matrices geometry while mapping it in a lower dimension. This is done by preserving the local structure's distance with respect to the local mean. In addition to it, this mapping minimize the geodesic distance in samples belonging to the same class as well as maximizes the geodesic distance between samples belonging to a different class. Davoudi et al. [97] adapted Harandi's geometry preserving the dimensionality reduction technique in an unsupervised manner. Similarly, Tanaka et al. [98] proposed graph fourier transform for reducing dimensionality of SPD matrices through Tangent space mapping. This method has shown improvement in performance for a small training dataset.

Enhancement of MI-BCI performance with combination of all

Li et al. [134] use the TPCT imaging method to fix the electrode positions and assign time-frequency feature values to each pixel in the MI-EEG image. This way promotes feature fusion from time, space and frequency domains respectively. These high dimensional images are fed to the modified VGG16 network [135]. Wang et al. [136] extracted a subset of channels from the motor imagery region. From these extracted channels a subject-specific time window and frequency band is obtained to extract CSP features for classification. Sadiq et al. [137] manually selected channels from the sensorymotor cortex area of the brain. The EEG signal from these selected channels is decomposed into ten IMFs using on adaptive empirical wavelet transform. The most sensitive mode out of ten is selected based on PSD and the Hilbert transform (HT) method extract instantaneous amplitude (IA) and instantaneous frequency (IF) from each channel. Statistical features are extracted from IF and IA components for classification. Selim et al. [138] used bio-inspired algorithm (attractor metagene (AM)) to select the optimal time interval and CSP features for classification. Furthermore, they used Bat optimization algorithm (BA)) to optimize SVM parameters to enhance the classifier's performance. Athif and Ren [139] proposed the waveCSP technique that used wavelet transform and CSP filtering technique to enhance the signal to noise ratio of the EEG signal and to obtain key features for classification. Li et al. [140] optimized the spatial filter by employing Fisher's ration in objective function. This not only avoids using regularization parameters but also selects optimal features for classification. Li et al [141] designed a spectral component

CSP algorithm that utilized ICA to extract relevant motor information from EEG amplitude features obtained from CSP. Liu et al. [142] proposed an adaptive boosting algorithm that selects the most suitable EEG channels and frequency band for CSP algorithm.

2.3.2 Reduce or Zero calibration time

Everyday a BCI user is required to go through a calibration phase for him/her to use BCI. This can be inconvenient, annoying, and frustrating. This section describes an on-going research solution to reduce the calibration phase or completely remove it. There are three categories of solutions: subject specific methods, transfer learning methods, and subject independent methods.

Subject-specific methods

Subject-specific methods for the reduction of calibration time aim mostly at extracting features more efficiently (i.e. with a small amount of training data). This can be achieved by particle swarm optimization based learning strategy to find optimal parameters for spiking neural model (SNM) (deep learning model) [143]. This method automatically adjusts the parameters, removes the need for manual tuning, and increases the efficiency of SNM. However, this requires very subject-specific optimization of parameters for best results [144]. Whereas, Zhao et al. [145] proposed the use of a framework that transforms EEG signals into three-dimensional space to preserve the temporal and spatial distribution of EEG signal and uses multi-branch 3D convolutional neural network to take advantage of temporal and spatial features in EEG signal. They showed that this approach significantly improves the accuracy under a small training dataset. Another approach of reducing calibration time is by a subject specific modification of the CSP algorithm. For example, Park and Chung [146] improved CSP by electing CSP features from good local channels rather than all channels. They selected good local channels based on variance ratio dispersion score (VRDS) and inter-class feature distance (ICFD). Furthermore, they extended this approach in Filter Bank CSP by selecting good local channels for each frequency band, whereas, Ma et al. [147] optimized SVM classifier's kernel and penalty parameters through a particle swarm optimization algorithm to obtain optimal CSP features. Furthermore, Costa et al. [148]

proposed an adaptive CSP algorithm to overcome the limitation of CSP in short calibration sessions. They iteratively update the coefficients of the CSP filters using a recursive least squares (RLS) approach. This algorithm can be enhanced based on right channel selection and training free BCI system by modifying the algorithm with unsupervised techniques. Kee et al. [41] proposed Renyi entropy as a new alternative feature extraction method for small sample setting MI-BCI. Their method outperforms conventional CSP and regularized CSP design in small training datasets. Lotte and Guan [14] proposed Weighted Tikhonov Regularization for CSP objective function that gives different penalties for different channels based on their degree of usefulness to classify a given mental state. They also extended the conventional CSP method for a small sample setting in [149] by penalizing the CSP objective function through prior information of EEG channels. Prior information of EEG channels was also used by Singh et al. [150] to obtain a smooth spatial filter to reduce the dimension of covariance matrices of trials under a small training set. They used MDRM for classification of covariance matrices. This approach had shown higher performance under a high dimensional small sample setting.

Transfer learning methods

An investigation on inter-session and inter-subject variabilities in multi-class MI-based BCI revealed the feasibility of developing calibration-free BCIs in subjects sharing common sensorimotor dynamics [151]. Based on this concept of using other subjects/sessions, transfer learning methods have been developed. Transfer learning methods aim to use other subjects data either to increase the amount of data that the classifier can be trained on or to regularize (prevent overfitting) the algorithm. The former can be seen in He and Wu [152], Hossain et al. [153], and Dai et al. [154]. He and Wu [152] used Euclidean-space alignment (EA) on top of CSP to enable transfer learning from other subjects. EA projects all subjects into a similar distribution using the Euclidean mean. Hossain et al. [153] extended FBCSP by adding selective informative instance transfer learning (SIITAL). The SIITAL trains the FBCSP with both source and target subjects by iteratively training the model and selecting the most relevant samples of the source subjects based on that model. Dai et al. [154] proposed unified cross-domain learning framework that uses the FBRCSP method [155] to extract features from source and target subjects. This is

achieved by ensemble classifiers that are trained on misclassified samples and contribute to the overall model based on their classification accuracy, while the latter can be seen in Azab et al. [156], Singh et al. [157, 158], and Jiao et al. [159]. Azab et al. [156] proposed a logistic regression-based transfer learning approach that assigns different weights to a previously recorded session or source subject to represent similarities between these sessions/subjects features distribution and the new subject features distribution. Based on Kullback-Leibler divergence (KL) metrics similar source/session feature space to target subject is chosen to obtain subject-specific common spatial patterns features for classification. Singh et al. [157, 158] proposed a framework that takes advantage of both Euclidean and Riemannian approaches. They used a Euclidean subject to subject transfer approach to obtain optimized spatial filter for the target subject and employed Riemannian geometry-based classification to take advantage of the geometry of covariance matrices. Park et al. [155] extended the FBCSP with regularization. They obtained an optimized spatial filter for each frequency band using information from other subjects trials. The CSP features from each frequency band are obtained and lastly, based on mutual information most discriminate CSP features are selected for classification. Jiao et al. [159] proposed Sparse Group Representation Model for reducing calibration time. In their work, they constructed a composite dictionary matrix with training samples from source subjects and target subject. A sparse representation-based model is then used to estimate the most compact representation of a target subject samples for classification by explicitly exploiting within-group sparse and group-wise sparse constraints in the dictionary matrix. The former has the advantage of being applicable to all trained subjects over the latter.

Subject independent methods

Subject-independent methods aim to eliminate the calibration stage, allowing the user to plug and play the BCI device. One way of achieving this is by projecting all different subjects/sessions' data into a unified space. Rodrigues et al. [160] proposed the Riemannian Procrustes Analysis as a projection based method. It transforms subject-specific data into a unified space by applying a sequence of geometrical transformations on their SCMs. These geometrical transformations aim to match the distribution of all subjects in high-dimensional space. These

geometrically transformed SCMs are then fed to the MDRM classification model to discriminate MI tasks. However, this method still requires the creation of the geometrical transformations based on the targets' session thus is not entirely calibration-free but pave way for fully subject independent MI-BCIs. Another way of achieving subject-independence is to create a universal map which can take in any subject data and output the command. Zhu et al. [161] proposed a deep learning framework for creating a universal neural network called separate channel CNN (SCCN). SCCN contains 3 blocks: CSP block, Encoder block, and recognition block. The CSP block was used to extract the temporal features from each channel. The encoder block then encodes those extracted features followed by a concatenation of the encoded features and feeding them into the recognition block for classification. Joadder et al. [162] also proposed a universal MI-BCI map that extracts sub-band energy, fractal dimension, Log Variance and Root Mean Square (RMS) features from spatial filtered EEG signal (CSP) for Linear Discriminant Analysis (LDA) classification model. They evaluated their design on a different time window after cue, different frequency band and different number of EEG channels and obtained good performance compared to existing subject-dependent methods. Although both Zhu et al. [161] and Joadder et al. [162] classifiers are subject-independent, the CSP extracted features are not. Zhao et al. [163] hypothesised that there exists a universal CSP that is subject-independent. They used a multi-subject multi-subset approach where they took each subject in the training data and randomly picked samples to create multiple subsets and calculated a CSP on each subset. This was followed by a fitness evaluation-based distance between these CSP vectors (density and distance between highly dense vectors). They also proposed a semi-supervised approach as a classifier, however, unlike the universal CSP, it required unlabelled target data. In the same vein, Kwon et al. [164] followed the same universal CSP concept. Unlike Zhao et al. [163], they trained only one CSP on all of the available source subject's data and, since they had a larger dataset, they assumed it would find the universal CSP. Mutual information and CNN was then used for a complete subject-independent algorithm.

2.3.3 BCI illiteracy

BCI illiteracy subject is defined as the subject who cannot achieve classification accuracy higher than 70% [29, 165–170]. BCI illiteracy indicates that the user is

unable to generate required oscillatory pattern during MI task. This leads to poor performance of MI-BCI. Some researchers focus on predicting whether a user falls under BCI illiterate category or not. This can help us to design better algorithm for decoding MI or design better training protocol to improve user skills. For instance, Ahn et al. [171] demonstrated that self-assessed motor imagery accuracy prediction has a positive correlation with actual performance. This can be valuable information to find BCI inefficiency in the user. While, Shu et al. [166] in their work proposed two physiological variables, that is, laterality index (LI) and cortical activation strength (CAS), to predict MI-BCI performance prior to clinical BCI usage. Their proposed predictors exhibited a linear correlation with BCI performance, whereas, Darvishi et al [172] proposed a simple reaction time (SRT) as the BCI performance predictor. SRT is a metrics that reflects the time required for a subject to respond to a defined stimulus. Their results indicate that SRT is correlated with BCI performance and BCI performance can be enhanced if the feedback interval is updated in accordance with the subject's SRT. In the same vein, Müller et al. [173] has theoretically shown that adaptation that is too fast may confuse the user while an adaptation that is too slow might not be able to track EEG variabilities due to learning. They created an online co-adaptation BCI system by ever-changing feedback according to the user and the system's learning. In the same vein, co-adaptive approach to address BCI illiteracy has also been proposed by Acqualagna et al. [167]. Their paradigm composed of two algorithms: a pre-trained subject independent classifier based on simple features, and a supervised subject optimized algorithm that can be modified to run in an unsupervised setting based. Acqualagna et al. approach is based on Vidaurre et al. [174] classification of users. Vidaurre et al. [174] in their study classify users in three categories: For category I users (Cat I), the classifier can be successfully trained and they gain good BCI control in the online feedback session. For Category II users (Cat II) the classifier can be successfully trained however, good performance cannot be achieved in the feedback phase. For Category III users (Cat III), successful training of the classifier is not achieved. In the same vein, Lee et al. [175] found that that a universal BCI illiterate user does not exist (i.e. all participants were able to control at least one type of BCI system). Their study paves way to design BCI system based on user's skill. Another way of addressing BCI illiteracy problem is to design novel solutions

that can improve performance even in the case of BCI illiterate user. Like, Zhang et al [176] addressed BCI illiteracy through a combination of CSP and brain network features. They constructed a task-related brain network by calculating the coherence between EEG channels, the graph-based analysis showed that the node degree and clustering coefficient have intensity differences between left and right-hand motor imagery. Their work suggests that there is a need to explore more feature extraction methods to address the BCI illiteracy problem. Furthermore, Yao et al. [165] proposed a hybrid BCI system to address the BCI inefficiency based on somatosensory attentional (SA) and motor imagery (MI) modalities. SA and MI are generated by attentional concentration intention (at some focused body part) and mentally simulating the kinesthetic movement respectively. SA and MI are reflected through EEG signals at the somatosensory and motor cortices respectively. In their work, they demonstrate that the combination of SA and MI would provide distinctive features to improve performance and increase the number of commands in a BCI system. In the same vein, Sannelli et al. [177] created an ensemble of adaptive spatial filters to increase BCI performance for BCI inefficient users. External factors can also improve BCI accuracy. For instance, Vidaurre et al. [178] proposed assistive peripheral electrical stimulation to modulate activity in the sensorimotor cortex. It is proposed that this will elicit short-term and long-term improvements in sensorimotor function thus improve BCI illiteracy among users.

2.3.4 Asynchronised MI-BCI

MI-based BCI is usually trained in a synchronous manner, that is, there exists a sequence of instructions (or cue) which user follows to produce an ERD/ERS phenomenon. However, in a real-world application, user want to execute control signal at his own will rather than waiting for cue. Therefore, there has been an increasing interest in creating an asynchronous MI. That is, MI-based BCI can detect that the user has an intention to do a Motor imagery then classifies MI task. This is done by splitting the incoming data into segments with overlapping periods. Each segment represents a potential MI command. One way of determining whether this potential MI command is an actual MI command is to build a classifier for that purpose. For example, Yu et al. [179] study presents the self-paced operation of a brain-computer interface (BCI),

which can be voluntarily used to control a movement of a car (starting the engine, turning right, turning left, moving forward, moving backward and stopping the engine). The system involved two classifiers: control intention classifier (CIC), and left/right classifier (LRC). The CIC is implemented in the first phase to identify the user intention being "idle" or "MI task-related". If an MI task-related is identified, a second phase follows the first phase by classifying it. Similarly, both Cheng et al. [180] and Antelis et al. [181] proposed a deep learning method that is trained to distinguish between resting state, transition state, and execution state. However, Cheng proposed a convolutional neural network followed by a fully connected network (CNN-FC) while Antelis proposed Dendrite morphological neural networks (DMNN). Another approach is to let the subject achieve a set number of consistent right/left classification within a set period for an action to be taken thus confirming the command and avoiding randomness [182], both adding a classifier and classifying multiple times, adds computational time and complexity to the system, the latter also adding time required for classification. Sun et al. [183] suggested a method that avoids these constraints by using a threshold on an existing classifier that separates idle from MI task-related. He et al. [184] proposed a similar approach for continuous application such as mouse movement. This is achieved through moving the object (in this case a mouse) by confidence level of the classifier. The threshold-based method of addressing this challenge requires defining a threshold which could be difficult and user-dependent. This brings us to the last methodology of addressing this challenge, that is, by adding an idle class into the classifier [185–188]. All of the above-motivated methods, except Yousefi et al. [188] proposed method, use a target-oriented paradigm where the user is asked to perform a task and the algorithm is evaluated based on the user's ability to achieve that task. However, Yousefi et al. [188] tested their algorithm by giving the user a specified time interval to perform any task the user desired and after the time has passed, the user provides feedback as to whether the algorithm responded to his commands. In conclusion, all algorithm can run asynchronously given they have a reasonable run time.

2.3.5 Increase number of commands

More diverse and complex applications like spellers etcetera can be developed with high ITR and increased number of classes in MI-BCI. Traditionally, MI-BCI was

designed as binary class (left and right) problem. The first way to extend MI-BCI into multi-class is by employing a hybrid approach during which the MI paradigm is complemented with another mental strategies. For example, Yu et al. [189] proposed a hybrid asynchronous brain-computer interface based on sequential motor imagery (MI) and P300 potential to execute eleven control functions for wheelchairs. The second way to achieve multi-class MI-BCI is algorithmically. For example, the traditional CSP algorithm is extended to recognize four MI tasks [190]. In the similar manner, Wentrup and Buss [191] proposed information theoretic feature extraction frameworks for CSP algorithm by extending it for multiclass MI-BCI system. In the same vein, Christensen et al. [192] extended FBCSPs for five class MI-BCI system. Similarly, Razzak et al. [193] proposed novel multiclass support matrix machine to handle multiclass MI imagery tasks. Likewise, Barachant et al. [194] presented a new classification method based on Riemannian geometry that uses covariance matrices to classify multi-class BCI. Faiz and Hamadani [195] controlled humonoid robotic hand gatures through five class online MI BCI using a commercial EEG headset. They user AR and CSP feature extractions and PCA to reduce the dimension of AR features. Lastly, CSP and AR features are concatenated and trained by a SVM classifier to achieve multi-class recognition.

2.3.6 Adaptive BCI

One of the issues still being worked in EEG based MI-BCI is the consistency of the accuracy of the classifier during long sessions. This is because EEG is a non-stationary signal that get impacted over time as well as when there is change in recording environment and state of mind (e.g. fatigue, attention, motivation, emotion etc). Adaptive methods have been proposed to address this challenge. For instance, Aliakbaryhosseinabadi et al [196] demonstrated that it is possible to detect a user's attention diversion during a MI task, whereas, Dagaev et al. [197] extracted target state (LH, RH) from background state (environment, emotional and cognitive condition etc.). This was achieved by asking subjects in the training stage to open and close their eyes during the trials. These instructions act as the two different background conditions. Methods that detect cause of change in user signals other than MI task could pave the way for adaptive MI-BCI by giving both the user real-time neurofeedback and giving the adaptive algorithm additional information to work

with while decoding MI task.

Another way to address this challenge is to modify the training protocol or extracting more information during it. Both Mondini et al. [198] and Schwarz et al. [199] modified the training protocol. By creating an adaptive training protocol, Mondini et al. [198] fulfil three tasks: (a) adapt the training session based on the subject's ability, that is, make the training short and restart the training from the beginning with different motor imagery strategy if the system performance is lower than a certain threshold; (b) present training cue (left/right) in a biased manner that is present left cue more often manner if left imagery performance is low compared to the right; (c)keep challenging the performance of the user by only giving feedback if it exceeds an adaptive threshold. Schwarz et al. [199] by proposed a co-adaptive online learning BCI model that uses the concept of semi-supervised retraining. Schwarz model uses a few initial supervised calibration trials per MI tasks and then perform recurrent retraining by using artificially generated labels. This ensures feedback to the user after a very short training and engages the user in mutual learning with the system. Information gathered during training protocol such as command delivery time (CDT), and the probability of the next action could be used to address this challenge. CDT has been used by Saeedi et al. [200] to provide a system that delivers adaptive assistance that is, if the current trial is long, then the system will slow down to give enough time to the user to execute the MI tasks. Their study suggests that brain pattern is different for short, long and time-out commands. They were able to differentiate between command type using only one second before the trial started, while Perdakis et al. [201] proposed to use probability of next action to adapt the classifier. Specifically, they implemented an online speller based on the BrainTree MI text-entry system that uses probabilistic contextual information to adapt an LDA classifier. The final method observed in the literature to address this challenge was to create an adaptive classifier. Faller et al. [202] proposed an online adaptive MI-BCI that auto-calibrates. Their system in regular interval not only discriminates features for classifier retraining but also learns to reject outliers. Their system starts to provide feedback after minimal training and keeps improving by learning subject-specific parameters on the run. Raza et al. [203] proposed unsupervised adaptive ensemble learning algorithm that tackles non-stationary based co-variate shifts between two BCI sessions. This algorithm

paves the way for online adaption to variabilities between BCI sessions. In the same vein, Rong et al. [204] proposed an online method that handles the statistical difference between sessions. They used an adaptive fuzzy inference system.

2.3.7 Online MI-BCI

After an adaptive BCI, one key factor that determines MI based system's usability and efficacy is BCI mode. MI-BCI systems are operated in offline or online mode through cue-based paradigms where as self-placed (asynchronous) are mostly online systems. Mostly, literature proposed improvements in offline mode of MI-BCI systems; very few test their proposed algorithms in online environment. In online BCI studies, Sharghian et al. [205] proposed MI-EEG which uses sparse representation-based Classification (SRC). Their approach obtains an online dictionary learning scheme from extracted band power from a spatial-filtered signal. This dictionary leads to reconstruction of sparse signal for classification. In the same vein, Zhang et al. [206] proposed an incremental linear discriminant analysis algorithm that extract AR features from preferable incoming data. Their method paved way for fully auto-calibrating an online MI-BCI system. Similarly, Yu et al. [185] proposed an asynchronous MI BCI system to control wheelchair navigation. Perez [207] extended the fuzzy logic framework for adaptive online MI-BCI system and evaluated it through realistic navigation of a bipedal robot. Ang and Guan [208] introduced an adaptive strategy that continuously compute the subject-specific model during an online phase. Abdalsalam et al. [209] controlled the screen cursor through a four class MI-BCI system. Their results suggest that online feedback increase ERDs over μ ($8 - 10Hz$) and upper beta ($18 - 24Hz$) band which results into a higher cursor control success rate. Many studies have demonstrated the efficiency of virtual reality (VR) and gaming environment in a online BCI [210]. Achanccaray et al. [211] in the same vein, verified that virtual reality based online feedback has positive effects on the subject. It has been observed that motor cortex increases its activation level (in alpha and beta band) due to an immersive VR experience. This is very helpful in supporting upper limb rehabilitation of post-stroke patients. Similarly, Alchalabi and Faubert [212] used VR based neurofeedback in the Online MI-BCI session. Cubero et al. [213] proposed an online system based on an endless running game that runs on three

class MI-BCI. They used graphic representation of EEG signals for multi-resolution analysis to take advantage of spatial dimension along with temporal and spectral dimensions .

2.3.8 Training Protocol

Similar to other normal user skills, BCI control is also a skill that can be learned and improved with proper training. A typical BCI training protocol is a combination of user instructions, cues on screen to modulate the user's neural activity in a specific manner and, lastly, a feedback mechanism that represents confidence of the classifier in recognition of the mental task to user. Unfortunately, standard training protocol does not satisfy the psychology of human learning; being usually boring and very long. Meng and He [214] studied effect of MI training on users. They found out that with a few hours of MI training, there is change in electrophysiological properties. Their study suggested design engaging training protocol and multiple training sessions rather than a long training session for low BCI performers. In the same vein, Kim et al. [215] proposed a self placed training protocol in which the user performs MI task continuously without an inter-stimulus-interval. During each trial the user has to imagine a single MI task (e.g. RH for 60 sec). The results of this protocol showed that it reduces calibration time compared to conventional MI training protocol. Jeunet et al. [216] surveyed the cognitive and psychological factors related to MI-BCI and categorized these factors into three categories (a) user-technology relationship,(b) attention and (c) spatial abilities. Their work is very useful to design new training protocol that take advantage of these factors. Furthermore, in another study Jeunet et al. [29] found that spatial ability plays an important role in BCI performance of a subject. They suggested having pre-training sessions to explore spatial ability for BCI training.

Many studies proposed new training strategies which uses other mental strategies to compliment MI training (kinesthetic imagination of limbs). For instance, Zhang et al. [217] proposed a new BCI training paradigm which combine conventional MI training protocol with covert verb reading. This improves the performance of MI-BCI and paves the way for utilizing semantic processing with motor imagery. Along the same lines, Wang et al. [218] proposed hybrid MI-paradigm that uses speech imagery with motor imagery. In this paradigm, the user repeatedly and silently reads move (left/right) cues

during imagination. Standard training protocols are fixed that is not tailored made to user's need and experience. With respect to this, Wang et al. [219] proposed MI training with visual-haptic neurofeedback. Their findings validate that their approach improves cortical activations at the sensorimotor area, thus leads to an improvement in BCI performance. Liburkina et al. [220] proposed a MI training protocol that gives cue to perform and feedback to the user through vibration. Along the same lines, Pillette et al. [221] designed an intelligent tutoring system that provides support during MI training and enhance user experience/performance on MI-BCI system. Skola et al. [222] proposed a virtual reality-based MI-BCI training that uses a virtual avatar to give feedback. Their training helps in maintaining high levels of attention and motivation. Furthermore, their proposed method improves BCI skills of first time users.

2.4 Conclusion

In this chapter, we have provided an extensive review of methodologies for designing an MI-BCI system. In doing so, we have created a generic framework and mapped literature related to different components (data acquisition, MI training, preprocessing, feature extraction, channel and feature selection, classification and performance metrics) in it. This will help in visualizing gaps to be filled by future studies in order to further improve BCI usability.

Despite many outstanding developments in MI-BCI research, some critical issues still need to be resolved. Mostly, studies are on synchronised MI-BCI in offline mode. There is a need to have more studies on online BCI. Typically, researchers use performance evaluation metrics as per their convenience. It would be better to have general BCI standards that can be widely adhered by researchers. Our literature survey found that enhancing the performance is still a critical issue even after two decades of research. Due to availability of high computational resources, present studies employ methods based on deep learning and Riemannian geometry more than traditional machine learning methods. With current advancement in algorithms, future research should concentrate more on eliminating or reducing long calibration in MI-BCI. Future studies should focus on more diverse BCI applications which can be developed with increased number of commands. Our review shows that BCI illiteracy is


critical issue that can be addressed either by using better training protocol that suit users' requirements or through smart algorithms. Finally, EEG is a non-stationary signal that changes over time as user's state of mind changes. This cause inconsistency in BCI classifier's performance, thus it is important to make progress in development of adaptive methods to address this challenge in an online settings.



GRADUATE
RESEARCH
SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Amardeep Singh
Name/title of Primary Supervisor:	Prof. Hans W. Guesgen
In which chapter is the manuscript /published work: Chapter 3	
Please select one of the following three options:	
<input checked="" type="radio"/> The manuscript/published work is published or in press <ul style="list-style-type: none"> • Please provide the full reference of the Research Output: Amardeep Singh, Sunil Lal and Hans W. Guesgen, "Reduce calibration time in motor imagery using spatially regularized symmetric positives-definite matrices based classification", <i>Sensors</i> 2019, 19 (379), 2879. 	
<input type="radio"/> The manuscript is currently under review for publication – please indicate: <ul style="list-style-type: none"> • The name of the journal: • The percentage of the manuscript/published work that was contributed by the candidate: 80.00 • Describe the contribution that the candidate has made to the manuscript/published work: Conceptualization, Methodology, Investigation, Formal analysis, Validation, Writing - Original Draft, Writing - Review & Editing, Visualization. 	
<input type="radio"/> It is intended that the manuscript will be published, but it has not yet been submitted to a journal	
Candidate's Signature:	Amardeep Singh <small>Digitally signed by Amardeep Singh Date: 2021.02.26 13:13:11 +13'00'</small>
Date:	26-Feb-2021
Primary Supervisor's Signature:	
Date:	26-Feb-2021

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/publication or collected as an appendix at the end of the thesis.

Chapter 3

Reduce Calibration Time Using Spatially Regularized Symmetric Positive-Definite Matrices

Quick Summary

One major challenge in the development of a brain–computer interface is to reduce calibration time or completely eliminate it. To address this problem, existing approaches use covariance matrices of electroencephalography (EEG) trials as descriptors for decoding BCI but do not consider the geometry of the covariance matrices, which lies in the space of Symmetric Positive Definite (SPD) matrices. This inevitably limits their performance. We focus on reducing calibration time by introducing SPD based classification approach. However, SPD-based classification has limited applicability in small training sets because the dimensionality of covariance matrices is large in proportion to the number of trials. To overcome this drawback, we propose a new framework that transforms SPD matrices in lower dimension through spatial filter regularized by prior information of EEG channels.

Related Paper

- **Amardeep Singh**, Sunil Lal and Hans W. Guesgen. "Reduce calibration time in motor imagery using spatially regularized symmetric positive-definite matrices based classification" *Sensors* 2019, 19(379), 2879

3.1 Introduction

Electroencephalogram (EEG) based brain–computer interfaces (BCI) detect neural activity from brain scalp and translate them into control commands for external devices [2]. EEG based BCI systems can be categorized as exogenous or endogenous, according to paradigm used to generate neural activity [223]. An exogenous BCI derives its output from neural activity (EEG signals) generated due to attentional selection of an external stimulus among many [22]. An endogenous BCI derives its outputs from oscillatory neural activity, which is spontaneously controlled by the user [22]. Endogenous BCI does not require external stimulus to generate specific neural pattern for BCI, in fact the user can control BCI system voluntarily. Motor imagery (MI) is one such endogenous BCI paradigm where neural activity is generated at the sensorimotor cortex due to the kinaesthetic imagination of a body part (left/right hand) movement [71]. During MI, there is a rhythmic power decrease or increase in measured EEG signals from the sensorimotor cortex. These phenomena are also known as event related desynchronization (ERD) and event related synchronization (ERS), respectively [224]. BCIs distinguish different MI tasks through spatial and temporal properties of measured EEG signals [63]. Therefore, to increase the spatial and temporal resolution, electroencephalogram (EEG) signals are recorded with multi-channel electrodes system with high sampling rate. This results in high-dimensional signals.

MI-based BCIs are indeed very promising for people suffering from neuromuscular disorder, but still lack adoption as access modalities outside laboratories. The main reason that prevents MI-based BCIs from widely being used is high performance variations among and within subjects. These performance variations are due to change in the external (user’s muscle movements, recording condition and machine related causes) and internal (user’s cognitive state of mind) state of the user [225]. Therefore, it requires extensive training compared to exogenous BCI systems. During training and calibration phase, new subjects learn to voluntarily regulate oscillatory EEG pattern and training trials are collected to obtain discriminative features that are fed into machine learning algorithms for MI classification [16].

The standard feature extraction techniques for motor imagery use covariance matrices of trials. One such technique is common spatial pattern (CSP) that aims to determine optimal spatial filters that discriminate two MI task

(left/right) [226]. CSP requires large number of subject specific calibration trial sessions to achieve good MI classification. These calibration sessions are very time consuming and not user-friendly. Thus, it is desirable to reduce or remove the calibration entirely.

However, in the case of a small EEG calibration trials set, these covariance matrices poorly estimate MI and therefore lead to poor performance of CSP. To address this, Lu et al. [227] proposed Regularized CSP, which uses other subjects' trials to construct MI classes spatial covariance matrices for new target subjects that will be used to extract CSP features. In the same vein, Dai et al. [228] also employed transfer learning technique to learn domain invariant CSP features from source and target subjects. Both approaches rely on other subjects (source) to subject (target) transfer learning, which, in the worst case, might hurt the performance of the target subject. This situation is often called negative transfer [229].

Unlike the above methods, Arvaneh et al. [230] suggested a technique that does not rely on source subjects trials. Rather, this approach optimizes obtained CSP filters by using channels from brain regions that have high variances between MI classes, and attenuates the noisy channels from regions with low and irregular variances. Similarly, Lotte et al. [149] used spatial information of electrodes as prior knowledge to regularize objective function of the CSP algorithm to obtain spatial filters. In similar manner, Park and Chung [231] used electrodes from certain brain regions to extracted diverse CSP features and obtain high accuracy compared to standard CSP under small training samples (trials).

The efficiency of spatial filter is sensitive to individual's temporal and frequency characteristics. To address subject specific frequency characteristics issue, Ang et al. [226] proposed filter bank CSP (FBCSP) that uses multiple bandpass filters to extract CSP features. FBCSP may lose important frequency information, as it uses fixed partition of the frequency (frequency width of 4 Hz, varying from 4 Hz to 30 Hz). To address this problem, Yang et al. [232] proposed CSP feature extraction based on varying partition of the frequency bands with different bandwidth to cover as many bands as possible. In a similar way, Park and Lee [233] extended FBCSP by regularizing CSP features obtained from multiple filter banks. They used other subjects' trials covariance matrices to regularize filter bank CSP features. Zhang et al. [234] proposed a

method that simultaneously optimizes filter bands and time window used to obtain CSP features to further boost classification accuracy of MI. Filter bank-based methods result in a high dimensional CSP feature set, therefore requires a feature selection algorithm to select discriminative CSP features for MI classification. To address the feature selection problem, Selim et al. [138] used bio-inspired optimization algorithm for feature selection. They also selected optimal time interval for each subject to extract CSP features. Unlike the above methods, Tang et al. [235] used a convolutional neural network model to classify MI tasks based on spatiotemporal characteristics of EEG. Furthermore, Tabar and Halici [236] combined convolutional neural network and stacked autoencoders to classify EEG Motor Imagery signals.

All methods discussed above use covariance matrices of trials to extract CSP features (log variance) into a vector in Euclidean space. Furthermore, pattern recognition metrics used to classify features also lies in Euclidean space. As covariance matrices lie in the symmetrical positive definite (SPD) matrices manifold, these methods fail to notice distinct characteristic of EEG data such as their interrelation across the manifold dimensions [17].

The effectiveness of data treatment based on the concept of geometrical properties was proved by Barachant et al. [194]. They proposed minimum distance to Riemannian mean (MDRM) classification technique that adopts Riemannian distance as pattern recognition metric to classify test trials. MDRM outperforms standard CSP approach, but performance of MDRM declines as the size of covariance matrices grows. Under small training set, the size of covariance matrices are larger than the number of trials. Therefore, MDRM algorithms encounter the curse of dimensionality problem [97].

To address dimensionality problem, Horev et al. [72] adapted PCA to the space of SPD matrices, which conserves more data variance and maps covariance matrices to a lower-dimensional SPD manifold. In a similar manner, Harandi et al. [237] learned mapping that maximizes the geodesic distances between inter-class samples and simultaneously minimizes the distances between intra-class samples. This was done via optimization on Grassmann manifolds. This algorithm tries to preserve the local structure of the data by preserving distance to local means, considers the geometry of SPD matrices, provides an implicit mapping and applies the supervised information for embedding to lower-dimensional space. Furthermore, Davoudi et al. [97] extended Harandi's

work by proposing another dimensionality reduction algorithm for the manifold of SPD matrices which preserves the local structure of data by preserving distance to local mean (DPLM). This algorithm can work in a supervised (sDPLM) or unsupervised (uDPLM) manner and projects a high-dimensional SPD manifold to a lower-dimensional one. In the same vein, Kumar et al. [238] also addressed dimensionality issue of covariance matrices by using spatial filtering. The drawback of this method is that it requires many subject-specific trials to optimize spatial filter performance. In this chapter, we propose a method that uses the best of both Euclidean and SPD space. We use prior information of EEG electrodes to obtain spatial filter that transform sample covariance matrices (SCM) into lower dimension. Then, Riemannian distance is used as pattern recognition metric for classification as it is invariant to any linear invertible transformation [152].

The rest of the chapter is organized as follows. In Section 3.2, we review the space of SPD matrices and MDRM classification approach. Section 3.3 presents our proposed SR-MDRM classification approach. Section 3.4 describes the experiment as well as datasets. In Section 3.5, we discuss and compare results of the experiment with existing studies. Section 3.6 draws the conclusions regarding proposed approach.

3.2 Geometry of SPD Matrices

An $n \times n$ square matrix C falls in the space of symmetric positive definite $SPD(n)$ if $\{C = C^T, u^T C u > 0, \forall u \in \mathbb{R}^n, \forall u \neq 0\}$. Symmetric positive definite matrices are always diagonalizable with strictly real positive eigenvalues. For SPD matrices in $SPD(n)$, the exponential matrix of C is obtained using the eigenvalue decomposition of C :

$$C = U \text{diag}([\lambda_1, \dots, \lambda_n]) U^T \quad (3.1)$$

where $\lambda_1 > \lambda_2 > \dots > \lambda_n$ are the eigenvalues and U the matrix of eigenvectors of C . It reads:

$$\exp(C) = U \text{diag}([\exp(\lambda_1), \dots, \exp(\lambda_n)]) U^T \quad (3.2)$$

The inverse operation is the logarithm of a SPD matrix [194] :

$$\log(C) = U \text{diag}([\log(\lambda_1), \dots, \log(\lambda_n)]) U^T \quad (3.3)$$

Equivalently, SPD matrices have the following properties [194]:

1. $\forall C \in SPD(n), C^{-1} \in SPD(n)$ i.e., SPD matrices are invertible.
2. $\forall C \in SPD(n), \det(C) > 0$

Covariance matrices of EEG trial lies in symmetric positive definite matrices manifold [239]. Covariance matrices hold spatial-temporal information for EEG trial and can directly be used for classification. SPD matrices lie on a differentiable Riemannian manifold.

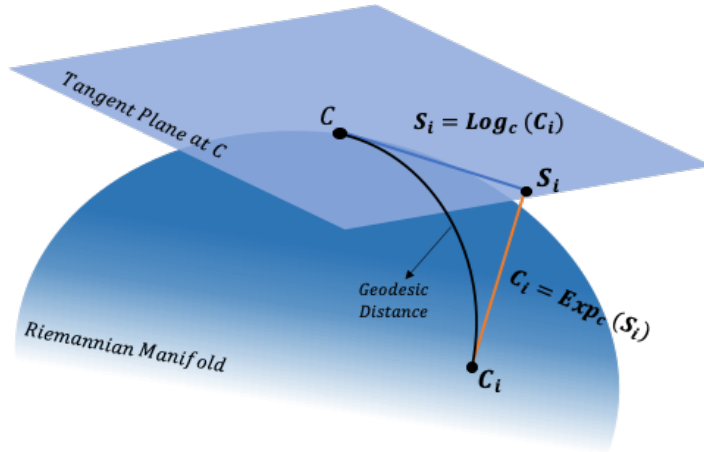


Figure 3.1: Illustration of Tangent space of manifold \mathcal{M} at point C . The Logarithmic map projects the matrix $C_i \in \mathcal{M}$ into the tangent space. The Exponential map projects the element of the tangent space S_i back to the manifold. Geodesic curve illustrate unique and shortest curve connecting C_i and C .

3.2.1 Riemannian Manifold

A Riemannian manifold \mathcal{M} is a differentiable manifold in which the tangent space at each point is a finite-dimensional Euclidean space [71]. That means, derivatives at a point C (i.e. covariance matrix of MI trial) on manifold lies in a vector space $\mathcal{T}_C \mathcal{M}$, which is the tangent space at that point. The manifold \mathcal{M} and the tangent space are $m = n(n + 1)/2$ dimensional [17]. This tangent space is Euclidean and

locally homomorphic to the manifold and Riemannian distance computations in the manifold can be well approximated by Euclidean distance computations in the tangent space [71]. A point (matrix) C_i is projected to tangent space using logarithmic mapping $Log_C(C_i)$ as

$$S_i \in \mathcal{T}_C \mathcal{M} = Log_C(C_i) = C^{1/2} \log_m(C^{-1/2} C_i C^{-1/2}) C^{1/2} \quad (3.4)$$

where C is a reference point in the manifold where the tangent plane is mapped and $\log_m(\cdot)$ is logarithm of SPD matrix given in Equation (3.3). Furthermore, the tangent vector S_i from tangent space is projected back to manifold using exponential mapping $Exp_C(S_i)$ given by

$$C_i \in \mathcal{M} = Exp_C(S_i) = C^{1/2} \exp_m(C^{-1/2} S_i C^{-1/2}) C^{1/2} \quad (3.5)$$

where $\exp_m(\cdot)$ is exponential of SPD matrix, as shown in Equation (3.2). Figure 3.1 illustrate logarithmic mapping operator and exponential mapping operator on Riemannian Manifold and corresponding local tangent space at C .

3.2.2 Riemannian Distance

The minimum length curve connecting two points on the manifold is geodesic, and the Riemannian distance between the two points is given by the length of this curve as illustrated in Figure 3.1. Mathematically, Riemannian distance between two SPD matrices C_1 and C_2 in the Riemannian manifold \mathcal{M} is given by:

$$R_d(C_1, C_2) = \|\log(C_1^{-1/2} C_2 C_1^{-1/2})\|_F = \left(\sum_{i=1}^n \log^2 \lambda_i \right)^{1/2} \quad (3.6)$$

where $\|\cdot\|_F$ is the Frobenius norm and λ_i 's are the positive eigenvalues of $C_1^{-1/2} C_2 C_1^{-1/2}$. The Riemannian distance $R_d(C_1, C_2)$ is invariant to any linear invertible transformation [152]:

$$R_d(A^T C_1 A, A^T C_2 A) = R_d(C_1, C_2) \quad (3.7)$$

where A is an invertible matrix. The Riemannian distance between two points in manifold can be approximated in tangent space by approximating the distance between projected tangent vectors through a reference point C . To obtain a good approximation of the Riemannian distance, reference point C needs to be close to two points in the manifold. Usually, the Riemannian mean $\Pi_{\mathcal{M}}$ is the most suitable choice for the reference point.

3.2.3 The Choice of a Reference SPD matrix

The reference SPD matrix is free parameter that defines the point in the manifold \mathcal{M} where the tangent plane is computed [71]. The most common choice of reference matrix C_{ref} is the average of the whole set of covariance matrices. All the averaging approaches such as arithmetic, geometric, and harmonic mean estimate the centrality of manifold \mathcal{M} [240]. In this thesis, we used the geometric mean (also known as Riemannian mean) of SPD matrices that minimizes the sum of squared Riemannian distances. It is given by:

$$\Pi_{\mathcal{M}}(C_1, \dots, C_I) = \underset{C \in SPD(n)}{arg\ min} \sum_{i=1}^I R_d^2(C_i, C) \quad (3.8)$$

The mean of $I \geq 1$ SPD matrices such as EEG trials covariance matrices keeps shifting due to the non-stationarity of EEG signals. Therefore, it needs to be iteratively computed whenever any new trials are collected. The computation of the Riemannian mean $\Pi_{\mathcal{M}}$ goes through the following steps until it converges. Firstly, projecting the covariance matrices in the tangent space, estimating the arithmetic mean in the tangent space and projecting the arithmetic mean back in the manifold. Then iterate the three above steps until convergence. Chebbi and Moakher [241] provide a detail discourse on the computation of Riemannian mean.

3.2.4 Minimum Distance to Riemannian Mean (MDRM)

MDRM is a classification approach that uses the Riemannian mean of each class and its Riemannian distance to test covariance matrix of the trial to predict a label for it. In this approach, the Riemannian mean is calculated for each class using its labeled training trials, and then the Riemannian distance of each class is calculated with respect to test trial's covariance matrix. The class mean that is closest to test trial covariance becomes the trial's label.

$$pred(C_x) = \underset{\varphi=1,2..K}{arg\ min} R_d(C_x, C_{\Pi}) \quad (3.9)$$

where C_x is the covariance matrix of the test trial, C_{Π} is the Riemannian mean of Class φ and $pred(C_x)$ is the prediction of its class label. The MDRM approach is not robust to noise [97], therefore, some filtering over SPD matrices

is required. Barachant et al. [242] suggested geodesic filtered MDM (FGMDM) approach, which computes set of filters by applying a supervised Fisher geodesic discriminant analysis (FGDA) to the tangent (Euclidean) space projection of covariance matrices. The obtained filters are applied through geodesic filtering approach [242] over SPD matrices. This filtering operation do not change any dimension of the SPD matrices. Finally, the filtered SPD matrices are used for MDRM classification.

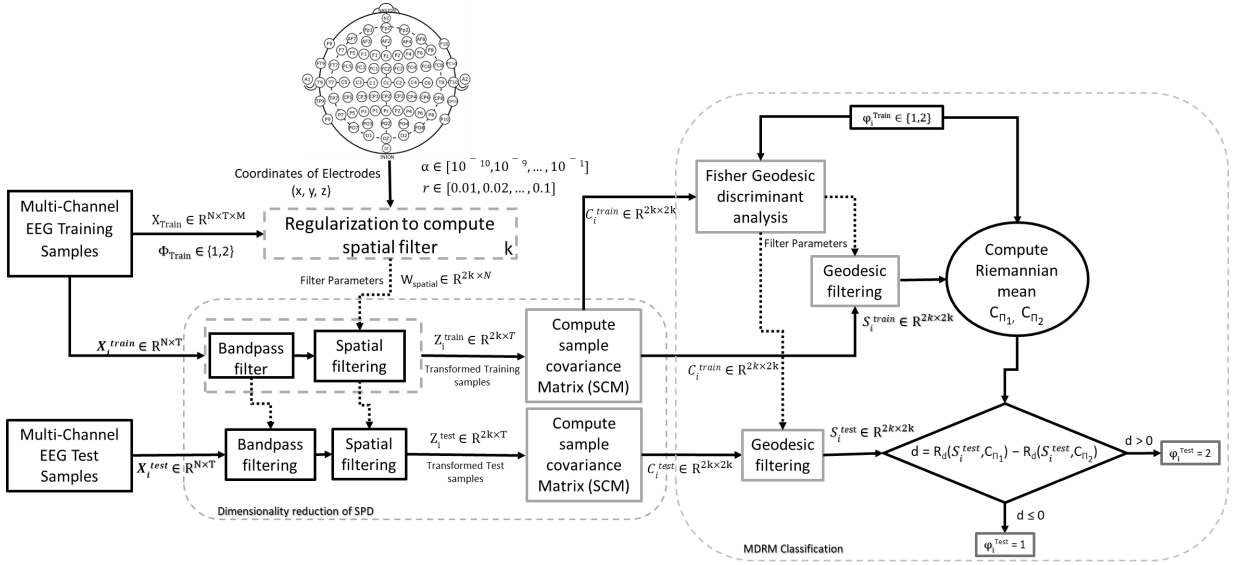


Figure 3.2: Framework for proposed approach.

3.3 Methodology

The conceptual framework of our proposed methodology is shown in Figure 3.2. EEG signals are often divided into trials based on the label given during training phase. Let $X_i \in (\mathbb{R}^{N \times T})$ be a bandpassed EEG trial where N is number of electrodes and T is sampled time points in the trials and trial labels $\varphi_i \in (1, 2)$. Therefore, the training set can be given as $\{X_i, \varphi_i\}_{i=1}^M$ where M is the total number of training trials. The normalized sample covariance matrix C_i [243] for trial X_i is calculated as follows:

$$C_i = \frac{X_i X_i^T}{tr(X_i X_i^T)} \quad (3.10)$$

where $tr(\cdot)$ denotes the trace operator of the matrix, and the superscript T denotes the transpose of the matrix. Normalization on covariance matrices to reduce remaining sources of variability such as variations related to electrode impedances [243]. The sample covariance matrix for class can be obtained by taking sum of sample covariance matrices for M trials that belong to it. It is calculated as follows:

$$C_\varphi = \sum_{m=1}^M C_{(\varphi,m)} \quad (3.11)$$

where M is the total trial number of each class, and m is the index of the trial ($m \leq M$). φ denotes the class index, and we consider only two classes ($\varphi \in \{1, 2\}$) in this paper. Spatial filters w through CSP are obtained by extremizing the following function:

$$J(w) = \frac{w^T C_1 w}{w^T C_2 w} \quad (3.12)$$

This is an optimization problem that can be solved by Lagrange multiplier method using Equation (3.13):

$$L(\lambda, w) = w^T C_1 w - \lambda(w^T C_2 w - 1) \quad (3.13)$$

The filters w extremizing L are such that the derivative of L with respect to w equals 0:

$$\begin{aligned} \frac{\partial L}{\partial w} &= 2w^T C_1 - 2\lambda w^T C_2 = 0 \\ &\Leftrightarrow C_1 w = \lambda C_2 w \\ &\Leftrightarrow C_2^{-1} C_1 w = \lambda w \end{aligned} \quad (3.14)$$

Equation (3.14) is a standard eigenvalue problem. The matrix $C_2^{-1} C_1$ containing the first $k/2$ vectors explains the maximum of the variance of class 1 and the minimum of class 2, while the last $k/2$ vectors explains the maximum of the variance of class 2 and the minimum of class 1. if k is too small, the classifier would fail to fully capture the discrimination between two classes after spatial filtering. To get optimal results, $k = 6$, i.e., three eigenvectors from both ends of $C_2^{-1} C_1$ are used as the spatial filters w . Alternatively one can choose the eigenvectors according to different criterion or use cross-validation to determine the number of components [244]. From a neuro-physiological point of view, neighboring neurons tend to function similarly, so if the two electrodes are close

enough could measure similar brain signals [245] (As there are thousands of neuron comes under the area of single electrode) . Thus, we can expect that neighboring channels of the spatial filter should have similar weights (i.e., smooth spatial filter). To obtain smooth spatial filter, we use the spatial information of electrodes as a prior knowledge [149] to penalize objective function of CSP algorithm. Smooth spatial filters w can be obtained by extremizing the following functions:

$$J_{P_1}(w) = \frac{w^T C_1 w}{w^T C_2 w + \alpha P(w)} \text{ and } J_{P_2}(w) = \frac{w^T C_2 w}{w^T C_1 w + \alpha P(w)} \quad (3.15)$$

The penalty term $P(w)$ measures the spatial smoothness of the spatial filters w , where $P(w) = w^T K w$ with $K = D - G$. G is a Gaussian Kernel such that $G_{ij} = \exp -\frac{1}{2}(\frac{\|v_i - v_j\|^2}{r^2})$, with v_i a vector containing 3D coordinates of the i th electrode. D is a diagonal matrix such as $D_{ii} = \sum_j G_{ij}$. Therefore, $w^T K w = w^T (D - G) w = \sum_{i,j} G_{ij} (w_i - w_j)^2$. There are two hyperparameters (r, α) in regularized objective function. The first hyperparameter r defines how far two electrodes can be to be still considered as close to each other and the second hyperparameter α defines the level of spatial smoothness the filters should reach. Equation (3.15) becomes:

$$J_{P_1}(w) = \frac{w^T C_1 w}{w^T C_2 w + \alpha w^T K w} \text{ and } J_{P_2}(w) = \frac{w^T C_2 w}{w^T C_1 w + \alpha w^T K w} \quad (3.16)$$

using Lagrangian multiplier method, the solution is obtained as,

$$M_1 = (C_2 + \alpha K)^{-1} C_1 \text{ and } M_2 = (C_1 + \alpha K)^{-1} C_2 \quad (3.17)$$

We construct the projection matrix $W_p \in \mathbb{R}^{2N \times N}$ using Equation (3.18)

$$W_p = [M_1 M_2] \quad (3.18)$$

The EEG signal trial is transformed with $W_{spatial}$ made from the first and last k columns of W_p by using Equation (3.19)

$$Z_i = W_{spatial} X_i \quad (3.19)$$

where $Z_i \in \mathbb{R}^{2k \times T}$ is transformed signal corresponding to X_i . Sample covariance

matrices C_i^{train} of filtered EEG trials Z_i from the target’s training set are calculated using Equation (3.10). The MDRM classification algorithm is based on shortest Riemannian distance between the test covariance matrix and two MI class means. The main limitation of this approach is that due to noise related distance between two matrices can make class related information based distance get vanish. Therefore, it is preferable to perform some filtering over SPD matrices before applying MDRM classification. In order to discard irrelevant information and support class-related information geodesic filtering is applied on covariance matrices C_i^{train} in the tangent space using an algorithm, as mentioned in [242]. This algorithm is an extension of Fisher Linear Discriminant Analysis in the tangent space. After geodesic filtering, filtered SCMs (S_i^{Train} and S_i^{Test}) for target subject’s training and test trials are obtained. Finally, filtered SCMs of the target subject’s training set (S_i^{Train}) are used in calculating Riemannian mean $C_{\Pi_{R\varphi}}$ for both motor imagery classes. These Riemannian means are used for MDRM classification [242] of test trials (S_i^{Test}).

3.4 Data and Experiment

To assess the performance of our method for small training setting, we used the EEG Dataset IVa from BCI Competition III. Furthermore, we compared it with existing methods designed for small training set scenario. To confirm the robustness of proposed approach, we evaluated it over two publicly available datasets with a different number of EEG channels from BCI competition. A summary of the three datasets is given in Table 3.1.

Dataset IVa, BCI Competition III

Dataset IVa [246] contains EEG signals of binary (right hand and foot) motor imagery tasks from five healthy subjects. EEG signals were recorded using 118 electrodes at 100 Hz sampling rate. There are a total 280 trials per subject that are unevenly divided into training and testing set for each subjects, as shown in Table 3.1.

Dataset IIIa, BCI Competition III

Dataset IIIa [246] comprises of EEG signals of multi-class (right hand, left hand, tongue and foot) motor imagery (MI) tasks from three subjects (“k3b”, “k6b” and “l1b”). EEG signals were sampled at 250 Hz rate and recorded using 60 electrodes. In this study, we used EEG signals from trials corresponding to binary MI class (left/right). There are total 180 trials for subject “k3b” and 120 trials for subjects “k6b” and “lib”, respectively.

Dataset IIa, BCI Competition IV

Dataset IIa [247] contains of data recorded from 22 EEG channels and 3 EOG channels at sampling rate of 250 Hz. Dataset IIa contains multi class EEG signals from nine subjects, namely A01–A09. In this experiment, we considered data collected from 22 EEG electrodes corresponding to left and right MI class from each of nine subjects. Table 3.1 shows number of training and testing trials for all subjects.

Table 3.1: Summary of Dataset IVa, Dataset IIIa and Dataset IIa from BCI competitions (BCIC).

BCI Competition (BCI-C)	BCI-C III			BCI-C IV					
Dataset	Dataset IVa			Dataset IIIa			Dataset IIa		
ELECTRODES	118			60			22		
SAMPLING RATE	100 Hz			250 Hz			250 Hz		
SUBJECT	AA	AL	AV	AW	AY	K3B	K6B	L1B	A01–A09
TRAIN	168	224	84	56	28	90	60	60	144
TEST	112	56	196	224	252	90	60	60	144

3.4.1 Experimental Setup

This study was carried out using a Windows 10 computer with specification Intel (R) Core™ i5–6500 CPU @3.20 GHz with 8 GB RAM. All conventional methods (CSP and MDRM) and proposed algorithm were implemented in Matlab R2018a.

The study comprised six steps. Firstly, we used a time segment from 0.5 s to 2.5 s after the visual cue for all the datasets considered for this study [14, 248]. Thus, trials respective to Dataset IVa, Dataset IIIa and Dataset IIa comprised

200, 500 and 500 sampled time points. Secondly, all trials were filtered in frequency range within 7–30 Hz through fifth order Butterworth bandpass IIR filter. This frequency band was selected as it comprises the alpha and beta frequency bands, which have been shown to be most important for MI task classification [249, 250]. Thirdly, spatial filters were learned using hyper-parameters $\alpha \in [10^{-10}, 10^{-9}, \dots, 10^{-1}]$ and $r \in [0.01, 0.02, \dots, 0.09, 0.1]$. In the fourth step, the EEG signals were transformed into lower dimension using regularized spatial filter. In the fifth step, covariance matrices for training trials were employed to obtain FGDA filters. Lastly, after geodesic filtering, Riemannian mean for each MI class was calculated using training trials covariance matrices and labels were assigned to test trials based on their distance from the Riemannian mean of MI classes. For CSP and MDRM, we used the same time-segment, filter order and frequency band as described for the proposed method.

3.4.2 Evaluation Metrics

To evaluate the performance of proposed method, we used classification accuracy and kappa coefficient as evaluation metrics. In binary classification case, accuracy can be calculated as described in Equation (4.19).

$$Accuracy = \frac{a + b}{a + b + c + d} \quad (3.20)$$

where a is the number of positive samples correctly identified, b is the number of negative samples correctly identified, c is the number of negative cases incorrectly identified, and d is the number of positive cases incorrectly identified. Kappa coefficient compares the accuracy of the system to the accuracy of a random system. It is defined as

$$kappa = \frac{observedAccuracy - randomAccuracy}{1 - randomAccuracy} \quad (3.21)$$

where random accuracy [251] is given by

$$randomAccuracy = \left(\frac{a + b}{a + b + c + d}\right) * \left(\frac{a + c}{a + b + c + d}\right) + \left(\frac{c + d}{a + b + c + d}\right) * \left(\frac{b + d}{a + b + c + d}\right) \quad (3.22)$$

3.5 Results and Discussion

We evaluated the performance of the proposed approach (SR-MDRM) on the three datasets, and compared it with conventional (CSP and MDRM) methods as well as benchmark results reported in the literature. Table 3.2 shows hyper-parameters used in SR-MDRM classification for all subjects belonging to different datasets.

Table 3.2: Hyper-parameters α, r of SR-MDRM for all subjects belonging to different datasets.

Parameters	Dataset IVa					Dataset IIIa			Dataset IIa									
	aa	al	av	aw	ay	k3b	k6b	l1b	A01	A02	A03	A04	A05	A06	A07	A08	A09	
α	10^{-1}	10^{-4}	10^{-5}	10^{-5}	10^{-1}	10^{-4}	10^{-3}	10^{-10}	10^{-4}	10^{-2}	10^{-3}	10^{-3}	10^{-4}	10^{-2}	10^{-3}	10^{-10}	10^{-10}	
r	0.06	all r values		0.07	0.07	0.08	0.1	0.08	0.01	0.06	0.07	0.05	0.09	0.07	0.04	0.06	0.01	0.01

3.5.1 Dataset IVa, BCI Competition III

Table 3.3 shows the classification accuracy proposed method, winner of BCI Competition III on Dataset IVa, CSP method and other benchmark results reported in the literature on Dataset IVa. As shown in Table 3.3, our method outperformed the existing studies in the literature, except for the winner. In this study, we used same approach for all subjects; on the contrary, winner [246] did not use the same approach for all subjects. Wang et al. (winner) [246] used an ensemble classifier based on CSP, autoregressive (AR) and Temporal waves of readiness potential (RP). Only CSP method was applied for subject “al”, “aw” and “ay” but for subject “aa” and “av” combination of all three methods (CSP–AR–RP) was used. Moreover, for subjects with fewer training data (“aw” and “ay”), they used former classified test sample as extended training samples, whereas our proposed approach used only training samples even for subjects with limited training trials. Therefore, it is unfair to compare our simple methods with the first winner.

Table 3.3: Classification accuracy (Mean and Standard deviation in percent) of the proposed approach and other MI classification approaches on Dataset IVa, BCI Competition III.

Studies	Methods	Year	aa	al	av	aw	ay	Mean	Std
Conventional Method	CSP		66.07	96.43	47.45	71.88	49.6	66.28	19.83
Belwafi et al. [252]	WOLA-CSP	2018	66.07	96.07	52.14	71.43	50	67.29	18.54
Arvaneh et al. [230]	SSCSP	2011	72.32	96.42	54.10	70.53	73.41	73.35	15.09
Lotte and Guan [149]	SRCSPP	2010	72.32	96.43	60.20	77.68	86.51	78.63	13.77
Selim et al. [253]	RMS/LDA	2016	69.64	89.29	59.18	88.84	86.90	78.77	13.65
Dai et al. [228]	TKCSP	2018	68.10	93.88	68.47	88.40	74.93	79.17	11.78
Park and Lee [233]	SBRCSP	2017	86.61	98.21	63.78	89.05	73.81	82.69	13.53
Park and Chung [231]	SSS-CSP	2018	74.11	100	67.78	90.07	89.29	84.46	13.05
Selim et al. [138]	CSP/AM-BA-SVM	2018	86.61	100	66.84	90.63	80.95	85.00	12.30
Proposed Method	SR-MDRM		79.46	100	73.46	89.28	88.49	86.13	10.15
Wang et al. [246]	Winner		96.00	100	81.00	100	98.00	94.20	8

Selim et al. [138] used subject specific optimal time interval for CSP feature extraction. Furthermore, they used hybrid bio-inspired algorithms for feature selection and classifier optimization. They achieved 85% classification accuracy, which is slightly (1.13%) less than the proposed approach. One drawback of this approach is that the classifier optimization takes a very long time. Park and Chung [231] used a set of various local channels region to extract CSP features. They used eigenvalue disparity score to select CSP features from the local channel region and support vector machine (SVM) classifier to classify extracted features. They obtained 84.46% accuracy, which is less than the proposed approach by 1.67%. Park and Lee [233] (SBRCSP) focused on regularizing CSP features from filter bank using other subjects training trials. Their results were less than the proposed approach by 3.44%. In the same vein, Dai et al. [228] implemented a “Transfer Kernel CSP” (TKCSP) approach to learn a domain-invariant kernel by directly matching distributions of source subjects and target subjects. Similar to our approach, they employed all 118 channels to obtain 82.69% which is less than our approach by 6.96. Both TKCSP and SBRCSP have the same drawback, as they rely on other subjects’ (source) training trials.

Selim et al. [253] used root mean square (RMS) features for LDA classifier to obtain 78.77% accuracy with 7.36% less than that of proposed approach. Lotte

and Guan [149] penalized CSP objective function to obtain smooth filters to extract features and achieved 78.63%, which is less than proposed approach by 7.50%. Similarly, Arvaneh et al. [230] implemented “Spatially Sparsed CSP” (SSCSP) filters to extract CSP features. Their results were less than the proposed approach by 12.63%. Belwafi et al. [252] used weighted overlap-add (WOLA) algorithm to perform dynamic filtering of EEG-signals for filter bank CSP method. Their method achieved 67.29% classification accuracy, which is less than our approach by 19.85%. Our method improved the mean classification accuracy by 19.85% compared to CSP method.

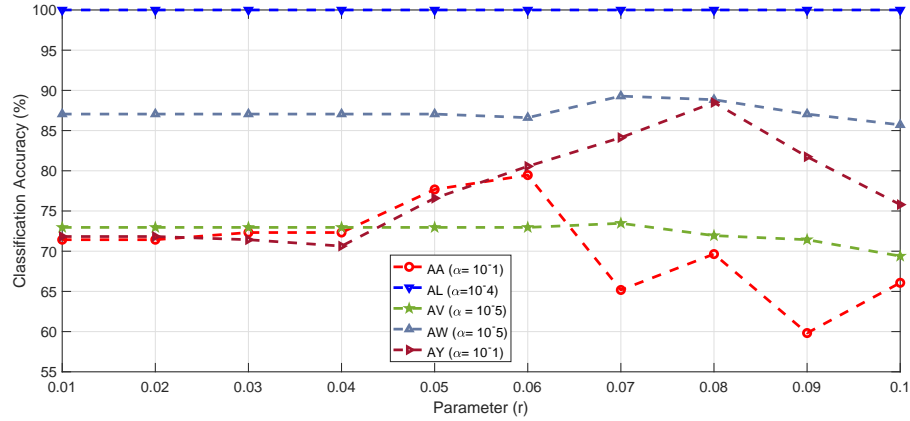


Figure 3.3: Classification accuracy of all subjects from Dataset IVA with respect to hyper-parameter r and best value of α .

Our proposed method shows the highest classification accuracy for subject “al”. Lotte and Guan [149] identified subject AV as BCI illiterate with CSP method because their performance was below 55% (close to random). However, with SR-MDRM, subject “av” achieved 73.46% classification accuracy. Subject “av” would no longer be identified as illiterate. Subject “ay”’s accuracy improved drastically with only 28 training trials. Thus, we might hypothesize that adding spatial prior along with geometry based classification increases accuracy despite the limited amount of training data.

The hyper-parameter r controls the trade off between accuracy and filters sparsity. Therefore, the optimal r value must be selected to increase the accuracy. Figure 3.3 shows the effect of r values on the classification accuracy values of all subjects with fixed (best) α value. Subject “al” reached maximum

accuracy independent from value of r parameter. This is because “al” had sufficient training data. Other subjects performance showed dependence on the value of r and reached maximum accuracy for particular r and α value.

Table 3.4 shows performance of SPD manifold based classification methods for all the subjects in kappa values. As shown in Table 3.4, our method outperformed all existing methods. In addition, SR-MDRM obtained highest kappa value for subjects “al”, “av” and “aw”. As Dataset IVa represents a small sample setting, results obtained on it signify that SR-MDRM is suitable for small sample scenarios.

Table 3.4: The performance of proposed approach and existing Riemannian geometry based approaches on Dataset IVa of BCI Competition III in terms of kappa values.

Studies	Year	AA	AL	AV	AW	AY	MEAN
Barachant et al. [194] MDRM		0.22	0.86	0.25	0.13	0	0.29
Harandi et al. [237] MDRM	2014	0.23	1.00	0.40	0.53	0.82	0.59
Horev et al [72] MDRM	2017	0.62	0.96	0.42	0.68	0.60	0.65
Davoudi et al. [97]-uDPLM	2017	0.57	1.00	0.39	0.64	0.72	0.66
Davoudi et al. [97]-sDPLM	2017	0.63	1.00	0.46	0.66	0.78	0.70
SR-MDRM		0.58	1	0.47	0.79	0.77	0.72

3.5.2 Dataset IIIa, BCI Competition III

Dataset IIIa is also a good test environment for proposed approach, as it also has limited training samples and high EEG signals dimensionality. Table 3.5 presents classification accuracy of proposed method and other existing methods on Dataset IIIa.

Table 3.5: Classification accuracy (Mean and Standard deviation in percent) of the proposed approach and other MI classification approaches on Dataset IIIa, BCI Competition III.

Studies	Methods	Year	k3b	k6b	l1b	Mean	std
Proposed Method	SR-MDRM		100	76.67	100	92.22	13.46
Zhang et al. [234]	TSGSP	2018	99.2	67.2	96.5	87.63	17.74
Belwafi et al. [252]	WOLA-CSP	2018	97.77	61.66	93.33	84.25	19.69
Conventional Method	CSP		95.56	61.67	93.33	83.52	18.95
Horev et al. [72]	HOREV-MDRM	2017	95.56	68.33	85	82.96	13.72
Barachant et al. [194]	MDRM		96.66	60	88.33	81.66	19.21
Lotte and Guan [14]	SRCSP	2011	96.67	53.33	93.33	81.11	24.11

In Figure 3.4, the SR-MDRM method shows a higher mean classification accuracy than the six other methods. In addition, the SR-MDRM method shows the highest classification accuracy for individual subjects. That is, Figure 3.4 clearly shows that the SR-MDRM method is more efficient for binary motor imagery classification than the other six methods. Zhang et al. [234] proposed

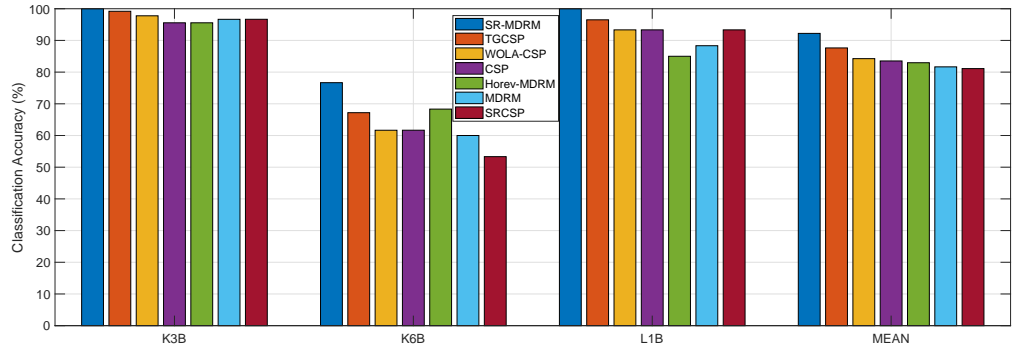


Figure 3.4: Classification accuracy of the proposed approach and other MI classification approaches on Dataset IIIa, BCI Competition III.

temporally constrained sparse group spatial pattern (TSGSP) method; their performance was slightly less than our method. In their study, they simultaneously optimized filter bands and time window to extract CSP features for classification to obtain a mean accuracy of 87.63%. Dataset IIIa is recorded with (60) electrodes, thus covariance matrices dimensionality is less compared to Dataset IVa (118×118). MDRM method's performance improved due to the small size of covariance matrices. As shown in Figure 3.4, it is marginally less

(1.3%) than Horev’s MDRM [72] method, which adapted PCA to map covariance matrices to a lower-dimensional SPD manifold. Interestingly, standard CSP performed better than spatially regularized CSP method proposed [14] on Dataset IIIa. Subject “k6b”’s performance improved drastically with our proposed approach. Subject “k6b” gained 9.47% classification accuracy more than the state-of-the-art method TGCSP [234]. Figure 3.5 shows the classification accuracy values of subject “k6b” according to the hyper-parameters r and α , respectively. Subject “k6b” reached maximum accuracy at $\alpha = (10^{-3})$ and $r = 0.08$ values. It proves our hypothesis that spatial prior and geometry based treatment of data helps achieve the highest classification accuracy under small training sample.

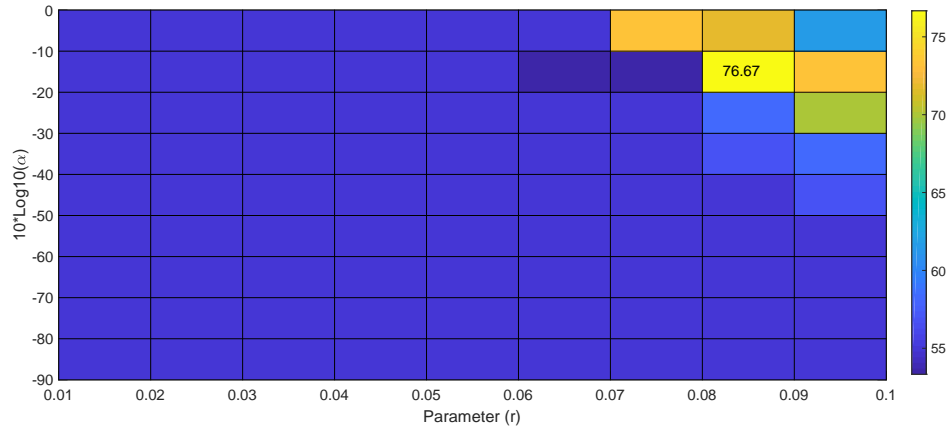


Figure 3.5: Classification accuracy according to hyper-parameter r and $10 \log_{10}(\alpha)$ of the proposed approach on subject $K6B$ from Dataset IIIa, BCI Competition III.

3.5.3 Dataset IIa, BCI Competition IV

Dataset IIa has sufficient training trails per subjects and EEG signals are low dimensional (22 channels), as shown in Table 3.1. It is a good test environment to check for the robustness of our proposed approach under low dimensional and sufficient training samples.

Table 3.6 shows the classification accuracy of existing methods and proposed method (SR-MDRM) on Dataset IVa. For subjects A02, A03, A04 and A07, the proposed method achieved highest accuracy compared to existing methods in

the literature. Similar to our proposed approach, Gaur et al. [254] used Riemannian geometry to classify features obtained through subject specific multivariate empirical mode decomposition method (SS-MEMD). They achieved higher accuracy for subject A09 and mean accuracy was slightly less than proposed approach. Due to lower dimensionality of covariance matrices in Dataset IIa, MDRM method outperformed other methods for subjects A01 and A05.

Table 3.6: Classification accuracy (mean and standard deviation in percent) of the proposed approach and other approaches on Dataset IIa, BCIC IV.

Studies	Methods	Year	A01	A02	A03	A04	A05	A06	A07	A08	A09	Mean	Std
Proposed Method	SR-MDRM		90.21	63.28	96.55	76.38	65.49	69.01	81.94	95.14	93.01	81.22	12.43
Gaur et al. [254]	SS-MEMDBF	2018	91.49	60.56	94.16	76.16	58.52	68.52	78.57	97.01	93.85	79.93	14.14
Barachant et al. [194]	MDRM		91.61	57.03	90.21	73.61	73.94	68.31	75	95.14	90.21	79.45	12.92
Belwafi et al. [252]	WOLA-CSP	2018	86.81	63.19	94.44	68.75	56.25	69.44	78.47	97.91	93.75	78.85	15.15
Lotte and Guan. [14]	SR-CSP	2011	88.89	63.19	96.53	66.67	63.19	63.89	78.47	95.83	92.36	78.78	14.77
standard Method	CSP		88.89	51.39	96.53	70.14	54.86	71.53	81.25	93.75	93.75	78.01	17.01
Raza et al. [255]	TL-CSP ₁	2016	90.28	54.17	93.75	64.58	57.64	65.28	62.5	90.97	85.42	73.84	15.93
Raza et al. [255]	TL-CSP ₂	2016	90.28	57.64	95.14	65.97	61.11	65.28	61.11	91.67	86.11	74.92	15.42

3.6 Conclusion


We propose spatially regularized Symmetric positive definite (SPD) matrices based motor imagery classification method. This method incorporates prior information of EEG electrodes to obtain spatial filters that transform sample covariance matrices into lower dimension and maximize the variance between two motor imagery task in small sample setting. The proposed method takes advantage of geometrical properties of covariance matrices by employing Riemannian distance as pattern recognition metric for classification as it is invariant to any linear invertible transformation. The efficacy of the proposed approach was validated on three public datasets from BCI competition. Our proposed method’s mean classification accuracy is better than other approaches in existing studies on all three datasets.



GRADUATE
RESEARCH
SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Amardeep Singh
Name/title of Primary Supervisor:	Prof. Hans W. Guesgen
In which chapter is the manuscript /published work: Chapter 4	
<p>Please select one of the following three options:</p> <p><input checked="" type="radio"/> The manuscript/published work is published or in press</p> <ul style="list-style-type: none"> • Please provide the full reference of the Research Output: Amardeep Singh, Sunil Lal and Hans W. Guesgen. "Small Sample MotorImagery Classification Using Regularized Riemannian Features" in IEEEAccess, vol. 7, pp. 46858 - 46869, 2019. <p><input type="radio"/> The manuscript is currently under review for publication – please indicate:</p> <ul style="list-style-type: none"> • The name of the journal: • The percentage of the manuscript/published work that was contributed by the candidate: 80.00 • Describe the contribution that the candidate has made to the manuscript/published work: Conceptualization, Methodology, Investigation, Formal analysis, Validation, Writing - Original Draft, Writing - Review & Editing, Visualization. <p><input type="radio"/> It is intended that the manuscript will be published, but it has not yet been submitted to a journal</p>	
Candidate's Signature:	Amardeep Singh <small>Digitally signed by Amardeep Singh Date: 2021.02.26 13:13:11 +13'00'</small>
Date:	26-Feb-2021
Primary Supervisor's Signature:	
Date:	26-Feb-2021

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/publication or collected as an appendix at the end of the thesis.

Chapter 4

Small Sample Motor Imagery Classification Using Regularized Riemannian geometry Features

Quick Summary

Motor imagery-based electroencephalogram brain-computer interface (BCI) performance suffers from huge variations within and across subjects. This is due to different spatial and temporal characteristics among the subjects. To address these variabilities, a large number of labeled subject specific training trials are collected to calibrate systems for new subjects. This results in long calibration time that limits the BCI usage in practice. We focus on reducing calibration time by introducing a Riemannian approach. However, in Riemannian approach the performance degrades in small sample scenario as the dimensionality of covariance matrices is large in comparison to the number of trials. To overcome this limitation, we proposed a new framework that transforms covariance matrices into a lower dimension through spatial filter regularized by data from other subjects.

Related Paper

- **Amardeep Singh**, Sunil Lal and Hans W. Guesgen. "Small Sample Motor Imagery Classification Using Regularized Riemannian Features" in IEEE Access, vol. 7, pp. 46858 - 46869, 2019.

4.1 Introduction

The electroencephalogram (EEG) based brain computer interface (BCI) translates neural activity in the brain into commands to control external devices [2]. Motor imagery (MI) is a BCI paradigm where neural activity is generated at the sensorimotor cortex due to the kinaesthetic imagination of a body part (left/right hand) movement [71]. This neural response varies spatially and temporally across the subjects even for the same motor imagery task [63], [234]. Therefore, for every new subject calibration, trials are collected to understand their neural response pattern to different motor imagery tasks.

The standard approach to do this, first captures temporal and spatial characteristics of a subject in each MI calibration trial with high spatial and temporal resolution. It is done with a multi-channel electrodes system with a high sampling rate. This also results in a high-dimensional EEG signal. Second, through different signal processing techniques, discriminative features from high-dimensional EEG trials are obtained that are fed into machine learning algorithms for MI classification [16].

One of the most effective techniques for feature extraction in motor imagery are common spatial pattern (CSP) filters. CSP aims to determine optimal spatial filters that discriminate two MI task classes [226]. For this, CSP uses sample covariance matrices (SCM) of EEG trials as it encode the spatial information of neural responses to the MI task. Although good performance can be achieved with CSP [256] and deep learning based approaches [257, 258] but it requires a large number of subject specific calibration trial sessions. These calibration sessions are very time consuming and not so user-friendly [259]. Thus, it is highly desirable to reduce or remove the calibration entirely.

In the case of a small EEG calibration trials set, these covariance matrices poorly estimate MI and therefore lead to poor performance of CSP. To address this, Lu *et al.* [227] suggested subject to subject transfer approach (RCSP) that uses information from other subjects involved in similar motor imagery tasks to construct a spatial covariance matrix for the target subject. In the same vein, Dai *et al.* [228] (TKCSP) also used transfer learning approach to learn domain invariant CSP features from source and target subjects. Both these approaches uses data from other subjects. Unlike the above method, Arvaneh *et al.* [230] proposed (SSCSP) a method that does not require data from other subjects to obtain CSP features rather this approach emphasize more on those regions that

have high variances between MI imagery classes, and attenuates the regions with low variances. Similarly, Park and Chung [231] have also focused on those channels that belong to the region that contributes to extracting diverse CSP features and obtaining high accuracy compared to standard CSP. In the same vein, Lotte *et al.* [149] extended conventional CSP method by adding prior information of EEG channels. Their approach emphasizes the fact that neighboring channels measure similar brain signals thus have a similar contribution in the spatial filter. Lotte *et al.* proposed Spatial regularized CSP (SRCSP) aims at regularizing the CSP objective function by penalizing filters which are not spatially smooth, that is for neighboring electrodes which have very different weights.

The performance of the spatial filter depends upon the subject specific frequency band. To address this problem, Ang *et al.* [226] proposed filter bank CSP (FBCSP) obtains features from multiple bandpass filters and spatial filters then employs a feature selection algorithm to select discriminative CSP features for classification. Belwafi *et al.* [252] (WOLA-CSP) choose subject specific frequency by analyzing state of the subject before and during MI task. Park *et al.* [233] extended the filter bank algorithm with regularization techniques (SBRCSP). They used information from other subjects trials to regularized CSP features from each filter bank and then employed feature selection algorithm to extract most discriminative CSP features. Combination of filter bank and regularization significantly improve performance of CSP technique in small sample setting. Selim *et al.* [138] extended CSP approach (CSP/AM-BA-SVM) by finding the optimal time interval for each subject to extract CSP features. Furthermore, they used bio-inspired optimization algorithm (bat algorithm) for feature selection and classifier optimization to boost classification accuracy. Even though it improved performance of CSP algorithm, it takes a very long time to optimize the classifier in the calibration phase. Zhang *et al.* [234], proposed (TGCSP) a method which simultaneously optimizes filter bands and time window within CSP to further boost classification accuracy of MI EEG.

Unfortunately, all the approaches discussed above extract CSP features (log variance) from sample covariance matrices (SCM) into a vector in the Euclidean space. Furthermore, pattern recognition and dissimilarity metrics between features are built only for vectors in Euclidean space. Thus, these approaches fail to notice a very distinctive characteristic of data: their structure

(geometry), or more specific, the manifold space and the interrelation across the manifold space [17]. Data treatment based on the concept of manifolds have been proved to be more effective and adopted in many applications. Barachant *et al.* [194], proposed a method named minimum distance to Riemannian mean (MDRM) that uses geometric properties of the covariance matrix as the EEG descriptor and adopts the Riemannian distance to discriminate among them. The Riemannian distance follows congruence invariance [152], meaning it remains unchanged under linear invertible transformation. MDRM performance is better than the conventional CSP method, but starts deteriorating as the size of covariance matrices increase (i.e. large number of EEG channels). Usually, the dimensionalities of covariance matrices in BCI applications are large in proportion to the number of trials and therefore, MDRM algorithms encounter the curse of dimensionality problem [97].

To address this issue, Horev *et al.* [72] proposed an adaptation of PCA to the space of SPD matrices, which extends the standard definition from Euclidean to Riemannian geometries, preserves more data variance and also maps to a lower-dimensional SPD manifold. Harandi *et al.* [237] learned mapping that maximizes the geodesic distances between inter-class samples and simultaneously minimizes the distances between intra-class samples. This was done via optimization on Grassmann manifolds. This algorithm tries to preserve the local structure of the data by preserving the distance to the local means by considering the geometry of SPD matrices and providing an implicit mapping. In addition it applies the supervised information for embedding to lower-dimensional space. Furthermore, Davoudi *et al.* [97] extended Harandi's work, proposed another dimensionality reduction algorithm for the manifold of SPD matrices which preserves the local structure of data by preserving the distance to the local mean (DPLM). This algorithm can work in a supervised (sDPLM) or unsupervised (uDPLM) manner and projects a high-dimensional SPD manifold to a lower-dimensional one.

From our literature review, we found that CSP is good spatial filtering algorithm which transforms the sample covariance matrices (SCM) of EEG trials into lower dimension and promotes the variance between two motor imagery classes. However, it requires a large number of subject specific calibration trial sessions. A promising technique for reducing or eliminate these calibration session is transfer learning (TL) [259]. Transfer learning is a practice

of using data recorded in one task to increase the performance in another, related task (for a comprehensive review, see [229]). For transfer learning, there is a prior assumption that there exists some shared structure between source and target tasks. The aim is to learn some representation of this structure so in the future similar tasks can be solved easily.

In the context of BCIs, transfer learning become more important as EEG signals are non-stationary, and so in a way every trial can be considered as a new task. The challenge is determining how to transfer some kind of knowledge between different trials. Looking at the literature, transfer learning can be done in two general ways. First way is to find some structure in how the decision rules differ between different sessions or subjects. This is known as rule adaptation based transfer learning technique. Second way, attempts to find some structure in the data that is invariant across data sets, this is known as domain adaptation technique. BCI literature exclusively dominated by domain adaptation based transfer learning approaches [229]. Many transfer learning approaches are tried with CSP, largely based on assumption that there exists a set of linear filters that are invariant across either subjects or sessions. Since covariance matrices are used in CSP, whereas the target domain does not have enough labeled samples to reliably estimate them, a direction to incorporate TL into CSP is to utilize the source domain covariance matrices to obtain regularized covariance matrix for each class(left/right), which is used to find an invariant subspace on which to project the data of new subjects [15]. In this chapter, we propose a method that uses the best of both Euclidean and Riemannian approaches. We used subject to subject transfer properties of Euclidean approach [227] to obtain spatial filter under small sample scenario that reduces the dimension of covariance matrices for Riemannian geometry based classification. We used Riemannian distance as a pattern recognition metric as it follows congruence invariance i.e. it is invariant to linear invertible transformation.

The rest of the chapter is organized as follows. In section II, we discuss the space of the sample covariance matrices (SCM) and the MDRM classification approach in the Riemannian space. Section III presents our proposed classification approach (Regularized MDRM). Section IV, describes the dataset as well as the experiment. In section V, the results of our experiments are discussed and compared to existing studies. Section VI draws the conclusions regarding our proposed method.

4.2 Geometry of SCM matrices

4.2.1 Data Model

Motor imagery BCI systems work in two phases, namely, calibration and feedback. During calibration, subjects were given visual cues to perform a motor imagery task. The acquired high-dimensional signal during calibration phase is then divided into trials based on the cues. These trials are processed through signal processing techniques [16] to extract features for the training classifier. This trained classifier is then used in the feedback phase to classify different motor imagery tasks of the user that are generated randomly by the user without any cue.

A filtered single EEG trial X_i as shown in Figure 4.1 is a matrix of size $N \times T$ where N is the number of Channels and T the number of sampled time points which depend on the sampling frequency of the acquisition device. From this single trial X_i , the normalized sample covariance matrix C_i is calculated as follows:

$$C_i = \frac{X_i X_i^T}{tr(X_i X_i^T)} \quad (4.1)$$

where $tr(\cdot)$ denotes the trace operator of the matrix, and the superscript T denotes the transpose of the matrix. Covariance matrices encode spatial-temporal information of the trial and lie in the symmetric positive definite (SPD) space.

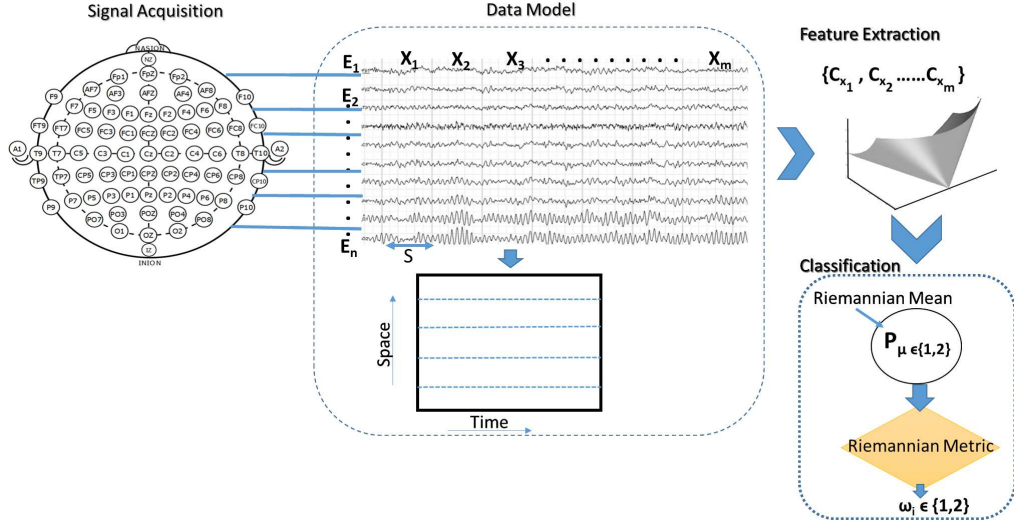


Figure 4.1: Framework for MDRM approach

4.2.2 Riemannian Manifold

An $n \times n$ square matrix C lies in the symmetric positive definite $SPD(n)$ matrices space, if $\{C = C^T, u^T C u > 0, \forall u \in \mathbb{R}^n, \forall u \neq 0\}$. Equivalently, SPD matrices have the following properties:

1. $\forall C \in SPD(n)$, eigenvalues are positive i.e. $\lambda(C) > 0$
2. $\forall C \in SPD(n), C^{-1} \in SPD(n)$ i.e. SPD matrices are invertible

SPD matrices in $SPD(n)$ forms a Riemannian manifold of $m = n(n + 1)/2$ dimensional. A Riemannian manifold \mathcal{M} is a differentiable manifold in which the tangent space at each point is a finite-dimensional Euclidean space [71]. That means, derivatives at a matrix C on manifold lies in a vector space $\mathcal{T}_C \mathcal{M}$, which is the tangent space at that point. The tangent space is lying in the space of symmetric matrices ($S(n), S^T = S$) [71]. A matrix (point) C_i is projected to tangent space $\mathcal{T}_C \mathcal{M}$ using logarithmic mapping $Log_C(C_i)$ as:

$$S_i \in \mathcal{T}_C \mathcal{M} = Log_C(C_i) = C^{1/2} \log(C^{-1/2} C_i C^{-1/2}) C^{1/2} \quad (4.2)$$

where $\log(\cdot)$ is logarithm of SPD matrix and $C \in \mathcal{M}$ is a reference point where the tangent plane $\mathcal{T}_C \mathcal{M}$ is mapped. Similarly, exponential mapping $Exp_C(S_i)$ is used to project back tangent vector S_i from tangent space to manifold. It is given by:

$$C_i \in \mathcal{M} = Exp_C(S_i) = C^{1/2} \exp(C^{-1/2} S_i C^{-1/2}) C^{1/2} \quad (4.3)$$

where $\exp(\cdot)$ is exponential of SPD matrix. Figure 4.2 illustrate Riemannian Manifold and corresponding local tangent space at C .

Riemannian Distance

Riemannian Distance is defined as a length of shortest (called geodesic for curved spaces) and unique path connecting two points C_1 and C_2 in the Riemannian manifold \mathcal{M} . It is given by:

$$R_d(C_1, C_2) = \|\log(C_1^{-1/2}C_2C_1^{-1/2})\|_F = \left(\sum_{i=1}^n \log^2 \lambda_i\right)^{1/2} \quad (4.4)$$

where λ_i 's are the positive eigenvalues of $C_1^{-1/2}C_2C_1^{-1/2}$ and $\|\cdot\|_F$ is the Frobenius norm of the matrix. As shown in Equation (4.5), Riemannian distance (metric) $R_d(C_1, C_2)$ possesses congruence invariance property [152] that means Riemannian metric is invariant to any linear invertible transformation.

$$R_d(A^T C_1 A, A^T C_2 A) = R_d(C_1, C_2) \quad (4.5)$$

where A is an invertible matrix. The Riemannian distance between two matrix in manifold can be approximated in tangent space by approximating the distance between projected tangent vectors through a reference point. To obtain a good approximation of the Riemannian distance, reference point needs to be close to two points in the manifold. Usually, the mean is the most suitable choice for the reference point.

The choice of reference point

The reference SPD matrix is free parameter that defines the point in the manifold \mathcal{M} where the tangent plane is computed [71]. The most common choice of reference matrix C_{ref} is the average of the whole set of covariance matrices. All the averaging approaches such as arithmetic, geometric, and harmonic mean estimate the centrality of manifold \mathcal{M} [240]. In this thesis, we used the geometric mean (also known as Riemannian mean) of SPD matrices that minimizes the sum of squared Riemannian distances. It is given by:

$$\Pi_{\mathcal{M}}(C_1, C_2, \dots, C_I) = \arg \min_{C \in SPD(n)} \sum_{i=1}^I R_d^2(C_i, C) \quad (4.6)$$

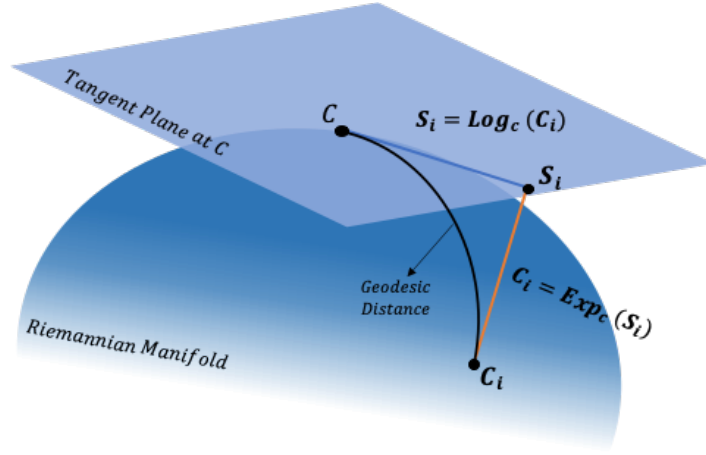


Figure 4.2: Illustration of Tangent space of manifold \mathcal{M} at point C . Tangent vector S_i is the projection of C_i . Geodesic curve illustrate unique and shortest curve connecting C_i and C . The length of geodesic curve is Riemannian distance between C_i and C .

For a manifold of non-positive sectional curvature like $\text{SPD}(n)$, such local minimum exists and is unique. However, there is no closed-form expression to compute the mean and optimisation algorithms must be employed. An efficient iterative algorithm to compute the Riemannian mean of $I \geq 1$ SPD matrices. Chebbi and Moakher [241] provide a detail discourse on the computation of Riemannian mean.

4.2.3 Minimum Distance to Riemannian Mean (MDRM)

MDRM is a classification technique that utilizes the Riemannian mean of each class and its Riemannian distance to covariance matrix of the new trial to anticipate a label for it. The framework of MDRM is shown in Figure 4.1. During this approach, the Riemannian mean is determined for each class utilizing its labelled training trials, then the Riemannian distance of each class is obtained with respect to new trial's covariance matrix. The class mean, which is nearest to new trial's covariance, becomes its label.

$$\text{pred}(C_x) = \underset{\varphi=1,2..C}{\text{arg min}} R_d(C_x, C_\varphi^{\text{II}}) \quad (4.7)$$

where C_φ^{II} is the Riemannian mean of Class φ , C_x is the covariance matrix of

the new trial and $pred(C_x)$ is the prediction of its class label. Algorithm 1 is the pseudo code of the MDRM classification [242]. The MDRM approach is not robust to noise [97], therefore, some filtering over SPD matrices is required. Barachant *et al.* [242] suggested Geodesic filtered MDM (FGMDM) approach which computes set of filters by applying a supervised Fisher geodesic discriminant analysis (FGDA) to the tangent (Euclidean) space projection of covariance matrices. The obtained filters are applied through Geodesic filtering approach [242] over SPD matrices. This filtering operation do not change any dimension of the SPD matrices. Finally, the filtered SPD matrices are used for MDRM classification.

Algorithm 1: MDRM classification [242]

Input: Train data set $\{(X_i, \omega_i)_{m=1}^M, X_i \in \mathbb{R}^{N \times T}$ and Class label $\omega_i \in 1, 2\}$, Test data set $\{(X_i)_{n=1}^N, X_i \in \mathbb{R}^{N \times T}$

Input: Test trial $X_i \in \mathbb{R}^{N \times T}$

Output: The class label (φ_*) for target test data set

- 1 Compute the normalized sample covariance matrices (SCM) C_i of target subject training trials by Eq. (4.1)
- 2 Computer Riemannian mean of class 1 $C_{\pi_1} = \Pi_{\mathcal{M}}(C_i)$ with $i|\omega = 1$
- 3 Computer Riemannian mean of class 2 $C_{\pi_2} = \Pi_{\mathcal{M}}(C_i)$ with $i|\omega = 2$
- 4 Compute the sample covariance matrices (SCM) C_i^{test} of target subject test trial by Eq. (4.1)
- 5 $\delta = R_d(C_i^{test}, C_{\pi_1}) - R_d(C_i^{test}, C_{\pi_2})$
- 6 **if** $\delta \leq 0$ **then**
- 7 $\varphi_x = 1$;
- 8 **else**
- 9 $\varphi_x = 2$;
- 10 **return** φ_x

4.3 Methodology

The conceptual framework of our proposed methodology is shown in Figure 4.3 and the high level pseudocode is described in algorithm 2.

A normalized sample covariance matrix (SCM) [243] C_i for a single trial $X_i \in \mathbb{R}^{N \times T}$ can be obtained using equation (4.1) and its size is $N \times N$. The covariance matrix for each class can be obtain by taking the sum of normalized covariance

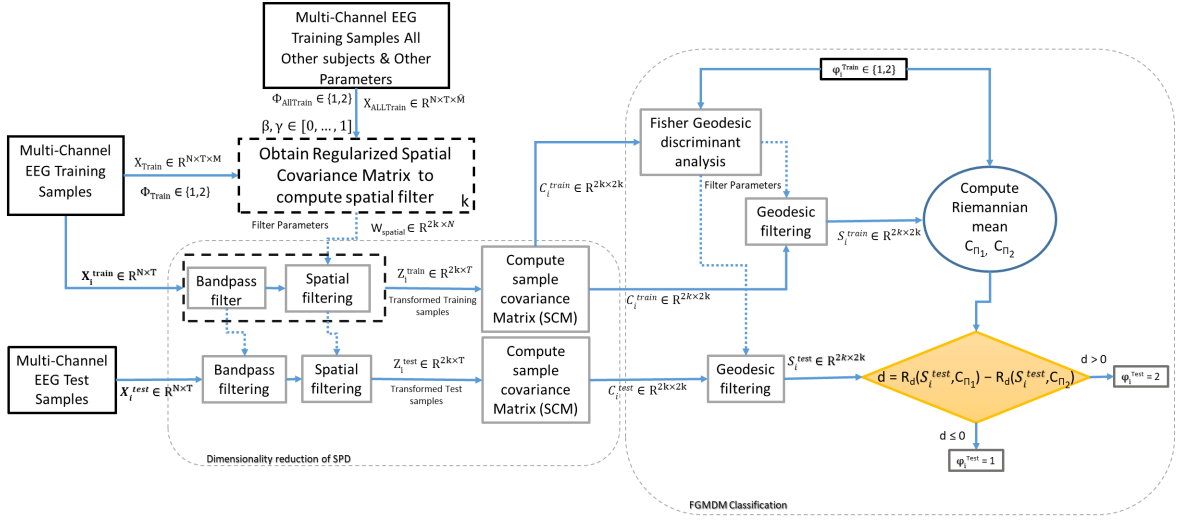


Figure 4.3: A block diagram of proposed approach

matrices for M trials that belong to it. It is calculated as follows:

$$C_\varphi = \sum_{m=1}^M C_{(\varphi,m)} \quad (4.8)$$

where M is the total trial number of each class, and m is the index of the trial ($m \leq M$). φ denotes the class index, and we consider only two classes ($\varphi \in \{1, 2\}$) in this chapter. The generic covariance matrix \hat{G}_φ for regularization is calculated using the EEG collected for the other subjects as follows:

$$\hat{G}_\varphi = \sum_{\hat{m}=1}^{\hat{M}} C_{(\varphi,\hat{m})} \quad (4.9)$$

where \hat{M} is the number of trials in class φ from the other subjects. Based on generic learning, \hat{G}_φ reduces the variance in calculating the regularized covariance matrix. In addition, it provides more stable results [227]. In equation (4.9), \hat{G}_φ is calculated using EEG samples from other subjects excluding the target subject. In equation (4.10), $\hat{C}_\varphi(\beta)$ is calculated using covariance matrices of target subjects training samples as well as other subjects training samples covariance matrices.

$$\hat{C}_\varphi(\beta) = \frac{(1 - \beta) \cdot P_\varphi + \beta \cdot \hat{G}_\varphi}{(1 - \beta) \cdot M + \beta \cdot \hat{M}} \quad (4.10)$$

where β is the first hyper-parameter that controls the variance of the estimated covariance. As the final regularization process, the regularized covariance matrix $\hat{R}_\varphi(\beta, \gamma)$ for each class can be obtained as follows:

$$\hat{R}_\varphi(\beta, \gamma) = (1 - \gamma) \cdot \hat{C}_\varphi(\beta) + \frac{\gamma}{N} \text{tr}[\hat{C}_\varphi(\beta)] \cdot I \quad (4.11)$$

where γ is the second hyper-parameter, that reduces bias caused by small samples. Like CSP, the mixed covariance matrix $\hat{R}(\beta, \gamma)$ is calculated using each class regularized covariance matrix. Then, eigenvalue decomposition is performed as follows:

$$\hat{R}(\beta, \gamma) = \hat{R}_1(\beta, \gamma) + \hat{R}_2(\beta, \gamma) = U\Lambda U^T \quad (4.12)$$

where $\hat{R}(\beta, \gamma)$ is symmetric, U is the eigenvector matrix and Λ is the diagonal matrix with its element arranged in a descending order. The whitening transformation matrix $Q \in \mathbb{R}^{N \times N}$ is computed using equation (4.13) and covariance matrices of each class is transformed using equation (4.14)

$$Q = \Lambda^{-1/2} U^T \quad (4.13)$$

$$S_0 = Q\hat{R}_1(\beta, \gamma)Q^T \text{ and } S_1 = Q\hat{R}_2(\beta, \gamma)Q^T \quad (4.14)$$

Transformed covariance matrices satisfy

$$S_0 + S_1 = I \quad (4.15)$$

where $I \in \mathbb{R}^{N \times N}$ is an identity matrix. Therefore, eigen-decomposition is performed as shown in equation (4.16).

$$S_0 = U_s D_0 U_s^T \text{ and } S_1 = U_s (I - D_0) U_s^T \quad (4.16)$$

Equation (4.16) shows both transformed covariance matrices S_0 and S_1 share the same eigenvectors (U_s) and associated eigenvalues (D_0) in reverse order [260]. We construct the projection matrix $W_p \in \mathbb{R}^{N \times N}$ using equation (4.17)

$$W_p = U_s^T Q \quad (4.17)$$

The EEG signal trial is transformed with $W_{spatial}$ made from the first and last k columns of W_p corresponding to k largest and smallest eigenvalues of S_0 (or S_1)

by using equation (4.18)

$$Z_i = W_{spatial} X_i \quad (4.18)$$

where $Z_i \in \mathbb{R}^{2k \times T}$ is transformed signal corresponding to X_i .

Algorithm 2: Pseudo-code for proposed methodology

- Input:** Target subject's train data set $\{(X_i, \varphi_i)_{m=1}^M, X_i \in \mathbb{R}^{C \times S} \text{ and } \varphi_i \in \{0, 1\}\}$,
 Target subject's test data set $\{(X_i)_{n=1}^N, X_i \in \mathbb{R}^{C \times S}$
- Input:** Other Subjects Training data $\{(X_i, \varpi_i)_{m=1}^M\}$, Regularization coefficients
 $(0 \leq \beta, \gamma \leq 1)$ and parameter k is the no. of pairs spatial filter
- Output:** The class label (φ_*) for target test data set
- // Step I: Obtain regularized spatial filters
- 1 Compute the sample covariance matrices C_i of target subject training trials by Eq. (4.1)
 - 2 Compute the covariance matrix of each class P_φ of target subject training trials by Eq. (4.1) and Eq. (4.8)
 - 3 Compute the generic covariance matrix \hat{G}_φ of each class from other subject's training data according to Eq. (4.1) and Eq. (4.8)
 - 4 Obtain k pairs regularized spatial filter by the following Eq. (4.10), (4.11), (4.12), (4.14), (4.15), (4.16) and Eq. (4.17)
- // Step II: Reduce dimensionality of SCMs
- 5 Obtain the transformed signal for target subject's train Z_i^{train} and test $Z_i^{test} \in \mathbb{R}^{k \times S}$ trials respectively by using Eq. (4.18)
 - 6 Compute the sample covariance matrices (SCM) of transformed target subject's training C_i^{train} and test C_i^{test} trials respectively by Eq. (4.1)
- // Step III: Classifying the SPD matrices
- 7 Obtain FGDA filters using target subject's training trials SCMs C_i^{train} and labels φ_i by Algorithm in [242]
 - 8 Using a Geodesic filtering approach [242], filter target subjects training S_i^{train} and testing trials S_i^{test} covariance matrices are obtained.
 - 9 Using Algorithm 1, classify the filtered SPD matrices S_i^{test} of target subject's test trials.
-

Sample covariance matrices C_i^{train} of filtered EEG trials Z_i from the target's training set are calculated using equation (4.1). These SCMs C_i^{train} are used to obtain FGDA filter for geodesic filtering [242]. After geodesic filtering, filtered SCMs (S_i^{Train} and S_i^{Test}) for target subject's training and test trials are obtained. Finally, filtered SCMs of the target subject's training set (S_i^{Train}) are used in calculating Riemannian mean $C_{\Pi_R \varphi}$ for both motor imagery classes. These Riemannian means are used for MDRM classification [242] of test trials S_i^{Test} .

4.4 Data and Experiments

In order to evaluate the performance of the proposed algorithm in small sample setting, we used the EEG data set IVa from BCI competition III. To ensure the robustness of our method, we evaluated it over two other public benchmark data sets from BCI competitions with a different number of EEG channels. A summary of the three datasets is given in Table 4.1.

Table 4.1: Summary of dataset IVa, dataset IIIa and dataset IIa from BCI competitions (BCIC)

BCI COMPETITION (BCIC)	SUBJECT	TRAIN	TEST
Dataset IVa ,BCIC-III Channels:118 Sampling Rate: 100 Hz	AA	168	112
	AL	224	56
	AV	84	196
	AW	56	224
	AY	28	252
Dataset IIIa,BCIC-III Channels: 60 Sampling Rate: 250 Hz	k3b (45 Trials/task)	90	90
	k6b (30 Trials/task)	60	60
	l1b (30 Trials/task)	60	60
Dataset IIa, BCIC-IV Channels: 22 Sampling Rate: 250 Hz	A01 - A09 (9 Subjects)	144	144

Data set IVa, BCI competition III:

This dataset [246] contains an EEG signals recorded using 118 channels from five healthy subjects (“aa”, “al”, “av”, “aw” and “ay”). This data set contain EEG signals of binary (right hand and foot) motor imagery tasks from each of the five subjects. There are a total 280 trials per subject, among which 140 trials belong to each class. Furthermore, trials are divided into uneven training and testing sets for each subject as shown in Table 4.1 .

Data set IIIa, BCI competition III:

This dataset [246] comprises of 60 channels EEG signals from three subjects (“k3b”, “k6b” and “l1b”) who performed multi-class (right hand, left hand, tongue

and foot) motor imagery (MI) tasks. EEG signals in data set IIIa was sampled at 250 Hz rate with Notchfilter on between frequency 1 and 50 Hz. Subjects “k6b” and “11b” have the same number of trials (i.e. 60 trials/MI task) for each MI task but subject “k3b” has more trials/MI task (90 trials/MI task). In this study, we are using trials only from the left and right hand MI task. The trials from the left/right hand task are divided equally into training and testing set for each subject as shown in Table 4.1.

Data set IIa, BCI competition IV

The BCI competition IV dataset IIa [247] comprised of EEG data from nine subjects, namely (A01 – A09) that perform four types of motor imagery tasks. From the four classes, we considered only two classes (left and right hand) in this study. During the tasks, EEG signals are recorded and sampled at the rate of 250 Hz using 22 EEG and 3 EOG channels. Only EEG channels are selected for this study. All subjects performed two sessions, one for training and the other for evaluation. The total number of trials per session are 288, with 72 trials per class. As we are considering a binary problem. Both sessions are therefore comprised of only 144 trials.

4.4.1 Experimental Setup

This study was carried out using a Windows 10 computer with specification Intel(R) Core TM i5 – 6500 CPU @3.20 GHz with 8 GB RAM. All the conventional methods (CSP , R-CSP and MDRM) including the proposed algorithm, were implemented in Matlab R2018a.

The work comprised of six steps. Firstly, for the dataset mentioned above, we used a time segment from 0.5 second to 2.5 seconds after the visual cue instructing subjects for the MI task. Therefore, each trial of data set IVa, data set IIIa and data set IIa contains 200, 500 and 500 sample points respectively. Secondly, through fifth order Butterworth bandpass IIR filter each trial of all datasets is filtered in frequency range 7 – 30Hz [249]. Thirdly, for each subject three pairs of regularized spatial filters are learnt from the first and last rows using hyper-parameters $\beta, \gamma \in [0, 0.1, \dots, 1]$ and training data available from other subjects. In the fourth step, the signal is projected using a learned regularized spatial filter; this step reduces the dimensionality of the signal. In

the fifth step, we computed covariance matrices for each trial in the projected space to obtain FGDA filters and Riemannian mean for each MI class. Lastly, both training and testing trials covariance matrices were filtered through FGDA filter and finally the label is assigned to each test trial based on its covariance matrices minimum distance from the respective Riemannian mean of MI classes. For CSP, RCSP and MDRM, we used the same time-segment, filter order and frequency band as described for the proposed method

4.4.2 Evaluation Metrics

To evaluate the performance of proposed method, we used classification accuracy and kappa coefficient as evaluation metrics. In binary classification case, accuracy can be calculated as described in Equation (4.19).

$$Accuracy = \frac{a + b}{a + b + c + d} \quad (4.19)$$

where a is the number of positive samples correctly identified, b is the number of negative samples correctly identified, c is the number of negative cases incorrectly identified, and d is the number of positive cases incorrectly identified. Kappa coefficient compares the accuracy of the system to the accuracy of a random system. It is defined as

$$kappa = \frac{observedAccuracy - randomAccuracy}{1 - randomAccuracy} \quad (4.20)$$

where random accuracy [251] is given by:

$$randomAccuracy = \left(\frac{a + b}{a + b + c + d}\right) * \left(\frac{a + c}{a + b + c + d}\right) + \left(\frac{c + d}{a + b + c + d}\right) * \left(\frac{b + d}{a + b + c + d}\right) \quad (4.21)$$

4.5 Results and Discussion

We evaluated the performance of the proposed approach (R-MDRM) on the three datasets, and compared it with conventional CSP, MDM and RCSP methods result as well as benchmark results reported in literature.

Table 4.2: Classification accuracy of the proposed approach and other conventional approaches on dataset IVa, BCI Competition III

Dataset IVa				
SUBJECT	CSP	RCSP	MDRM	R-MDRM (β, γ)
AA	66.07	66.96	58.03	81.25
AL	96.43	96.43	89.28	100
AV	47.45	63.27	59.4	76.53
AW	71.88	71.88	61.16	87.05
AY	49.6	84.29	48.41	91.26
MEAN	66.28	76.56	63.25	87.21

4.5.1 Dataset IVa, BCI competition III

Table 4.2 shows the classification accuracy of all conventional methods and proposed method (R-MDRM) on dataset IVa. We have highlighted the highest accuracy for each subject. Table 4.3 shows regularization hyper-parameters used in R-MDRM classification for all subjects belonging to different datasets.

Table 4.3: Regularization hyper-parameters β, γ of R-MDRM for all subjects belonging to different datasets

PARAMETERS	Dataset IVa					Dataset IIIa			Dataset IIa								
	AA	AL	AV	AW	AY	K3B	K6B	L1B	A01	A02	A03	A04	A05	A06	A07	A08	A09
β	1	0	0.8	0	0.1	0.3	0.3	0	0	0	0.4	0.1	0.6	0	0	0.2	0
γ	0.3	0	0.3	0	0.1	0.4	0.9	0	0	0	0	0.2	0.1	0	0	0	0

As shown in Table 4.2, the R-MDRM method improves the mean classification accuracy by 20.93%, 10.65%, and 23.96% compared to CSP, RCSP and MDRM methods respectively. In addition, R-MDRM shows the highest accuracy for all subjects in dataset IVa. As shown in Table 4.2, “al” showed 100% accuracy with parameters $\beta, \gamma = 0$. This makes sense as “al” had sufficient training data, therefore, shows low dependence on covariance matrices from other subject’s trials. Interestingly, subject “aw” classification accuracy

does not improve with covariance matrices from other subjects as shown in Figure 4.4.

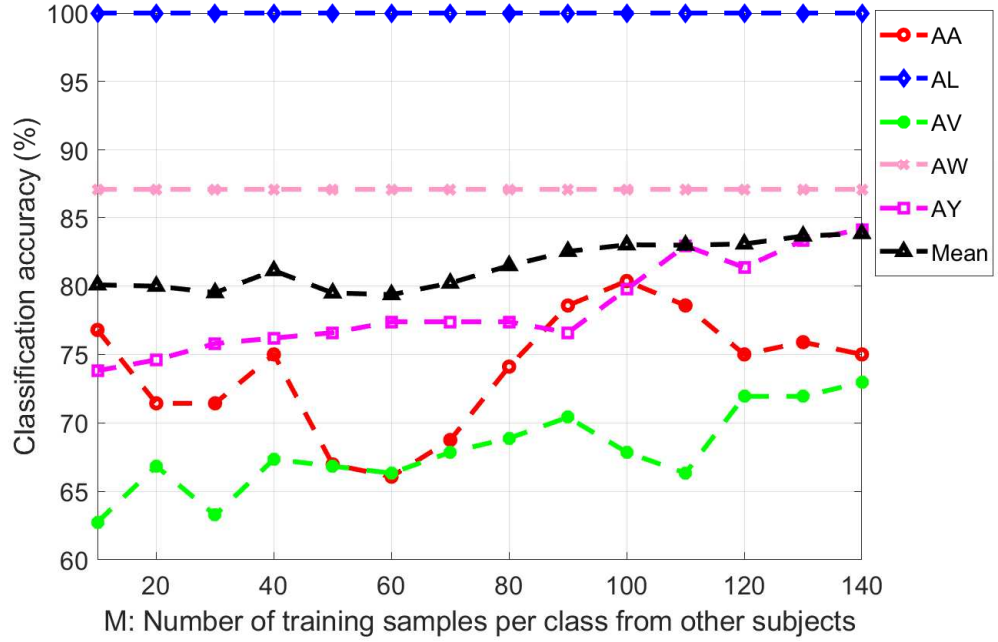


Figure 4.4: Accuracy according to number of trials (M) from other subjects

Figure 4.4 shows the classification accuracy values of all subjects according to the number of training trials per class used from other subjects to regularize the class covariance matrix. Subjects “al” and “aw” have shown low dependence on other subject’s trials and thus reach their maximum accuracy independent from value of M as shown in Figure 4.4. In the case of $M = 10$, our proposed method achieves the mean accuracy of 80.08% which is higher than all the conventional methods. For $M = 140$ per class, our method achieves a mean accuracy of 83.83%. At $M = 140$, subjects “ay” and “av” achieves 72.96% and 84.13% respectively. With more trials ($M > 140$) from other subjects, performance of R-MDRM further improves for subjects “aa”, “ay” and “av”. Subject “av” achieved 75.51% accuracy with the $\beta = 0.8$ and $\gamma = 0.3$ whereas subject “ay” achieved 91.26% classification accuracy with 28 training trials and hyper-parameters $\beta, \gamma = 0.1$.

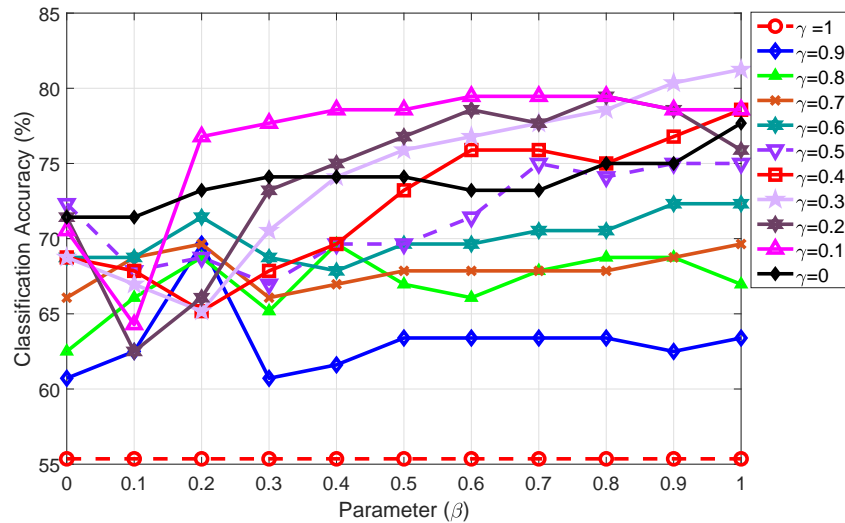


Figure 4.5: Accuracy of AA with respect to hyper-parameters β, γ values

Figure 4.5 shows R-MDRM classification accuracy for subject “aa” according to different β and γ hyper-parameter values. Subject “aa” achieved the highest accuracy at $\beta = 1$ and $\gamma = 0.3$. This shows that using geometrical properties of covariance matrices, along with subject to subject transfer helps in improving the performance. With 28 training instances subject “ay” had shown huge improvement compared to conventional and state of the art methods. so using t-SNE [261], we have visualized the effect of spatial filters obtained through transfer learning on covariance matrices of test trials from subject “ay” in Figure 4.6. Spatially filtering done through filters obtained by transfer learning maximizes the variance ratio between two MI classes in comparison to ordinary filter obtained by CSP algorithm and therefore improved the classification accuracy.

Figure 4.7 shows electrode weights of spatial filters obtained for subject “av” and “ay”. In general, these plots show that filters obtained through normal CSP algorithm appears messy, with large weights on various unexpected brain regions from a neurophysiological point of view. On the contrary, spatial filters obtained through transfer learning are generally smoother and physiologically more relevant, with strong weights over the motor cortex areas, as expected from the literature [262]. This suggests that transfer learning algorithms lead to filters that are neurophysiologically more plausible and as such more

interpretable.

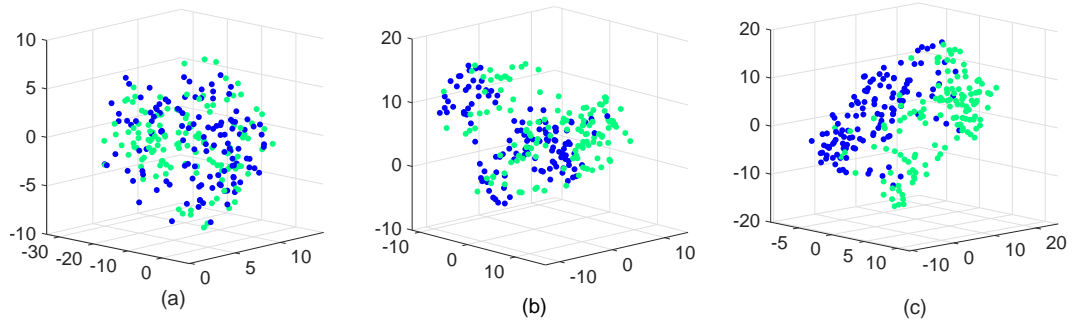


Figure 4.6: Covariance matrices of subject “ay” test trials (a)without spatial filtering,(b) After spatial filtering done through CSP filters and (c) After spatial filtering done through filters obtained by transfer learning, visualize using t-SNE

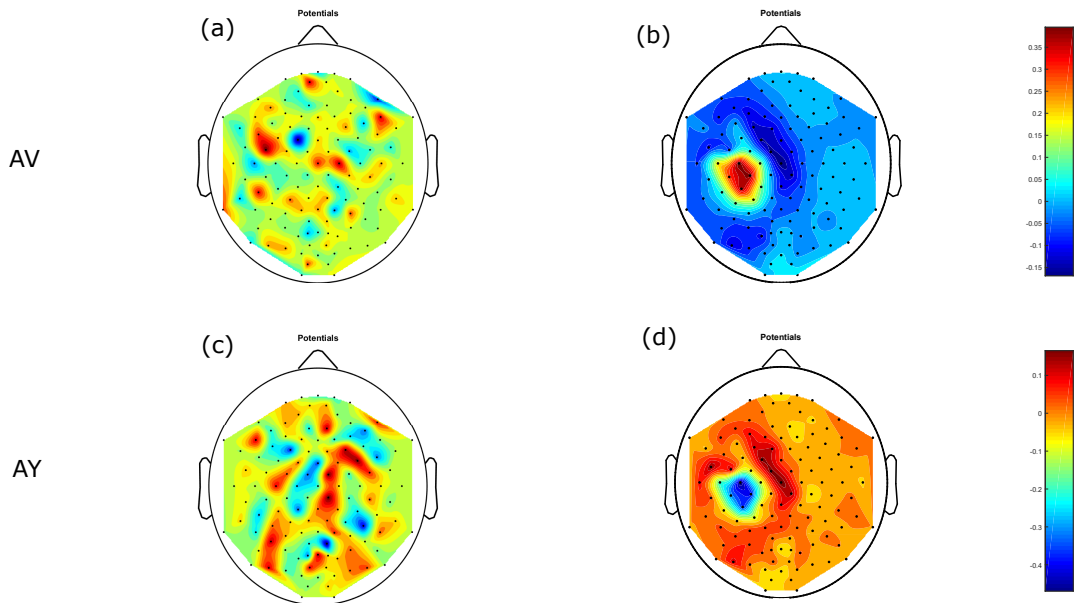


Figure 4.7: Electrode weights for filters obtained for different subjects, (a) and (c) shows electrode weights of filters obtained though CSP algorithm for subjects “av” and “ay” respectively. Similarly (b) and (d) shows electrode weights for filters obtained through transfer learning based CSP algorithm for subjects “av” and “ay” respectively

As shown in Table 4.2, RCSP has shown good performance in comparison to

CSP and MDRM in small sample situations. In particular, for subject “av” and “ay” case where training trials were limited to 84 and 28 instances. For “aa”, “al” and “aw”, RCSP has shown the same performance as CSP. MDRM was the worst performing method on dataset IVa, as shown in Table 4.2. The reason behind this is high dimensional covariance matrices resulting from use of 118 electrodes to record EEG signals. End results is 118×118 covariance matrix per trial for classification.

As shown in Table 4.4, our method outperformed the existing studies in the literature. Belwafi et al. [252] proposed weighted overlap-add (WOLA) algorithm for dynamic filtering of EEG-signals under filter bank based CSP(FBCSP) method. Their results were lower than the proposed approach by 19.92%. Arvaneh et al. [230] implemented “Spatially Sparsed CSP” (SSCSP). They extracted CSP features by emphasizing more on those brain regions that had high variances between MI imagery classes. They achieved 73.50% classification accuracy, which is less than proposed approach by 13.71%. In the same vein, Lotte and Guan [149] penalized CSP objective function through prior information of EEG channels to obtain smooth filters for MI classification. They achieved 78.63% accuracy, which is less than proposed approach by 8.58%. Selim et al. [253] used LDA classifier over root mean square (RMS) features to achieve 78.77% accuracy that is 8.44% lower than proposed approach.

Dai et al. [228] also proposed a transfer learning based CSP approach that learned a domain-invariant kernel by directly matching distributions of source and target subjects respectively. Like our proposed approach, they also employed all 118 EEG electrodes to achieve 82.69% classification accuracy, which is less than our approach by 8.04%. In the same vein, Park and Lee [233] employed transfer learning through regularizing CSP features obtained through filter bank by using other subjects training trials. Their results were less than the proposed approach by 4.52%. She et al. [263] employed CSP approach to filter raw EEG signals and used Fisher discrimination to obtain sparse coding coefficients from the filtered data. Theses discriminative coefficients feed to extreme learning machine (ELM) to achieve 80.68% classification accuracy. They even increased the hidden layers in ELM and achieve classification accuracy 79.33%, which less than proposed approach by 7.88%.

Table 4.4: Classification accuracy of the proposed approach and existing Euclidean approaches on dataset IVa, BCIC III

STUDIES	METHODS	YEAR	AA	AL	AV	AW	AY	MEAN
Belwafi et al. [252]	WOLA-CSP	2018	66.07	96.07	52.14	71.43	50	67.29
Arvaneh et al. [230]	SSCSP	2011	72.32	96.42	54.10	70.53	73.41	73.50
Lotte and Guan [149]	SRCSF	2010	72.32	96.43	60.20	77.68	86.51	78.63
Selim et al. [253]	RMS/LDA	2016	69.64	89.29	59.18	88.84	86.90	78.77
Dai et al. [228]	TKCSP	2018	68.10	93.88	68.47	88.40	74.93	79.17
She et al. [263]	H-ELM	2018	63.39	98.39	64.08	85.67	85.16	79.33
She et al. [263]	FDDL-ELM	2018	61.70	100	73.88	88.17	79.64	80.68
Park and Lee [233]	SBRCSF	2017	86.61	98.21	63.78	89.05	73.81	82.69
Park and Chung [231]	SSS-CSP	2018	74.11	100	67.78	90.07	89.29	84.46
Selim et al. [138]	AM-BA-SVM	2018	86.61	100	66.84	90.63	80.95	85.00
Singh et al. [150]	SR-MDRM	2019	79.46	100	73.46	89.28	88.49	86.13
Park et al. [155]	FBRCSP	2018	91.07	94.64	75	76.78	93.65	86.23
Proposed Method	R-MDRM		81.25	100	76.53	87.05	91.26	87.21

Park and Chung [231] extracted CSP features under small sample scenario by using set of local channels region. They selected useful CSP features from the local channel region by using eigenvalue disparity score in conjunction with support vector machine (SVM) to classify selected local channel features. They obtained 84.46% accuracy, which is less than the proposed approach by 2.75%. Selim et al. [138] extracted CSP features through subject specific time interval and used hybrid bio-inspired (bat) algorithms for feature selection and classifier optimization. They obtained 85% classification accuracy, which is less than the proposed approach by 2.21% . Even though it obtains good accuracy, this approach requires classifier optimization that takes a very long time. Singh et al. [150] used spatially regularized filter to reduce dimensionality of covariance matrices and employed Riemannian distance as pattern recognition metric to obtain 86.13% which is slightly (1.08%) less than our proposed approach. Park et al. [155] extracted regularized CSP features from filter bank then selected useful features based on mutual information for ensemble based classification. They obtained 86.23% classification accuracy, which is less than the proposed approach by is 0.98% .

Lastly, we compared our method with the existing Riemannian manifold based method used for small sample setting in literature. Table 4.5 shows that our method gave the highest kappa values for all the subjects except “aa” and “av”, where supervised dimensionality reduction technique (sDPLM) [97] and spatially regularized MDRM shows better performance respectively. As data set

Table 4.5: The performance of proposed approach and existing Riemannian geometry based approaches on Dataset IVa of BCI Competition III in terms of kappa values

STUDIES	AA	AL	AV	AW	AY	MEAN
Barachant et al. [194]-MDRM	0.22	0.86	0.25	0.13	0	0.29
Harandi et al. [237]MDRM	0.23	1.00	0.40	0.53	0.82	0.59
Horev et al [72] MDRM	0.62	0.96	0.42	0.68	0.60	0.65
Davoudi et al. [97]-uDPLM	0.57	1.00	0.39	0.64	0.72	0.66
Davoudi et al. [97]-sDPLM	0.63	1.00	0.46	0.66	0.78	0.70
Singh et al. [150] SR-MDRM	0.58	1	0.47	0.79	0.77	0.72
R-MDRM	0.62	1	0.53	0.74	0.82	0.74

Table 4.6: Classification accuracy of the proposed approach and other conventional approaches on dataset IIIa, BCIC III

Dataset IIIa				
SUBJECT	CSP	RCSP	MDRM	R-MDRM
K3B	95.56	98.89	96.66	97.78
K6B	61.67	45	60	75
L1B	93.33	93.33	88.33	100
MEAN	83.52	79.07	81.66	90.93

IVa represents a small sample setting, results obtained on it signifies that R-MDRM is suitable for small sample scenarios.

4.5.2 Dataset IIIa, BCI competition III

Dataset IIIa is also good test environment for proposed approach as it also have limited training samples and high EEG signals dimensionality. Table 4.6 reports the results of conventional methods and proposed method on dataset IIIa. On an average, R-MDRM has outperformed the CSP, RCSP and MDRM methods by 6.85%, 11.3%, and 8.71% respectively. Our proposed method achieved the highest accuracy for subject “k6b” and “l1b”. Subject “l1b” achieves 100% accuracy with regularization hyper-parameters $\beta, \gamma = 0$. When β, γ hyper-parameters are zero, it indicates that corresponding filter is obtained using normal CSP method (no

regularization), thus shown low dependence on other subjects. Subject “k6b” shows huge improvement compared to conventional methods as shown in Table 4.6. Therefore, we obtained the scalp plot of subject “k6b” to visualize the effect of transfer learning on filter quality. As shown in figure 4.8 filter obtained through transfer learning based algorithm have strong weights over the motor cortex areas in contrast to filter obtained through normal CSP.

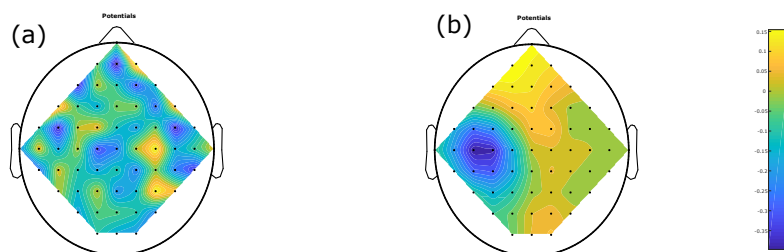


Figure 4.8: Electrode weights for filters obtained for subjects “k6b”, (a) shows electrode weights of filter obtained though CSP algorithm and (b) shows electrode weights for filter obtained through transfer learning based CSP algorithm

Table 4.7 presents the classification accuracy of our proposed approach in comparison with other existing state of the art methods in the literature. Our approach outperformed other subjects in case of subjects “k6b” and “l1b”. Subject “k6b” shown better performance under optimized filter band and time window as discussed in paper [234].

Table 4.7: Classification accuracy of the proposed approach and existing approaches on dataset IIIa, BCIC III

STUDIES	METHODS	YEAR	k3B	k6B	L1B	MEAN
R-MDRM			97.78	75	100	90.93
Zhang et al. [234]	TGCSP	2018	99.2	67.2	96.5	87.63
She et al. [263]	FDDL-ELM	2018	97.78	68.00	96.83	87.54
She et al. [263]	H-ELM	2018	98.56	60.00	98.33	85.63
Belwafi et al. [252]	WOLA-CSP	2018	97.77	61.66	93.33	84.25
Horev et al. [72]	MDRM	2017	95.56	68.33	85	82.96
Lotte and Guan [14]	SRCSF	2011	96.67	53.33	93.33	81.11

Table 4.8: Classification accuracy of the proposed approach and other conventional approaches on dataset Ila, BCIC IV

Subject	A01	A02	A03	A04	A05	A06	A07	A08	A09	MEAN
CSP	88.89	51.39	96.53	70.14	54.86	71.53	81.25	93.75	93.75	78.01
RCSP	86.11	60.42	93.75	56.94	49.31	65.28	81.25	93.75	88.19	75
MDRM	91.61	57.03	90.21	73.61	73.94	68.31	75	95.14	90.21	79.45
R-MDRM	91.61	63.28	97.02	72.91	64.08	69.71	81.25	96.52	92.30	80.98

Table 4.9: Classification accuracy of the proposed approach and existing methods in literature on dataset Ila, BCIC IV

STUDIES	METHODS	YEAR	A01	A02	A03	A04	A05	A06	A07	A08	A09	MEAN
Proposed Approach	R-MDRM		91.61	63.28	97.20	72.91	64.08	69.71	81.25	96.52	92.30	80.98
Gaur et al. [254]	SS-MEMDBF	2018	91.49	60.56	94.16	76.72	58.52	68.52	78.57	97.01	93.85	79.93
Belwafi et al. [252]	WOLA-CSP	2018	86.81	63.19	94.44	68.75	56.25	69.44	78.47	97.91	93.75	78.85
Lotte and Guan [149]	SRCSF	2011	88.89	63.19	96.53	66.67	63.19	63.89	78.47	95.83	92.36	78.78
Wang et al. [49]	AX-LSTM	2018	75.12	71.38	72.24	72.92	82.62	69.64	88.98	80.28	75.07	76.47
Raza et al. [255]	TLCS _{D1}	2016	90.28	54.17	93.75	64.58	57.64	65.28	62.5	90.97	85.42	73.84
Raza et al. [255]	TLCS _{D2}	2016	90.28	57.64	95.14	65.97	61.11	65.28	61.11	91.67	86.11	74.92
Lu et al. [264] [49]	FDBN	2017	71.08	55.56	76.87	65.62	69.08	64.98	71.68	92.37	83.38	72.18

4.5.3 Dataset Ila, BCI competition IV

Dataset Ila has sufficient training trials per subjects and EEG signals are low dimensional (*22 channels*) as shown in Table 4.1. It is a good test environment to check for the robustness of our proposed approach under low dimensional and sufficient training samples. Table 4.8 presents the results of standard methods and R-MDRM on dataset Ila. R-MDRM improved average classification accuracy by 2.97%, 5.98%, and 1.53% in comparison with CSP, RCSP and MDRM respectively. As all subjects have sufficient data therefore most of them (“A01”, “A02”, “A06”, “A07” and “A09”) have shown no dependence on other subjects covariance matrices as shown by regularization hyper-parameters in Table 4.3. Similarly, due to sufficient training samples, subjects “A03” and “A08” shown no dependence on parameter γ that reduces bias in small sample situation as $\gamma = 0$ according to Table 4.3. R-MDRM achieved highest accuracy for subjects “A02”, “A03”, “A07” and “A08” as shown in Table 4.8. MDRM performs better than CSP and RCSP in dataset Ila due to low dimensional EEG signal. For subjects “A01”, “A04” and “A05” MDRM outperforms all the methods as shown in Table 4.8. Table 4.9 reports the comparison of R-MDRM with existing methods in literature on dataset Ila under binary classification

(left/right). In case of 4 subjects R-MDRM performed better compared to other methods as shown in Table 4.9 where as for subjects “A04”, “A08” and “A09”, Gaur *et al.* [254] method outperforms R-MDRM by 3.25%, 0.49% and 1.55% respectively. Similar to our method, the subject specific multivariate empirical mode decomposition (SS-MEMD) [254] uses Riemannian geometry to classify features obtained through Euclidean method. subject ”A05” and ”A07” performance increase under the deep learning method proposed by Wang *et al.* [49].

4.6 Conclusion

In this chapter, we proposed a Symmetric positive definite (SPD) matrices based motor imagery classification method that not only is suitable for small sample setting, but also reduces the dimensionality of SPD matrices. This method incorporates inter and intra-subject variabilities through generic learning based regularization technique to obtain spatial filter. This spatial filter is used to maximize the variance ratio between two motor imagery task in small sample setting. The proposed method takes advantage of geometrical properties of covariance matrices in Riemannian Manifold to improve classification accuracy. We compared our method with conventional method and existing studies over all three datasets from BCI competition. Our method outperforms the conventional methods and other methods in literature in small sample setting. Moreover, R-MDRM has shown better performance on other datasets in comparison with existing state of the art.

Chapter 5

Statistical Testing and Discussion

Quick Summary

In the BCI pipeline, the calibration time is decreased by reducing the number of training trials for each subject. In this chapter we investigate the impact of limiting the training trials for each subject on the classification accuracy. In order to do so, a classical offline training-test framework is used where a bootstrap technique systematically reduces the number of trials, and we check our proposed pipeline performance with respect to the standard BCI pipeline. Moreover, we statistically verify our results and discuss the reasoning behind the performance of our proposed methods.

5.1 Introduction

All the manifold based pipelines discussed so far are designed with the goal of reducing the calibration time required by motor imagery (MI) based brain-computer interfaces (BCIs). We can gain meaningful insights into the performance of the proposed pipelines by systematically reducing training trials for each subject. In order to do so, we used classical offline train-test settings with bootstrap sampling (to choose training and test samples) rather than to assess the ability of the proposed methods in limited training trials.

The rest of the chapter is organized as follows. In Section II, we present our experiment methodology for reducing calibration settings. Section III, the results of the experiment are discussed and compared to existing state-of-the-art Riemannian methods. Section IV draws the conclusions.

5.2 Methodology

In order to evaluate the performance of the proposed algorithm in reduced calibration setting, we used bootstrap sampling technique. The bootstrap offers an easy and effective way to estimate the distribution of a statistic via simulation, by drawing (or generating) new samples from an existing sample with replacement.

Using the bootstrap, we can estimate accuracy with training trials 10, 20, 30, ... 80 and compute the standard error of the mean accuracy and confidence intervals. In a nutshell, the bootstrap procedure can be described as follows:

1. Draw a sample training trials with replacements from a subject's EEG trials.
2. Compute the accuracy from the test trials based on a machine learning model trained on drawn sample training trials.
3. Repeat steps 1-2 for 20 times, for each training size in 10, 20, 30, ... 80
4. Compute the standard deviation (standard error of the mean accuracy of the statistic)

In simple terms, we can interpret the bootstrap by means of drawing a potentially endless number of (new) samples from a population by re-sampling the original

dataset. This would have provided a more suitable framework to evaluate the superiority of the proposed methods as compared to the existing manifold method for all the subjects.

5.3 Results and Discussion

In order to evaluate the performance under reduced calibration settings, two proposed pipelines are tested on three publicly available databases comprising five, three and nine subjects, respectively, for a total of 17 subjects. All three public datasets are recorded on a different number of EEG channels (118, 60 and 22 respectively).

In the previous chapters, hyper-parameters have been optimized through a grid search for each subject that led to an increase in performance. In this chapter, the hyper-parameters are fixed for all the subjects in the different datasets. In the proposed sr-MDRM pipeline, we have used r hyper-parameter equal to 0.08 that defines the size of the neighborhood considered for smoothing filter and α equal to 10^{-1} that defines the level of spatial smoothness the filters should reach. Similarly, for the proposed r-MDRM pipeline, we have used γ equals to 0.2 and β equal to 0.3, respectively. Here γ hyper-parameter shrinks the initial covariance matrix estimate towards the identity matrix to counteract a possible estimation bias due to a small training set. β shrinks the initial covariance matrix estimate towards a generic covariance matrix (defined in equation 4.9), to obtain a more stable estimate. This matrix is typically built by using signals from several subjects whose EEG data has been recorded previously.

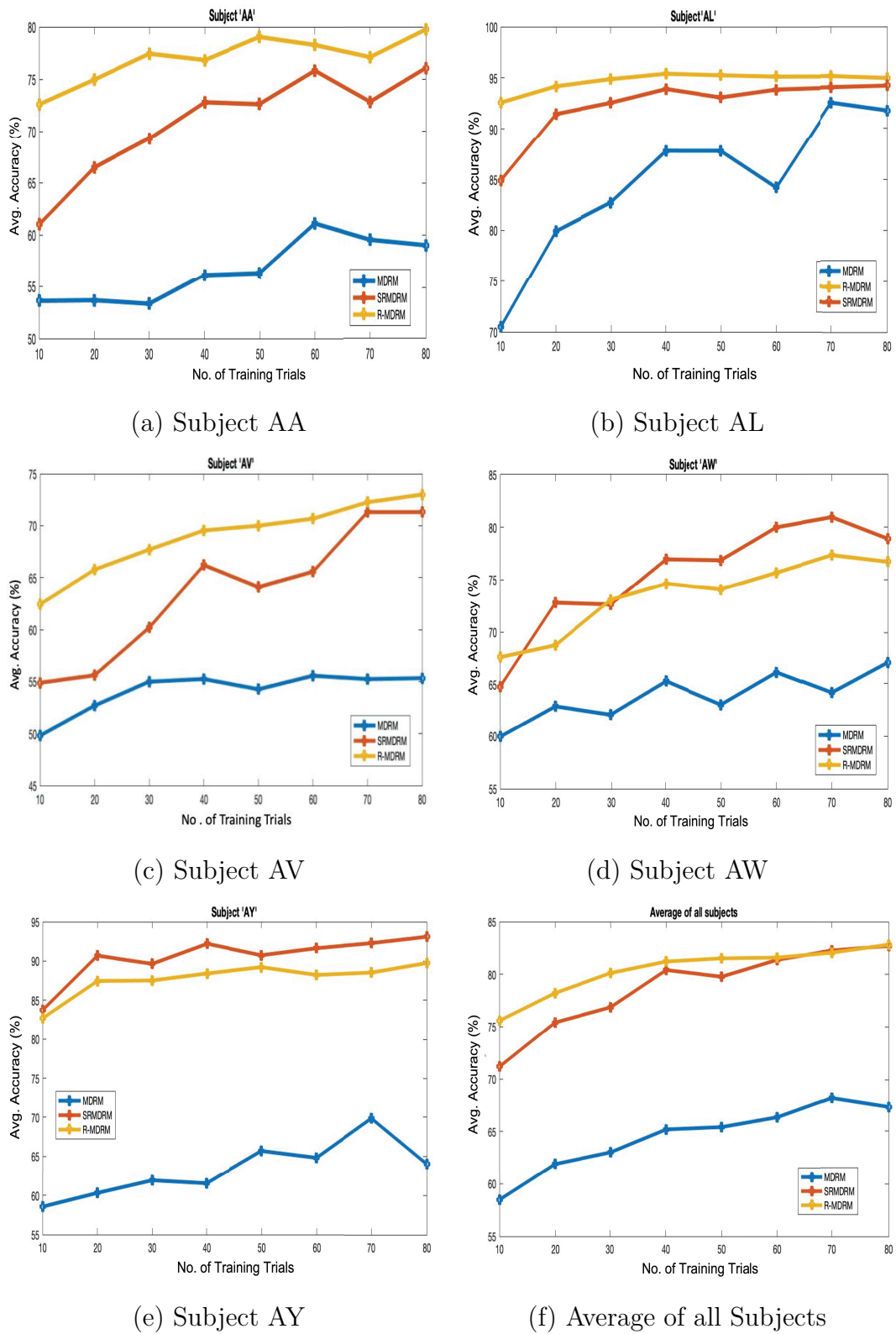


Figure 5.1: Average classification accuracies according to the number of training data available for the target subject.

5.3.1 Data set IVa, BCI competition III

I have compared the performances obtained by our proposed pipelines with the MDRM pipeline for dataset IVa in the Figure 5.1. On the data set used, my results suggested that proposed pipelines improve the accuracy of the BCI, especially when the amount of training data for the target subject is small. Interestingly for all training trials (except subject “AL”), MDRM average classification is below 70 percent which gives the false impression of BCI illiteracy in subjects [29, 165, 166]. This is because dataset IVa is recorded with 118 electrodes that result in a high-dimension covariance matrices (118×118). High-dimensional SPD matrices are ill-conditioned with respect to inversion, jeopardizing the numerical stability of all Riemannian geometry manipulations, which are based on spectral functions of the eigenvalues such as the logarithm, inverse, etc. [6, 19].

As shown in Figure 5.1 (f), the proposed pipeline’s average classification accuracy for all subjects is greater than or equal to 75 percent. This is because the spatial filters in the proposed pipelines reduces the dimension of the covariance matrices that lead to the numerical stability of all Riemannian geometry manipulations in the pipeline. At the same time, spatial filters maximizes the variance ratio between the two MI classes, which also results in a better performance of the manifold-based classification model.

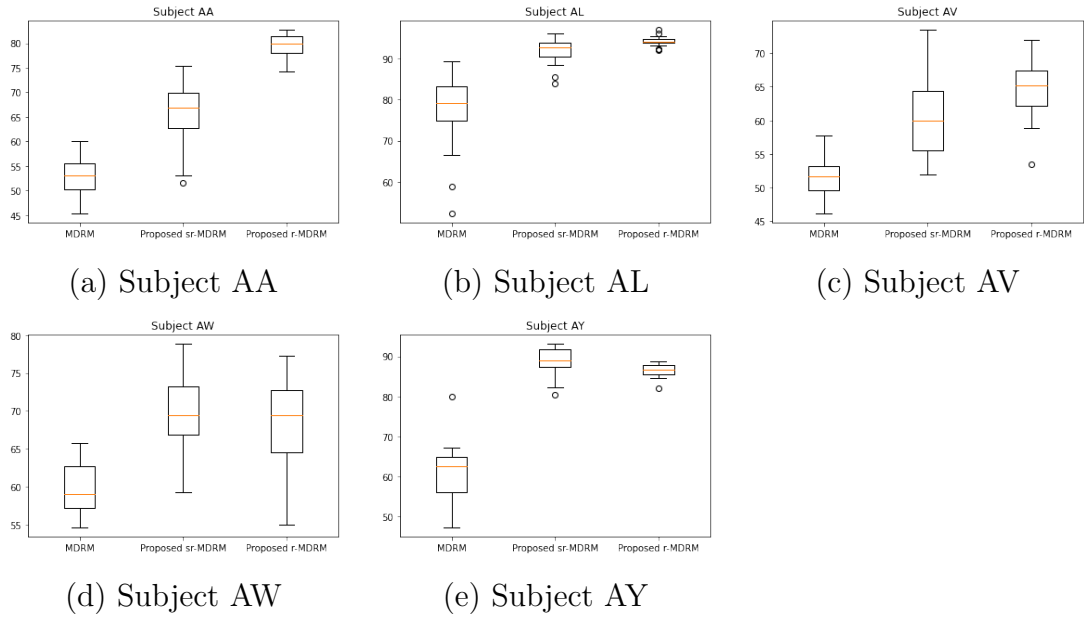
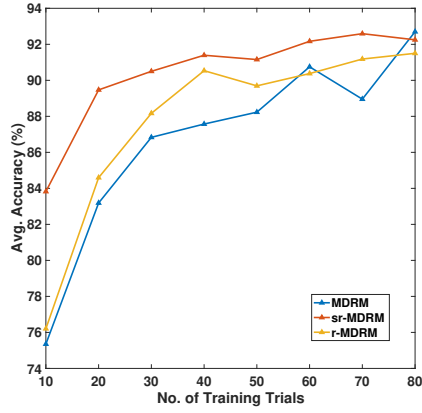
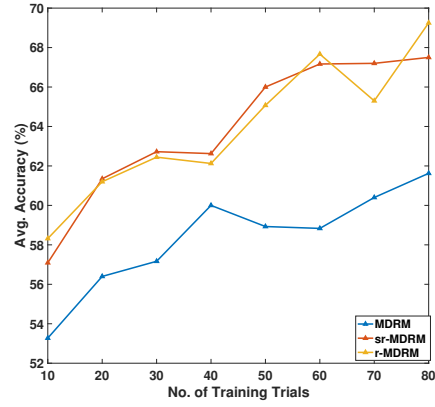


Figure 5.2: Box-plot of all the subjects at 20 training trials in dataset IVa.

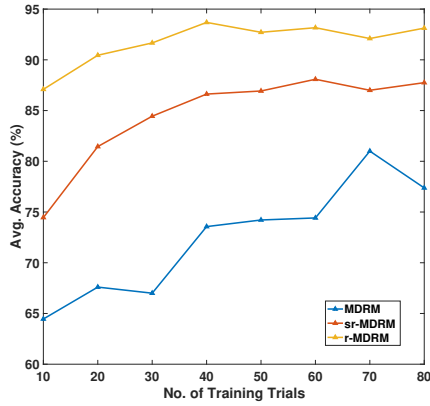
Mean classification accuracies in Figure 5.1 can sometimes be misleading because of outliers, therefore we have plotted a box-plot of classification accuracies obtained through different BCI pipelines using only 20 training trials in Figure 5.2. In this study, our focus was on using 20 trials (i.e. 10 trials per class) to achieve good performance because it will take only approximately 2.5 minutes training time (as the Graz protocol based single training trial takes eight seconds) compared to Blankertz et al., who suggested that 80 trials (i.e. approx. 11 minutes) are necessary to obtain a reasonable BCI performance [12]. In the box-plot shown in Figure 5.2, the middle red line indicates median classification accuracy and the box indicates the 25th to 75th percentile of classification accuracies of different BCI pipelines. Figure 5.2 shows that our proposed pipelines' median performance is consistently higher than the state of the art method (MDRM) for all the subjects in dataset IVa. In addition, a paired t-test gives very strong evidence ($p_{value} \ll 0.001$) that the average classification accuracy (over all subjects and all values of training trials) obtained by both of our proposed BCI pipelines was statistically higher



(a) Subject K3b



(b) Subject K6b



(c) Subject L1b

Figure 5.3: Average classification accuracies according to the number of training data available for the target subject.

than the standard MDRM pipeline. However, the difference is not statistically significant between the r-MDRM and the sr-MDRM pipelines (with $p = 0.0120$) at 1% significance level. We have chosen 1% significance value (critical value equals to 0.01) as it reduces the chance of a false positive, but it also makes it more difficult to reject the null hypothesis. Therefore, with a critical value of 0.01, the results are more trustworthy.

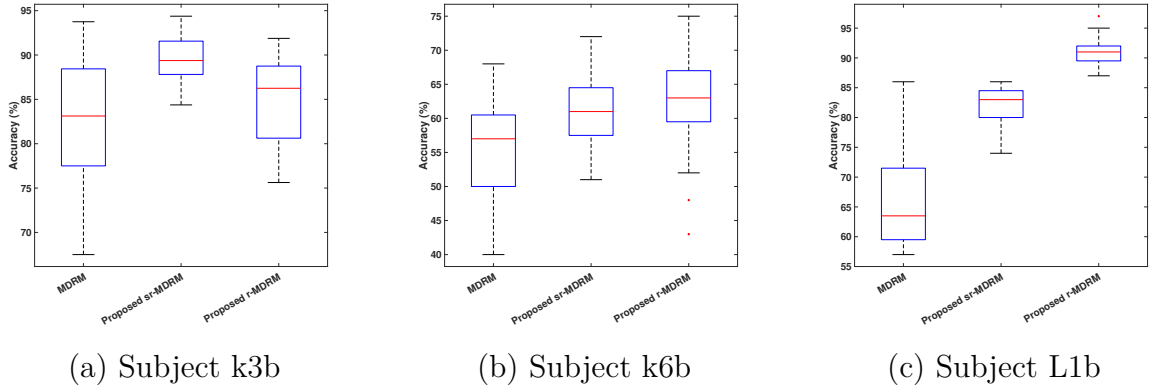


Figure 5.4: Box-plot of all the subjects at 20 training trials in dataset IIIa.

5.3.2 Dataset IIIa, BCI competition III

Dataset IIIa is recorded with 60 electrodes thus the covariance matrices obtained are of large size (60×60). Under a small training set, an MDRM pipeline encounters the curse of the dimensionality problem, as the size of covariance matrices are larger than the number of trials [17]. This is evident in Figure 5.3 which shows that standard MDRM pipeline records lower performance at training size 10, 20 and 30. Subject “K6b”’s average accuracy is very poor with all the pipelines but with our proposed pipelines the average accuracy increased a little bit in comparison with standard MDRM pipeline. Once we reached 50 training trials, we were able to achieve reasonable average accuracy. In addition to this, Figure 5.4 shows that the median accuracy of subject K6b at 20 training set is above 60 percent, where as standard MDRM pipeline median accuracy lies close to 55 percent. Similarly for subject *L1b* median accuracy for the MDRM pipeline is below 65 whereas proposed pipelines were able to achieve a median accuracy of 83 and 90.

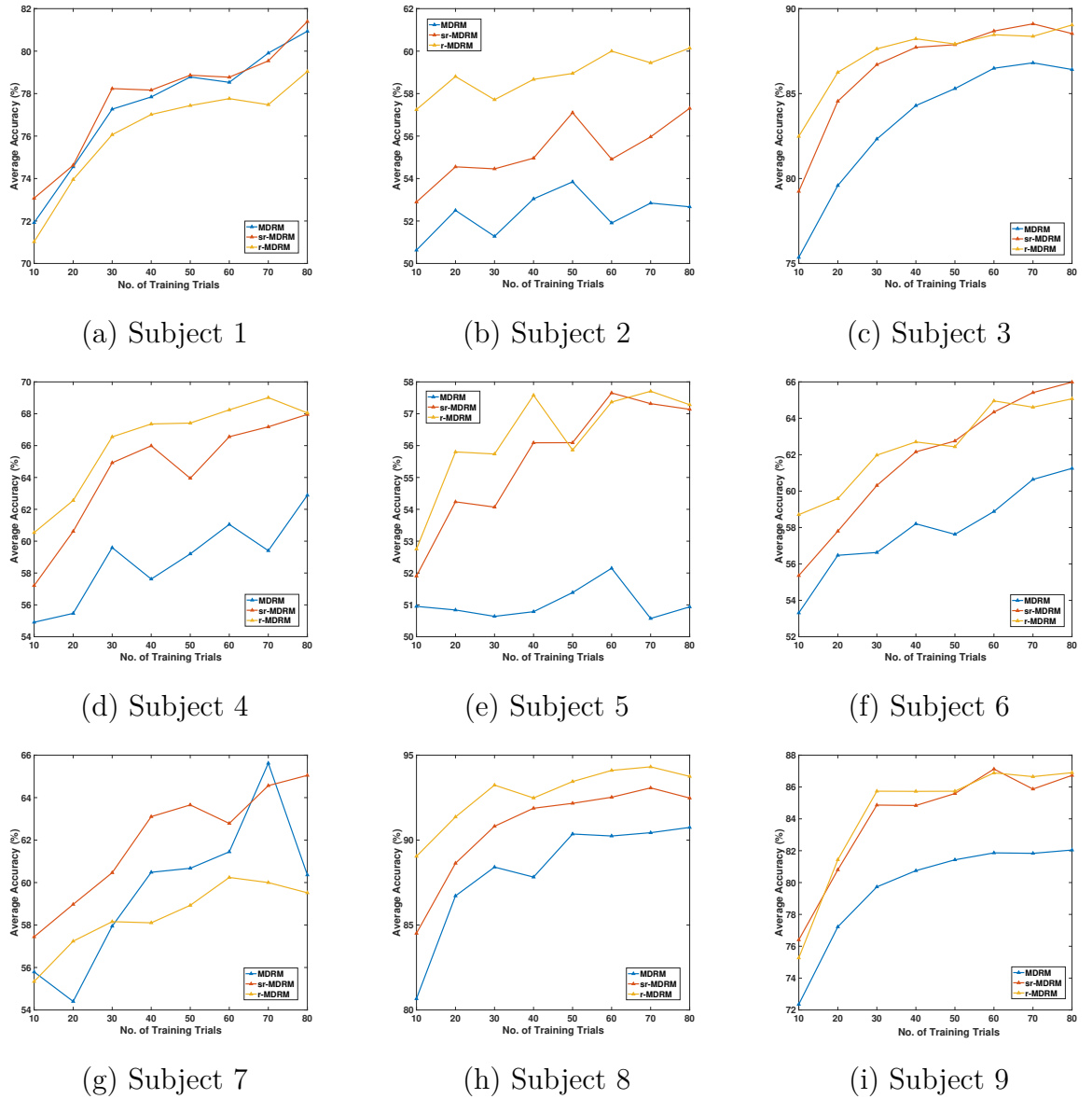


Figure 5.5: Average classification accuracies according to the number of training data available for the target subject.

The performance difference between the MDRM pipeline and the proposed BCI pipelines is statistically ($p \ll 0.001$) higher over all subjects and all training set values. Figure 5.6(a) shows that the average performance of the proposed pipelines is higher for all subjects at different training sizes. Interestingly, the average performance difference between r-MDRM and sr-MDRM is statistically significant ($p < 0.001$) over all subjects at different training trials. This performance difference is hard to explain as each subject's

external (subject's muscle movements, recording condition and machine-related causes) and internal (like fatigue, concentration and stress) state causes the BCI pipelines to respond differently.

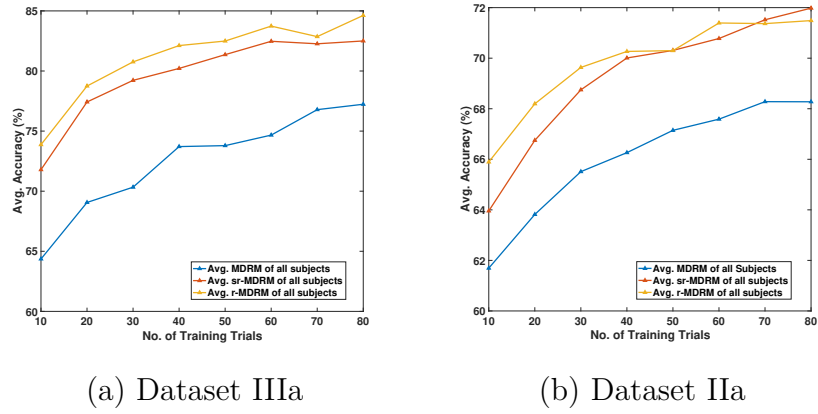


Figure 5.6: Average accuracies of all subjects in Dataset IIIa and Dataset IIa, respectively.

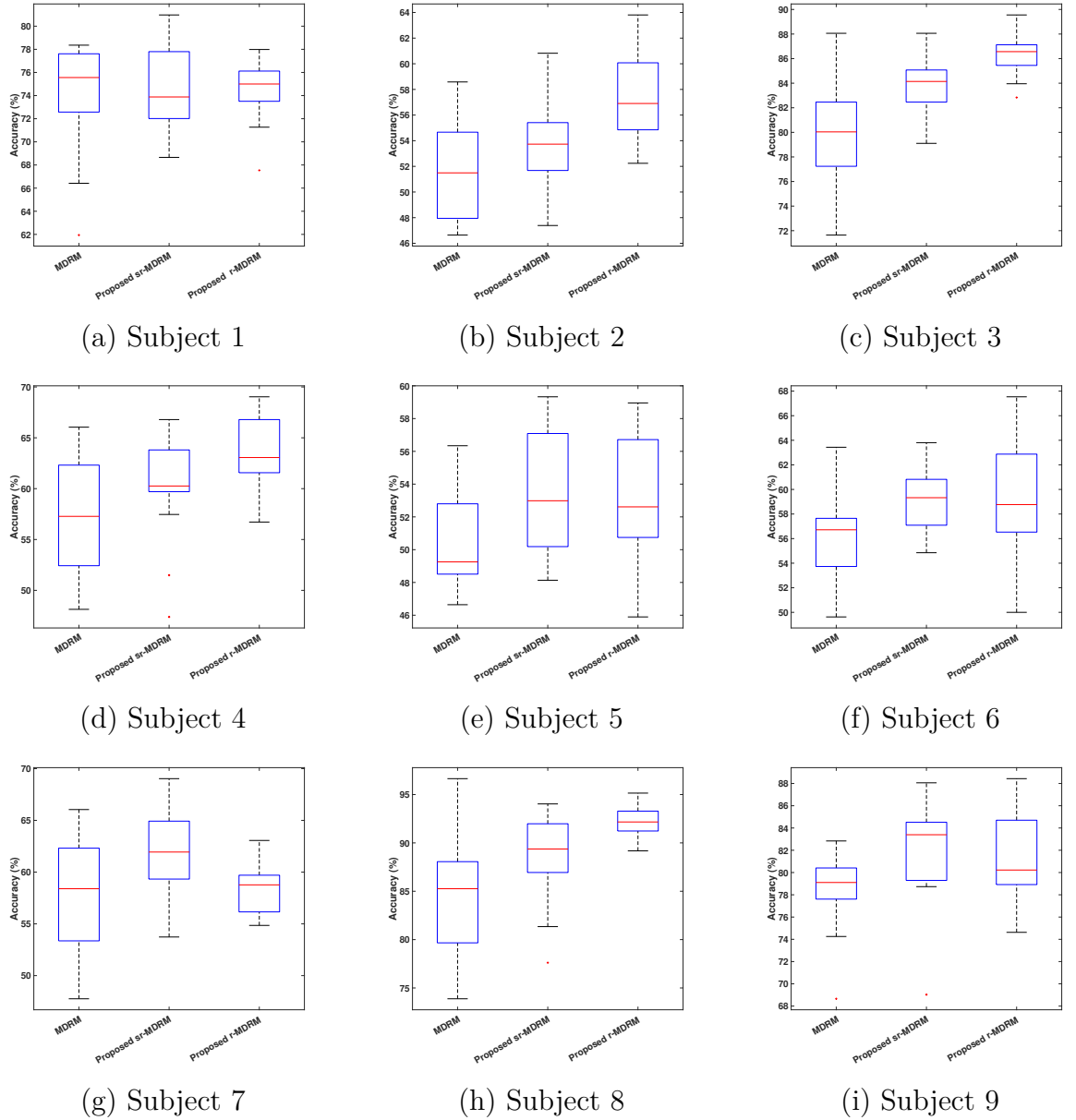


Figure 5.7: Box-plot of all the subjects at 20 training trials in Dataset IIa.

5.3.3 Dataset IIa, BCI competition IV

Dataset IIIa was recorded with 22 electrodes, thus covariance matrices obtained are comparatively small size (22×22). Therefore, the standard MDRM pipeline was able to perform better for subjects who were able to generate a distinguishable MI pattern for left/right MI task. Figure 5.5 shows average performance of all

the subjects under different training sizes. Figure 5.7 shows box plot of obtained accuracies at 20 training trials. From both plots, we can conclude that our proposed pipelines perform better than standard MDRM pipelines for all the subjects. This was also statistically verified using paired t-test. The paired t-test gave strong evidence that both proposed pipelines average accuracy over all subjects and training sizes is statistically higher than the MDRM pipeline. The difference in performance of r-MDRM and sr-MDRM pipeline is not significant at the 1% ($p_{value} = 0.0972$) level. Figure 5.6(b) shows average performance of all subjects at different training sizes. Some subjects (such as 2, 5, 6 and 7 respectively) were not able to achieve good performance even at 80 training trials and can be termed as BCI illiterate.

In addition, we compared the average training time of all three pipelines at 20 trials in Figure 5.8. For dataset IVa, the MDRM pipeline takes an average time of 12 seconds to train a Riemannian geometry based classifier, whereas both proposed pipelines take less than 0.05 seconds. Similarly with Dataset IIIa, MDRM takes an average time of around 0.4 seconds, whereas sr-MDRM and r-MDRM take less than 0.07 seconds. This is because computation of a Riemannian distance involves a matrix inversion and a matrix diagonalization operations whose computational time dramatically increases with the number of electrodes. The r-MDRM pipeline takes more time than sr-MDRM because it uses other subjects' trials to obtain generic covariance matrices per class for the target subject.

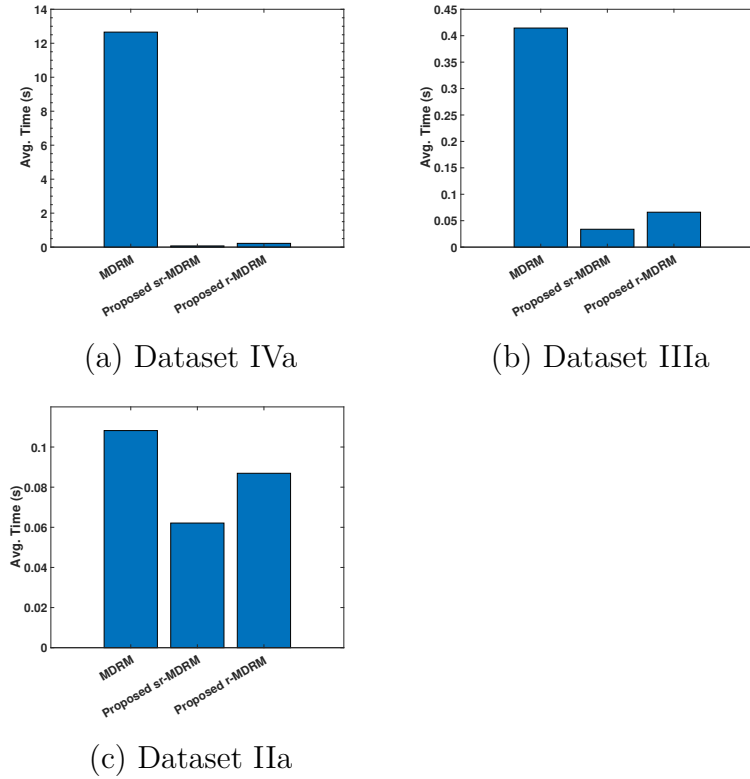


Figure 5.8: Average time in training algorithm on Single Subject in different datasets

5.4 Conclusion

The motivation of our thesis was to reduce calibration time as much as possible without losing performance. In this chapter we validated our approach in an offline training-test setting by using a bootstrap sampling technique to systematically reduce training size. The main finding of this chapter suggested that applying the proposed pipelines leads to reduction in the calibration time and enhances the average classification accuracy of the MI-BCI-based systems. This gain in performance is more evident when EEG signals are recorded with a large number of electrodes. From a theoretical point of view, this makes sense, as under high spatial and temporal resolution there is a high probability of noticing cognitive changes. But due to the curse of dimensionality, the standard MDRM pipeline failed to obtain good performance under small training sets. Our proposed pipelines were able to mitigate the issue of the numerical stability

of Riemannian geometry manipulations by reducing the size of covariance matrices with the use of spatial filters. This can be seen in results from Figure 5.1(g) and Figure 5.6) shows average accuracy of all subjects with proposed pipelines in comparison to standard pipeline. It can be seen at 20 trials reasonable performance is achieved i.e. shows approximately 77 percent drop in calibration time in comparison to 80 trials but at the cost of 9.6% drop in average accuracy for dataset IVa, 6.09% drop in average accuracy for dataset IIIa and 6.94% drop in average accuracy for dataset IIa for sr-MDRM pipeline. Similarly, r-MDRM average accuracy drop 6.02% for dataset IVa, 7.05% for dataset IIIa and 3.5% drop in dataset IIa. At the same time, both proposed pipelines perform better than Standard pipeline trained with 80 trials.

Chapter 6

Conclusion and Future Work

The work presented in this thesis aims to make BCI more reliable as a daily use system. Thus, our focus has been on developing novel approaches to reduce the calibration time of BCI with minimum accuracy loss. To achieve this objective two main challenges needed to be addressed based on the available training data from the current user and previous sessions or users. First, reducing intra-session non stationarity. Second, inter-subjects/sessions non-stationarity.

6.1 Conclusion

Through this thesis, we did a comprehensive review that equipped us with knowledge to propose two novel approaches to address the challenging issues and improve the usability of BCI. In Chapter 2, we provided a comprehensive overview of motor imagery-based BCIs in all relevant aspects. More precisely, the review is divided into two parts. In the first part, we discuss the various components of the MI-BCI system including data acquisition techniques, signal preprocessing techniques, feature extraction techniques, classification techniques, and finally performance evaluation metrics for BCIs or user performance. The second part of the study addresses current bottlenecks in the MI-BCI system and focuses on reducing the calibration time specific to the two main challenges.

In our literature review, we found that there are many research papers on feature extraction methods based on spatial filtering and source separation. These methods aim to decompose the sensor measurement into noise and EEG signals [265]. This clean EEG signal part is used for feature extraction and

classification. Mathematically, any linear spatial filtering is $y(t) = Bx(t)$ where $x(t)$ is an N -dimensional EEG measurement vector and the spatial filter is matrix B of size $P \times N$ with $0 < P \leq N$, realizing linear combinations of samples $x(t)$. Usually P is chosen smaller than the number of electrodes N , wherein the discarded $N - P$ components explain EEG energy not related to the task, i.e. the noise suppressed by the filter. While EEG scalp measurements are approximated by a linear mixture of brain dipolar sources i.e. $x(t) = As(t)$ where $s(t)$ is a vector holding the unknown source processes and A is the mixing matrix, assumed here invertible, with its left-inverse B named the demixing matrix.

Riemannian geometry gives an effective way to manipulate data points in the sensor space equivalent to source space of the same dimension. Mathematically, let S_i and S_j be the covariance matrix of the unknown source process for any two trials. Let two corresponding sensor covariance matrices be $C_i = AS_iA^T$ and $C_j = AS_jA^T$. Because of the congruence invariance property of the Riemannian distance if $P = N$ then the Riemannian distance between S_i and S_j is equivalent C_i and C_j . From a classification point of view, linear transformation does not impact the Riemannian distance between covariance matrices thus conserve a relatively good accuracy. High-dimensional covariance matrices ($N > 32$) are ill conditioned with respect to inversion, jeopardizing the numerical stability of all Riemannian geometry manipulations, which are based on spectral functions of the eigenvalues such as the logarithm, inverse, etc. [6, 19]. If we take $P < N$, meaning that we estimate fewer components than available sensors, we can still find a projection in a source sub-space enhancing the separation of the classes, that is, we can still improve the classification achieved by the MDRM as applied in the sensor space [6].

Equipped with this knowledge in Chapters 3 and 4, we adapted a CSP spatial filter in a preprocessing step before the Riemannian geometry based classification. CSP filter (F) is obtained by solving generalized eigenvector eigenvalue decomposition. Matrix (F) containing the first $P/2$ vectors explains the maximum of the variance of class A and the minimum of class B , while the last $P/2$ vectors explains the maximum of the variance of class B and the minimum of class A . Each associated eigenvector in matrix (F) projects the data covariance in a mono-dimensional space where the ratio of the variance in the two classes is maximized. That means CSP transformation not just

suppresses noise but also maximizes the ratio of the variance between two classes. Thus, the computation of the Riemannian distance between two filtered covariance matrices ($P < N$) takes less time, and also provides more noise suppression, making MDRM more robust, which increases the accuracy of the BCI system. In Chapter 3, we have addressed the intra-subject challenge by incorporating a CSP spatial filter [250] that is regularized by using apriori information about electrodes which are most likely to be useful in MI activity. MDRM classification is then applied on filtered covariance matrices. In the same vein, in Chapter 4, we have adapted a CSP filter version that uses common information across different subjects or sessions trials to obtain an optimal spatial filter for the target subject under a small sample setting to reduce the dimension of covariance matrices for classification. In both chapters, hyper-parameters of CSP filters were optimized using grid-search. This increases the computational cost but reduces the mental load as a limited number of trials were utilized. Lastly, in the Chapter 5 the proposed approaches were evaluated under reduced calibration settings. In order to do this, we utilized classical offline training-test framework where a bootstrap technique was systematically used to reduce the number of trials and check our proposed pipelines. Results from Chapter 5 suggested that with N small enough (< 32) the difference in accuracy obtained by the proposed pipelines and MDRM is very small, where as for large N , the proposed approaches proves superior because more and more irrelevant components are ignored by the CSP ($P \ll N$). Through our work, we exploited the advantages of both Euclidean and Riemannian approaches. The proposed pipelines were able to reduce calibration time 77 percent but at the cost of drop of accuracy in comparison to proposed pipelines when trained with 80 trials. At the same time, average classification accuracy proposed pipelines trained with 20 trials is better then standard MDRM pipeline calibrated with 80 trials. In summary, using the proposed algorithms we addressed the mentioned challenging issues, and consequently we achieved our objective to make BCI systems more robust with less calibration time.

6.2 Limitations and directions of future Work

The work presented in this thesis can potentially be extended to address the limitations we faced as well as some other challenges in BCI, or even other areas.

Some of these future extensions are listed below

- In this thesis, the results are obtained with an offline analysis. Although the results are promising and the analysis holds, it is recommended to perform an online evaluation of the proposed pipelines to verify the reliability of the results in real-time scenarios. For example, in the online evaluation, we need to explore methods (e.g. Riemannian potato [266], etc.) to reject bad trials before training the model and predicting the label of the new trial.
- The effectiveness of the proposed pipeline in Chapter 5 can be further improved by exploring new methods to measure the similarity between the previous subjects/sessions data and the few trials from the new subject.
- The proposed algorithms rely on different pre-processing methods. Thus, the classification results might be further improved, especially for subjects with poor BCI performance, by combining these algorithms and frameworks in a complementary way. Accordingly, different complementary combinations need to be explored in order to design a more accurate and effective BCI with minimum calibration time.
- This way higher accuracy may be obtained, however we need to estimate a spatial filter specific to the available training data, losing generalization power. Also, the Riemannian distance and mean are robust with respect to the noise if we do not filter the data whenever N is small.

In addition to the above-mentioned future works, the following long-term extensions might be of interest.

- In general, zero calibration is the ideal situation for a real-time BCI system that may be utilised for tasks that are part of daily life. A BCI user's learning process may be adversely effected by too frequent recalibrations, as well as leave the user perplexed about the received feedback. Therefore, recalibration needs to be done extremely carefully to avoid confusing the BCI user. To further enhance the BCI user learning process, research should be done at the human level to develop more advanced and effective user training methodologies. Future research in the BCI community would be highly interesting and important if it involved modelling human learning with signal processing and machine learning and utilising those models to determine when recalibration is required.

- It would be worthwhile to extend our proposed pipelines to include additional neurophysiological signals beyond motor imagery. Applying and generalising the proposed algorithms to additional domains that are impacted by noise and non-stationary data is very desirable in addition to minimising the calibration time and enhancing the accuracy and robustness of BCI.
- The discussions and analysis of the proposed pipelines in this study were limited to two classes of motor imagery tasks. It would be ideal to develop pipelines (algorithms) that can accurately classify more types of motor imagery tasks (classes). More accurate identification of motor imagery tasks translates to more commands issued by the BCI user to a communication device. In the future, it will be necessary to evaluate EEG signals from multi-class motor imagery tasks either by extending our proposed pipelines to multi-class paradigms or by developing new algorithms in order to increase the performance of multi-class BCIs with minimum calibration time.

References

- [1] J. Wolpaw, N. Birbaumer, D. Mcfarland, G. Pfurtscheller, and T. Vaughan, “Wolpaw, j.r.: Brain-computer interfaces for communication and control. ’clin. neurophysiol. 113, 767-791,” *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 113, pp. 767–91, 07 2002.
- [2] A. Singh, S. Lal, and H. W. Guesgen, “Architectural review of co-adaptive brain computer interface,” in *2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, Dec 2017, pp. 200–207.
- [3] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Z. Zhang, and K. Kreutz-Delgado, “Evolving signal processing for brain–computer interfaces,” *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1567–1584, 2012.
- [4] S. Saeedi, “Reliability and adaptive assistance in brain-computer interfaces,” p. 115, 2016. [Online]. Available: <http://infoscience.epfl.ch/record/218629>
- [5] K. A. McCreadie, D. H. Coyle, and G. Prasad, “Is sensorimotor bci performance influenced differently by mono, stereo, or 3-d auditory feedback?” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 3, pp. 431–440, May 2014.
- [6] M. Congedo, A. Barachant, and R. Bhatia, “Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review,” *Brain-Computer Interfaces*, vol. 4, no. 3, pp. 155–174, 2017.

References.

- [7] C. S. L. Tsui, J. Q. Gan, and H. Hu, “A self-paced motor imagery based brain-computer interface for robotic wheelchair control,” *Clinical EEG and neuroscience*, vol. 42, no. 4, pp. 225–229, 2011.
- [8] G. Pfurtscheller, C. Neuper, G. Muller, B. Obermaier, G. Krausz, A. Schlogl, R. Scherer, B. Graimann, C. Keinrath, D. Skliris, M. Wortz, G. Supp, and C. Schrank, “Graz-bci: state of the art and clinical applications,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 1–4, 2003.
- [9] A. Zimmermann-Schlatter, C. Schuster, M. A. Puhan, E. Siekierka, and J. Steurer, “Efficacy of motor imagery in post-stroke rehabilitation: a systematic review,” *Journal of neuroengineering and rehabilitation*, vol. 5, no. 1, pp. 1–10, 2008.
- [10] T. Li, J. Zhang, T. Xue, and B. Wang, “Development of a novel motor imagery control technique and application in a gaming environment,” *Computational Intelligence and Neuroscience*, vol. 2017, 2017.
- [11] U. Talukdar, S. M. Hazarika, and J. Q. Gan, “Adaptive feature extraction in EEG-based motor imagery BCI: tracking mental fatigue,” *Journal of Neural Engineering*, vol. 17, no. 1, p. 016020, jan 2020. [Online]. Available: <https://doi.org/10.1088/1741-2552/ab53f1>
- [12] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, and K.-R. Müller, “The berlin brain-computer interface: Accurate performance from first-session in bci-naive subjects,” *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 10, pp. 2452–2462, 2008.
- [13] A. Azab, “Novel transfer learning approaches for improving brain computer interfaces,” Ph.D. dissertation, University of Sheffield, 2019.
- [14] F. Lotte and C. Guan, “Regularizing common spatial patterns to improve bci designs: Unified theory and new algorithms,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, Feb 2011.
- [15] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, “Transfer learning in brain-computer interfaces,” *IEEE Computational Intelligence Magazine*, vol. 11, no. 1, pp. 20–31, Feb 2016.

References.

- [16] F. Lotte, “Signal processing approaches to minimize or suppress calibration time in oscillatory activity based brain computer interfaces,” *Proceedings of the IEEE*, vol. 103, no. 6, pp. 871–890, June 2015.
- [17] C. H. Nguyen and P. Artemiadis, “Eeg feature descriptors and discriminant analysis under riemannian manifold perspective,” *Neurocomputing*, vol. 275, pp. 1871 – 1883, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231217316648>
- [18] P. L. C. Rodrigues, “Exploring invariances of multivariate time series via riemannian geometry: validation on eeg data,” Ph.D. dissertation, Université de Grenoble, Ecole doctorale EEATS, spécialité Signal, Image . . . , 2019.
- [19] R. Bhatia, *Positive Definite Matrices*. USA: Princeton University Press, 2015.
- [20] H. Bashashati, R. K. Ward, A. Bashashati, and A. Mohamed, “Neural network conditional random fields for self-paced brain computer interfaces,” in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec 2016, pp. 939–943.
- [21] M. Rashid, N. Sulaiman, A. P. P. Abdul Majeed, R. M. Musa, A. F. Ab. Nasir, B. S. Bari, and S. Khatun, “Current status, challenges, and possible solutions of eeg-based brain-computer interface: A comprehensive review,” *Frontiers in Neurorobotics*, vol. 14, p. 25, 2020. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnbot.2020.00025>
- [22] R. A. Ramadan and A. V. Vasilakos, “Brain computer interface: control signals review,” *Neurocomputing*, vol. 223, pp. 26 – 44, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231216312152>
- [23] M. L. Martini, E. K. Oermann, N. L. Opie, F. Panov, T. Oxley, and K. Yaeger, “Sensor Modalities for Brain-Computer Interface Technology: A Comprehensive Literature Review,” *Neurosurgery*, 07 2019. [Online]. Available: <https://doi.org/10.1093/neuros/nyz286>
- [24] P. Bucci and S. Galderisi, *Physiologic Basis of the EEG Signal*. John Wiley and Sons, Ltd, ch. 2, pp. 7–12.

References.

- [25] B. Farnsworth, “Eeg (electroencephalography): The complete pocket guide,” 2019. [Online]. Available: <https://imotions.com/blog/eeg/>
- [26] G. Pfurtscheller and C. Neuper, “Motor imagery and direct brain-computer communication,” *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.
- [27] T. Otaiby, F. Abd El-Samie, S. Alshebeili, and I. Ahmad, “A review of channel selection algorithms for eeg signal processing,” *EURASIP Journal on Advances in Signal Processing*, vol. 2015, 08 2015.
- [28] X. Wan, K. Zhang, S. Ramkumar, J. Deny, G. Emayavaramban, M. Siva Ramkumar, and A. F. Hussein, “A review on electroencephalogram based brain computer interface for elderly disabled,” *IEEE Access*, vol. 7, pp. 36 380–36 387, 2019.
- [29] C. Jeunet, E. Jahanpour, and F. Lotte, “Why standard brain-computer interface (BCI) training protocols should be changed: An experimental study,” *Journal of Neural Engineering*, vol. 13, no. 3, 2016.
- [30] F. Cincotti, L. Kauhanen, F. Aloise, T. Palomäki, N. Caporusso, P. Jylänki, D. Mattia, F. Babiloni, G. Vanacker, M. Nuttin, M. Marciani, and J. d. R. Millan, “Vibrotactile feedback for brain-computer interface operation,” *Computational intelligence and neuroscience*, vol. 2007, p. 48937, 02 2007.
- [31] F. Lotte, J. Faller, C. Guger, Y. Renard, G. Pfurtscheller, A. Lécuyer, and R. Leeb, *Combining BCI with Virtual Reality: Towards New Applications and Improved BCI*, 07 2012, pp. 197–220.
- [32] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, “Brain-computer communication: Motivation, aim, and impact of exploring a virtual apartment,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 4, pp. 473–482, Dec 2007.
- [33] M. K. Islam, A. Rastegarnia, and Z. Yang, “Methods for artifact detection and removal from scalp eeg: A review,” *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 46, no. 4, pp. 287 – 305, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S098770531630199X>

References.

- [34] L. F. S. Uribe, C. A. S. Filho, V. A. de Oliveira, T. B. da Silva Costa, P. G. Rodrigues, D. C. Soriano, L. Boccato, G. Castellano, and R. Attux, “A correntropy-based classifier for motor imagery brain-computer interfaces,” *Biomedical Physics & Engineering Express*, vol. 5, no. 6, p. 065026, nov 2019. [Online]. Available: <https://doi.org/10.1088/2057-1976/5/6/065026>
- [35] B. Xu, L. Zhang, A. Song, C. Wu, W. Li, D. Zhang, G. Xu, H. Li, and H. Zeng, “Wavelet transform time-frequency image and convolutional network-based motor imagery eeg classification,” *IEEE Access*, vol. 7, pp. 6084–6093, 2019.
- [36] D. Suma, J. Meng, B. J. Edelman, and B. He, “Spatial-temporal aspects of continuous EEG-based neurobotic control,” *Journal of Neural Engineering*, vol. 17, no. 6, p. 066006, dec 2020. [Online]. Available: <https://doi.org/10.1088/1741-2552/abc0b4>
- [37] O. W. Samuel, Y. Geng, X. Li, and G. Li, “Towards efficient decoding of multiple classes of motor imagery limb movements based on eeg spectral and time domain descriptors,” *Journal of medical systems*, vol. 41, no. 12, p. 194, October 2017. [Online]. Available: <https://doi.org/10.1007/s10916-017-0843-z>
- [38] M. Hamed, S. Salleh, A. M. Noor, and I. Mohammad-Rezazadeh, “Neural network-based three-class motor imagery classification using time-domain features for bci applications,” in *2014 IEEE REGION 10 SYMPOSIUM*, 2014, pp. 204–207.
- [39] G. Rodríguez-Bermúdez and P. J. García-Laencina, “Automatic and adaptive classification of electroencephalographic signals for brain computer interfaces,” *J. Med. Syst.*, vol. 36, no. 1, p. 51–63, Feb. 2012. [Online]. Available: <https://doi.org/10.1007/s10916-012-9893-4>
- [40] U. Güçlü, Y. Güçlütürk, and C. K. Loo, “Evaluation of fractal dimension estimation methods for feature extraction in motor imagery based brain computer interface,” *Procedia Computer Science*, vol. 3, pp. 589–594, 2011, world Conference on Information Technology. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050910004734>

References.

- [41] C.-Y. Kee, P. S.G., and L. Chu Kiong, “Binary and multi-class motor imagery using renyi entropy for feature extraction,” *Neural Computing and Applications*, 01 2016.
- [42] S. Chen, Z. Luo, and H. Gan, “An entropy fusion method for feature extraction of eeg,” *Neural Computing and Applications*, vol. 29, pp. 857–863, 2016.
- [43] A. Adam, Z. Ibrahim, N. Mokhtar, M. I. Shapiai, P. Cumming, and M. Mubin, “Evaluation of different time domain peak models using extreme learning machine-based peak detection for eeg signal,” *SpringerPlus*, vol. 5, 07 2016.
- [44] C. M. Yilmaz, C. Kose, and B. Hatipoglu, “A quasi-probabilistic distribution model for eeg signal classification by using 2-d signal representation,” *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 187 – 196, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0169260718300956>
- [45] P. Batres-Mendoza, C. R. Montoro-Sanjose, E. I. Guerra-Hernandez, D. L. Almanza-Ojeda, H. Rostro-Gonzalez, R. J. Romero-Troncoso, and M. A. Ibarra-Manzano, “Quaternion-based signal analysis for motor imagery classification from electroencephalographic signals,” *Sensors*, vol. 16, no. 3, 2016. [Online]. Available: <https://www.mdpi.com/1424-8220/16/3/336>
- [46] C. A. Stefano Filho, R. Attux, and G. Castellano, “Can graph metrics be used for eeg-bcis based on hand motor imagery?” *Biomedical Signal Processing and Control*, vol. 40, pp. 359–365, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809417302355>
- [47] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggenberger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, “Deep learning with convolutional neural networks for eeg decoding and visualization,” *Human Brain Mapping*, vol. 38, no. 11, p. 5391–5420, Aug 2017. [Online]. Available: <http://dx.doi.org/10.1002/hbm.23730>
- [48] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “EEGNet: A compact convolutional neural network

References.

- for EEG-based brain-computer interfaces,” *Journal of Neural Engineering*, no. 5, 2018.
- [49] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, “Lstm-based eeg classification in motor imagery tasks,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 11, pp. 2086–2095, Nov 2018.
- [50] S. Aggarwal and N. Chugh, “Signal processing techniques for motor imagery brain computer interface: A review,” *Array*, vol. 1-2, p. 100003, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2590005619300037>
- [51] G. Rashkov, A. Bobe, D. Fastovets, and M. Komarova, “Natural image reconstruction from brain waves: a novel visual BCI system with native feedback,” *bioRxiv*, pp. 1–15, 2019.
- [52] C. D. V. G., J. H. S. A., J. M. Antelis, and L. E. Falcón, “Spiking neural networks applied to the classification of motor tasks in eeg signals,” *Neural Networks*, vol. 122, pp. 130 – 143, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0893608019303193>
- [53] S.-B. Lee, H.-J. Kim, H. Kim, J.-H. Jeong, S.-W. Lee, and D.-J. Kim, “Comparative analysis of features extracted from eeg spatial, spectral and temporal domains for binary and multiclass motor imagery classification,” *Information Sciences*, vol. 502, pp. 190 – 200, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0020025519305365>
- [54] Y. Chu, X. Zhao, Y. Zou, W. Xu, J. Han, and Y. Zhao, “A decoding scheme for incomplete motor imagery eeg with deep belief network,” *Frontiers in Neuroscience*, vol. 12, p. 680, 2018. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnins.2018.00680>
- [55] R. Zhang, P. Xu, R. Chen, F. Li, L. Guo, P. Li, T. Zhang, and D. Yao, “Predicting inter-session performance of smr-based brain-computer interface using the spectral entropy of resting-state eeg,” *Brain Topography*, vol. 28, pp. 680–690, 2015.

References.

- [56] Z. Gao, Z. Wang, C. Ma, W. Dang, and K. Zhang, “A wavelet time-frequency representation based complex network method for characterizing brain activities underlying motor imagery signals,” *IEEE Access*, vol. 6, pp. 65 796–65 802, 2018.
- [57] M. Ortiz, E. Iáñez, J. L. Contreras-Vidal, and J. M. Azorín, “Analysis of the eeg rhythms based on the empirical mode decomposition during motor imagery when using a lower-limb exoskeleton. a case study,” *Frontiers in Neurorobotics*, vol. 14, p. 48, 2020. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnbot.2020.00048>
- [58] Guo Xiaojing, Wu Xiaopei, and Zhang Dexiang, “Motor imagery eeg detection by empirical mode decomposition,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 2619–2622.
- [59] C. J. Ortiz-Echeverri, S. Salazar-Colores, J. Rodríguez-Reséndiz, and R. A. Gómez-Loenzo, “A new approach for motor imagery classification based on sorted blind source separation, continuous wavelet transform, and convolutional neural network,” *Sensors*, vol. 19, no. 20, 2019. [Online]. Available: <https://www.mdpi.com/1424-8220/19/20/4541>
- [60] M. Li, C. Zhang, S. Jia, and Y. Sun, “Classification of motor imagery tasks in source domain,” in *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, Aug 2018, pp. 83–88.
- [61] I. Rejer and P. Górski, “Eeg classification for mi-bci with independent component analysis,” in *Proceedings of the 10th International Conference on Computer Recognition Systems CORES 2017*, M. Kurzynski, M. Wozniak, and R. Burduk, Eds. Cham: Springer International Publishing, 2018, pp. 393–402.
- [62] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan, “Filter bank common spatial pattern (fbcsp) in brain-computer interface,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 2390–2397.
- [63] K. p thomas, C. Guan, C.-T. Lau, A. Vinod, and K. Ang, “A new discriminative common spatial pattern method for motor imagery

References.

- brain-computer interface,” *Biomedical Engineering, IEEE Transactions on*, pp. 2730 – 2733, 12 2009.
- [64] C. J. Ortiz-Echeverri, S. Salazar-Colores, J. Rodríguez-Reséndiz, and R. A. Gómez-Loenzo, “A new approach for motor imagery classification based on sorted blind source separation, continuous wavelet transform, and convolutional neural network,” *Sensors (Switzerland)*, vol. 19, no. 20, pp. 1–14, 2019.
- [65] Y. Li, X. Zhang, B. Zhang, M. Lei, W. Cui, and Y. Guo, “A channel-projection mixed-scale convolutional neural network for motor imagery eeg decoding,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 6, pp. 1170–1180, June 2019.
- [66] J. Yang, S. Yao, and J. Wang, “Deep fusion feature learning network for mi-eeg classification,” *IEEE Access*, vol. 6, pp. 79 050–79 059, 2018.
- [67] 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*, vol. 55, no. 6, pp. 1733–1743, 2008.
- [68] H. Suk and S. Lee, “A probabilistic approach to spatio-spectral filters optimization in brain-computer interface,” in *2011 IEEE International Conference on Systems, Man, and Cybernetics*, 2011, pp. 19–24.
- [69] P. Zhang, X. Wang, W. Zhang, and J. Chen, “Learning spatial–spectral–temporal eeg features with recurrent 3d convolutional neural networks for cross-task mental workload assessment,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 1, pp. 31–42, 2019.
- [70] J. S. Bang, M. H. Lee, S. Fazli, C. Guan, and S. W. Lee, “Spatio-spectral feature representation for motor imagery classification using convolutional neural networks,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2021.
- [71] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, “Classification of covariance matrices using a Riemannian-based kernel for BCI applications,” *Neurocomputing*, vol. 112, pp. 172–178, 2013.

References.

- [72] I. Horev, F. Yger, and M. Sugiyama, “Geometry-aware principal component analysis for symmetric positive definite matrices,” *Machine Learning*, vol. 106, no. 4, pp. 493–522, 2017.
- [73] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.
- [74] B. Venkatesh and J. Anuradha, “A review of feature selection and its methods,” *Cybernetics and Information Technologies*, vol. 19, no. 1, pp. 3 – 26, 01 Mar. 2019. [Online]. Available: <https://content.sciendo.com/view/journals/cait/19/1/article-p3.xml>
- [75] R. Battiti, “Using Mutual Information for Selecting Features in Supervised Neural Net Learning,” *IEEE Transactions on Neural Networks*, vol. 5, no. 4, pp. 537–550, 1994.
- [76] I. Homri and S. Yacoub, “A hybrid cascade method for EEG classification,” *Pattern Analysis and Applications*, vol. 22, no. 4, pp. 1505–1516, 2019. [Online]. Available: <https://doi.org/10.1007/s10044-018-0737-9>
- [77] A. C. Ramos, R. G. Hernandex, and M. Vellasco, “Feature Selection Methods Applied to Motor Imagery Task Classification,” in *2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 2016.
- [78] S. Kashef, H. Nezamabadi-pour, and B. Nikpour, “Multilabel feature selection: A comprehensive review and guiding experiments,” *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 2, p. e1240, 2018. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1240>
- [79] A. Atyabi, F. Shic, and A. Naples, “Mixture of autoregressive modeling orders and its implication on single trial EEG classification,” *Expert Systems with Applications*, vol. 65, pp. 164–180, 2016. [Online]. Available: <file:///C:/Users/CarlaCarolina/Desktop/Artigosparaacrescentarnaqualifica{\c{c}}{\~{a}}o/Theimpactofbirthweightoncardiovasculardiseaseriskinthe.pdf>
- [80] M. Z. Baig, N. Aslam, H. P. Shum, and L. Zhang, “Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG,” *Expert Systems with Applications*, vol. 90, pp. 184–195, 2017.

References.

- [81] S. Das and P. N. Suganthan, "Differential evolution: A survey of the state-of-the-art," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 4–31, 2011.
- [82] D. Karaboga, "An idea based on honey bee swarm for numerical optimization ," *Technical report-tr06, Erciyes university, engineering faculty, computer engineering department.*, vol. 200, pp. 1–10, 2005.
- [83] P. Rakshit, S. Bhattacharyya, A. Konar, A. Khasnobish, D. Tibarewala, and R. Janarthanan, "Artificial Bee Colony Based Feature Selection for Motor Imagery EEG Data," in *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)*, 2012, pp. 127–138. [Online]. Available: https://link.springer.com/chapter/10.1007/978-81-322-1041-2{_}11
- [84] L. van der Maaten, E. Postma, and H. Herik, "Dimensionality reduction: A comparative review," *Journal of Machine Learning Research - JMLR*, vol. 10, 01 2007.
- [85] A. Gupta, R. K. Agrawal, and B. Kaur, "Performance enhancement of mental task classification using eeg signal: a study of multivariate feature selection methods," *Soft Computing*, vol. 19, no. 10, pp. 2799–2812, Oct 2015. [Online]. Available: <https://doi.org/10.1007/s00500-014-1443-1>
- [86] V. Jusas and S. G. Samuvel, "Classification of motor imagery using a combination of user-specific band and subject-specific band for brain-computer interface," *Applied Sciences*, vol. 9, no. 23, 2019. [Online]. Available: <https://www.mdpi.com/2076-3417/9/23/4990>
- [87] S. Ayesha, M. K. Hanif, and R. Talib, "Overview and comparative study of dimensionality reduction techniques for high dimensional data," *Information Fusion*, vol. 59, pp. 44 – 58, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S156625351930377X>
- [88] B. Scholkopf, A. Smola, and K.-R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299–1319, 1996.

References.

- [89] D. Pei, M. Burns, R. Chandramouli, and R. Vinjamuri, “Decoding asynchronous reaching in electroencephalography using stacked autoencoders,” *IEEE Access*, vol. 6, pp. 52 889–52 898, 2018.
- [90] J. Tenenbaum, V. Silva, and J. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, pp. 2319–2323, 01 2000.
- [91] P. Iturralde, M. Patrone, F. Lecumberry, and A. Fernández, “Motor intention recognition in eeg: In pursuit of a relevant feature set,” in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, 09 2012, pp. 551–558.
- [92] A. Gramfort and M. Clerc, “Low dimensional representations of meg/eeg data using laplacian eigenmaps,” in *2007 Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging*, Oct 2007, pp. 169–172.
- [93] S. Lafon and A. B. Lee, “Diffusion maps and coarse-graining: a unified framework for dimensionality reduction, graph partitioning, and data set parameterization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1393–1403, Sep. 2006.
- [94] F. Lee, R. Scherer, R. Leeb, A. Schlögl, H. Bischof, and G. Pfurtscheller, “Feature mapping using pca, locally linear embedding and isometric feature mapping for eeg-based brain computer interface,” *Proceedings of the 28th AAPR Workshop, Austrian Computer Society*, p. 189–196, 2004.
- [95] M. Li, X. Luo, J. Yang, and Y. Sun, “Applying a locally linear embedding algorithm for feature extraction and visualization of mi-eeg,” *Journal of Sensors*, vol. 2016, pp. 1–9, 01 2016.
- [96] X. Xie, Z. L. Yu, H. Lu, Z. Gu, and Y. Li, “Motor imagery classification based on bilinear sub-manifold learning of symmetric positive-definite matrices,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 6, pp. 504–516, June 2017.

References.

- [97] A. Davoudi, S. S. Ghidary, and K. Sadatnejad, “Dimensionality reduction based on distance preservation to local mean for symmetric positive definite matrices and its application in brain–computer interfaces,” *Journal of Neural Engineering*, vol. 14, no. 3, p. 036019, 2017. [Online]. Available: <http://stacks.iop.org/1741-2552/14/i=3/a=036019>
- [98] T. Tanaka, T. Uehara, and Y. Tanaka, “Dimensionality reduction of sample covariance matrices by graph fourier transform for motor imagery brain-machine interface,” in *2016 IEEE Statistical Signal Processing Workshop (SSP)*, June 2016, pp. 1–5.
- [99] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, “A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update.” *Journal of neural engineering*, vol. 15, no. 3, p. 031005, 2018. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/29488902>
- [100] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons Inc, 2001.
- [101] E. Thomas, M. Dyson, and M. Clerc, “An analysis of performance evaluation for motor-imagery based bci,” *Journal of neural engineering*, vol. 10, p. 031001, 05 2013.
- [102] A. Schlögl, J. Kronegg, J. Huggins, and S. Mason, “Evaluation criteria for bci research,” *Toward Brain-Computer Interfacing*, 01 2007.
- [103] J. R. Wolpaw, H. Ramoser, D. J. McFarland, and G. Pfurtscheller, “Eeg-based communication: improved accuracy by response verification,” *IEEE Transactions on Rehabilitation Engineering*, vol. 6, no. 3, pp. 326–333, Sep. 1998.
- [104] T. Nykopp, “Statistical Modelling Issues for The Adaptive Brain Interface,” Ph.D. dissertation, Helsinki University of Technology, 2001.
- [105] F. Lotte and C. Jeunet, “Defining and quantifying users’ mental imagery-based BCI skills: a first step,” *Journal of Neural Engineering*, vol. 15, no. 4, p. 046030, jun 2018. [Online]. Available: <https://doi.org/10.1088/1741-2552/15/4/046030>

References.

- [106] J. Solé-Casals, C. Caiafa, Q. Zhao, and A. Cichocki, “Brain-computer interface with corrupted eeg data: a tensor completion approach,” *Cognitive Computation*, 07 2018.
- [107] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, “An automatic subject specific intrinsic mode function selection for enhancing two-class eeg-based motor imagery-brain computer interface,” *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6938–6947, Aug 2019.
- [108] M. M. Togha, M. R. Salehi, and E. Abiri, “Improving the performance of the motor imagery-based brain-computer interfaces using local activities estimation,” *Biomedical Signal Processing and Control*, vol. 50, pp. 52 – 61, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1746809419300084>
- [109] R. Sampanna and S. Mitaim, “Noise benefits in the array of brain-computer interface classification systems,” *Informatics in Medicine Unlocked*, vol. 12, pp. 88 – 97, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2352914818300923>
- [110] S. Kumar and A. Sharma, “A new parameter tuning approach for enhanced motor imagery eeg signal classification,” *Medical and biological engineering and computing*, vol. 56, 04 2018.
- [111] C. S. Kim, J. Sun, D. Liu, Q. Wang, and S. G. Paek, “Removal of ocular artifacts using ica and adaptive filter for motor imagery-based bci,” *IEEE/CAA Journal of Automatica Sinica*, pp. 1–8, 2017.
- [112] L. Sun, Z. Feng, B. Chen, and N. Lu, “A contralateral channel guided model for eeg based motor imagery classification,” *Biomedical Signal Processing and Control*, vol. 41, pp. 1 – 9, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1746809417302537>
- [113] H. Sagha, S. Perdakis, J. d. R. Millán, and R. Chavarriaga, “Quantifying electrode reliability during brain–computer interface operation,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 3, pp. 858–864, March 2015.

References.

- [114] J. K. Feng, J. Jin, I. Daly, J. Zhou, Y. Niu, X. Wang, and A. Cichocki, “An Optimized Channel Selection Method Based on Multifrequency CSP-Rank for Motor Imagery-Based BCI System,” *Computational Intelligence and Neuroscience*, vol. 2019, 2019.
- [115] A. Ramakrishnan and J. Satyanarayana, “Reconstruction of eeg from limited channel acquisition using estimated signal correlation,” *Biomedical Signal Processing and Control*, vol. 27, pp. 164 – 173, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1746809416300167>
- [116] Y. Yang, I. Bloch, S. Chevallier, and J. Wiart, “Subject-specific channel selection using time information for motor imagery brain–computer interfaces,” *Cognitive Computation*, vol. 8, no. 3, pp. 505–518, Jun 2016. [Online]. Available: <https://doi.org/10.1007/s12559-015-9379-z>
- [117] J. Ruan, X. Wu, B. Zhou, X. Guo, and Z. Lv, “An automatic channel selection approach for ica-based motor imagery brain computer interface,” *Journal of Medical Systems*, vol. 42, no. 12, p. 253, Nov 2018. [Online]. Available: <https://doi.org/10.1007/s10916-018-1106-3>
- [118] S.-M. Park, J.-Y. Kim, and K.-B. Sim, “Eeg electrode selection method based on bpsso with channel impact factor for acquisition of significant brain signal,” *Optik*, vol. 155, pp. 89 – 96, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0030402617312986>
- [119] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, “Correlation-based channel selection and regularized feature optimization for MI-based BCI,” *Neural Networks*, vol. 118, pp. 262–270, 2019. [Online]. Available: <https://doi.org/10.1016/j.neunet.2019.07.008>
- [120] X.-Y. Yu, J.-H. Yu, and K.-B. Sim, “Fruit fly optimization based eeg channel selection method for bci,” *Journal of Institute of Control, Robotics and Systems*, vol. 22, pp. 199–203, 03 2016.
- [121] N. Masood, H. Farooq, and I. Mustafa, “Selection of EEG channels based on Spatial filter weights,” in *Proceedings of 2017 International Conference on Communication, Computing and Digital Systems, C-CODE 2017*. IEEE, 2017, pp. 341–345.

References.

- [122] Y. Yang, S. Chevallier, J. Wiart, and I. Bloch, "Subject-specific time-frequency selection for multi-class motor imagery-based bcis using few laplacian eeg channels," *Biomedical Signal Processing and Control*, vol. 38, pp. 302 – 311, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1746809417301295>
- [123] R. Rajan and S. Thekkan Devassy, "Improving classification performance by combining feature vectors with a boosting approach for brain computer interface (bci)," in *Intelligent Human Computer Interaction*, P. Horain, C. Achard, and M. Malle, Eds. Cham: Springer International Publishing, 2017, pp. 73–85.
- [124] P. Shahsavari Baboukani, S. Mohammadi, and G. Azemi, "Classifying single-trial eeg during motor imagery using a multivariate mutual information based phase synchrony measure," in *2017 24th National and 2nd International Iranian Conference on Biomedical Engineering (ICBME)*, Nov 2017, pp. 1–4.
- [125] J. Wang, Z. Feng, N. Lu, L. Sun, and J. Luo, "An information fusion scheme based common spatial pattern method for classification of motor imagery tasks," *Biomedical Signal Processing and Control*, vol. 46, pp. 10 – 17, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1746809418301605>
- [126] A. Liu, K. Chen, Q. Liu, Q. Ai, Y. Xie, and A. Chen, "Feature selection for motor imagery eeg classification based on firefly algorithm and learning automata," *Sensors*, vol. 17, no. 11, 2017. [Online]. Available: <https://www.mdpi.com/1424-8220/17/11/2576>
- [127] S. Kumar, A. Sharma, and T. Tsunoda, "An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information," *BMC Bioinformatics*, vol. 18, pp. 125–137, 2017.
- [128] K. Samanta, S. Chatterjee, and R. Bose, "Cross subject motor imagery tasks eeg signal classification employing multiplex weighted visibility graph and deep feature extraction," *IEEE Sensors Letters*, pp. 1–1, 2019.
- [129] X. Xie, Z. L. Yu, Z. Gu, J. Zhang, L. Cen, and Y. Li, "Bilinear regularized locality preserving learning on riemannian graph for motor imagery bci,"

References.

- IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 3, pp. 698–708, March 2018.
- [130] Q. She, H. Gan, Y. Ma, Z. Luo, T. Potter, and Y. Zhang, “Scale-dependent signal identification in low-dimensional subspace: Motor imagery task classification,” *Neural Plasticity*, vol. 2016, pp. 1–15, 11 2016.
- [131] O. Özdenizci and D. Erdoğan, “Information theoretic feature transformation learning for brain interfaces,” *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 1, pp. 69–78, Jan 2020.
- [132] I. Razzak, I. A. Hameed, and G. Xu, “Robust sparse representation and multiclass support matrix machines for the classification of motor imagery eeg signals,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–8, 2019.
- [133] M. T. Harandi, M. Salzmann, and R. Hartley, “From manifold to manifold: Geometry-aware dimensionality reduction for spd matrices,” in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 17–32.
- [134] M. Li, J. Han, and L. Duan, “A novel mi-eeg imaging with the location information of electrodes,” *IEEE Access*, pp. 1–1, 2019.
- [135] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015.
- [136] H.-T. Wang, T. Li, H. Huang, Y.-B. He, and X.-C. Liu, “A motor imagery analysis algorithm based on spatio-temporal-frequency joint selection and relevance vector machine,” *Kongzhi Lilun Yu Yingyong/Control Theory and Applications*, vol. 34, pp. 1403–1408, 10 2017.
- [137] M. T. Sadiq, X. Yu, Z. Yuan, Z. Fan, A. U. Rehman, G. Li, and G. Xiao, “Motor imagery eeg signals classification based on mode amplitude and frequency components using empirical wavelet transform,” *IEEE Access*, vol. 7, pp. 127 678–127 692, 2019.
- [138] S. Selim, M. M. Tantawi, H. A. Shedeed, and A. Badr, “A csp am-ba-svm approach for motor imagery bci system,” *IEEE Access*, vol. 6, pp. 49 192–49 208, 2018.

References.

- [139] M. Athif and H. Ren, “Wavecsp: a robust motor imagery classifier for consumer eeg devices,” *Australasian Physical & Engineering Sciences in Medicine*, vol. 42, pp. 1–10, 01 2019.
- [140] X. Li, C. Guan, H. Zhang, and K. K. Ang, “A unified fisher’s ratio learning method for spatial filter optimization,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 11, pp. 2727–2737, Nov 2017.
- [141] L. Li, G. Xu, F. Zhang, J. Xie, and M. Li, “Relevant feature integration and extraction for single-trial motor imagery classification,” *Frontiers in Neuroscience*, vol. 11, p. 371, 2017. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnins.2017.00371>
- [142] Y. Liu, H. Zhang, M. Chen, and L. Zhang, “A boosting-based spatial-spectral model for stroke patients’ eeg analysis in rehabilitation training,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 1, pp. 169–179, Jan 2016.
- [143] A. Cachón and R. A. Vázquez, “Tuning the parameters of an integrate and fire neuron via a genetic algorithm for solving pattern recognition problems,” *Neurocomputing*, vol. 148, pp. 187 – 197, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231214009321>
- [144] R. Salazar-Varas and R. A. Vazquez, “Evaluating spiking neural models in the classification of motor imagery eeg signals using short calibration sessions,” *Applied Soft Computing*, vol. 67, pp. 232 – 244, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1568494618301133>
- [145] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, “A multi-branch 3d convolutional neural network for eeg-based motor imagery classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 10, pp. 2164–2177, Oct 2019.
- [146] Y. Park and W. Chung, “Frequency-optimized local region common spatial pattern approach for motor imagery classification,” *IEEE Transactions*

References.

- on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp. 1378–1388, July 2019.
- [147] Y. Ma, X. Ding, Q. She, Z. Luo, T. Potter, and Y. Zhang, “Classification of motor imagery eeg signals with support vector machines and particle swarm optimization,” *Computational and Mathematical Methods in Medicine*, vol. 2016, pp. 1–8, 01 2016.
- [148] A. Costa, J. Møller, H. Iversen, and S. Puthusserypady, “An adaptive csp filter to investigate user independence in a 3-class mi-bci paradigm,” *Computers in Biology and Medicine*, vol. 103, pp. 24 – 33, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S001048251830283X>
- [149] F. Lotte and C. Guan, “Spatially regularized common spatial patterns for eeg classification,” in *2010 20th International Conference on Pattern Recognition*, Aug 2010, pp. 3712–3715.
- [150] A. Singh, S. Lal, and H. W. Guesgen, “Reduce calibration time in motor imagery using spatially regularized symmetric positives-definite matrices based classification,” *Sensors*, vol. 19, no. 2, 2019. [Online]. Available: <https://www.mdpi.com/1424-8220/19/2/379>
- [151] S. Saha, K. I. U. Ahmed, R. Mostafa, L. Hadjileontiadis, and A. Khandoker, “Evidence of variabilities in eeg dynamics during motor imagery-based multiclass brain–computer interface,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 2, pp. 371–382, Feb 2018.
- [152] H. He and D. Wu, “Transfer learning for brain-computer interfaces: An euclidean space data alignment approach,” *CoRR*, vol. abs/1808.05464, 2018. [Online]. Available: <http://arxiv.org/abs/1808.05464>
- [153] I. Hossain, A. Khosravi, I. T. Hettiarachchi, and S. Nahavandhi, “Informative instance transfer learning with subject specific frequency responses for motor imagery brain computer interface,” in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct 2017, pp. 252–257.

References.

- [154] M. Dai, S. Wang, D. Zheng, R. Na, and S. Zhang, “Domain transfer multiple kernel boosting for classification of eeg motor imagery signals,” *IEEE Access*, vol. 7, pp. 49 951–49 960, 2019.
- [155] 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*, vol. 26, no. 2, pp. 498–505, Feb 2018.
- [156] A. M. Azab, L. Mihaylova, K. K. Ang, and M. Arvaneh, “Weighted transfer learning for improving motor imagery-based brain–computer interface,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp. 1352–1359, July 2019.
- [157] A. Singh, S. Lal, and H. W. Guesgen, “Motor imagery classification based on subject to subject transfer in riemannian manifold,” in *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, Feb 2019, pp. 1–6.
- [158] A. Singh, S. Lal, and H. W. Guesgen, “Small sample motor imagery classification using regularized riemannian features,” *IEEE Access*, vol. 7, pp. 46 858–46 869, 2019.
- [159] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, and A. Cichocki, “Sparse group representation model for motor imagery eeg classification,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 631–641, March 2019.
- [160] P. L. C. Rodrigues, C. Jutten, and M. Congedo, “Riemannian procrustes analysis: Transfer learning for brain–computer interfaces,” *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 8, pp. 2390–2401, Aug 2019.
- [161] X. Zhu, P. Li, C. Li, D. Yao, R. Zhang, and P. Xu, “Separated channel convolutional neural network to realize the training free motor imagery bci systems,” *Biomedical Signal Processing and Control*, vol. 49, pp. 396 – 403, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1746809418303264>

References.

- [162] M. Joadder, S. Siuly, E. Kabir, H. Wang, and Y. Zhang, “A new design of mental state classification for subject independent bci systems,” *IRBM*, vol. 40, no. 5, pp. 297 – 305, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1959031819300697>
- [163] X. Zhao, J. Zhao, W. Cai, and S. Wu, “Transferring common spatial filters with semi-supervised learning for zero-training motor imagery brain-computer interface,” *IEEE Access*, vol. 7, pp. 58 120–58 130, 2019.
- [164] O. Kwon, M. Lee, C. Guan, and S. Lee, “Subject-independent brain-computer interfaces based on deep convolutional neural networks,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2019.
- [165] L. Yao, X. Sheng, D. Zhang, N. Jiang, N. Mrachacz-Kersting, X. Zhu, and D. Farina, “A stimulus-independent hybrid bci based on motor imagery and somatosensory attentional orientation,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. PP, pp. 1–1, 03 2017.
- [166] X. Shu, S. Chen, L. Yao, X. Sheng, D. Zhang, N. Jiang, J. Jia, and X. Zhu, “Fast recognition of bci-inefficient users using physiological features from eeg signals: A screening study of stroke patients,” *Frontiers in Neuroscience*, vol. 12, p. 93, 2018. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnins.2018.00093>
- [167] L. Acqualagna, L. Botrel, C. Vidaurre, A. Kübler, and B. Blankertz, “Large-scale assessment of a fully automatic co-adaptive motor imagery-based brain computer interface,” *PLoS ONE*, vol. 11, no. 2, pp. 1–19, 2016.
- [168] X. Shu, S. Chen, G. Chai, X. Sheng, J. Jia, and X. Zhu, “Neural Modulation by Repetitive Transcranial Magnetic Stimulation (rTMS) for BCI Enhancement in Stroke Patients,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018, pp. 2272–2275.
- [169] C. Sannelli, C. Vidaurre, K. R. Müller, and B. Blankertz, “A large scale screening study with a SMR-based BCI: Categorization of BCI users and

References.

- differences in their SMR activity,” *PLoS ONE*, vol. 14, no. 1, pp. 1–37, 2019.
- [170] R. Zhang, X. Li, Y. Wang, B. Liu, L. Shi, M. Chen, L. Zhang, and Y. Hu, “Using Brain Network Features to Increase the Classification Accuracy of MI-BCI Inefficiency Subject,” *IEEE Access*, vol. 7, pp. 74 490–74 499, 2019.
- [171] M. Ahn, H. Cho, S. Ahn, and S. C. Jun, “User’s self-prediction of performance in motor imagery brain–computer interface,” *Frontiers in Human Neuroscience*, vol. 12, p. 59, 2018. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnhum.2018.00059>
- [172] S. Darvishi, A. Gharabaghi, M. C. Ridding, D. Abbott, and M. Baumert, “Reaction time predicts brain–computer interface aptitude,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, pp. 1–11, 2018.
- [173] J. Müller, C. Vidaurre, M. Schreuder, F. Meinecke, P. von Büna, and K.-R. Müller, “A mathematical model for the two-learners problem,” *Journal of neural engineering*, vol. 14, 02 2017.
- [174] C. Vidaurre, C. Sannelli, K. R. Müller, and B. Blankertz, “Machine-learning-based coadaptive calibration for Brain-computer interfaces,” *Neural Computation*, vol. 23, no. 3, pp. 791–816, 2011.
- [175] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, and S.-W. Lee, “EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy,” *GigaScience*.
- [176] R. Zhang, X. Li, Y. Wang, B. Liu, l. Shi, M. Chen, L. Zhang, and H. Yuxia, “Using brain network features to increase the classification accuracy of mi-bci inefficiency subject,” *IEEE Access*, vol. PP, pp. 1–1, 05 2019.
- [177] C. Sannelli, C. Vidaurre, K.-R. Müller, and B. Blankertz, “Ensembles of adaptive spatial filters increase BCI performance: an online evaluation,” *Journal of Neural Engineering*, vol. 13, no. 4, p. 046003, may 2016. [Online]. Available: <https://doi.org/10.1088/1741-2560/13/4/046003>
- [178] C. Vidaurre, A. R. Murguialday, S. Haufe, M. Gómez, K.-R. Müller, and V. Nikulin, “Enhancing sensorimotor bci performance with assistive

References.

- afferent activity: An online evaluation,” *NeuroImage*, vol. 199, pp. 375 – 386, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1053811919304707>
- [179] Y. Yu, Z. Zhou, E. Yin, J. Jiang, J. Tang, Y. Liu, and D. Hu, “Toward brain-actuated car applications: Self-paced control with a motor imagery-based brain-computer interface,” *Computers in Biology and Medicine*, vol. 77, pp. 148–155, 2016. [Online]. Available: <http://dx.doi.org/10.1016/j.compbiomed.2016.08.010>
- [180] P. Cheng, P. Autthasan, B. Pijarana, E. Chuangsuwanich, and T. Wilaiprasitporn, “Towards asynchronous motor imagery-based brain-computer interfaces: a joint training scheme using deep learning,” in *TENCON 2018 - 2018 IEEE Region 10 Conference*, Oct 2018, pp. 1994–1998.
- [181] G. Sanchez-Ante, J. Antelis, B. Gudiño-Mendoza, L. Falcon, and H. Sossa, “Dendrite morphological neural networks for motor task recognition from electroencephalographic signals,” *Biomedical Signal Processing and Control*, vol. 44, 04 2018.
- [182] Y. Jiang, N. T. Hau, and W. Chung, “Semiasynchronous bci using wearable two-channel eeg,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 3, pp. 681–686, Sep. 2018.
- [183] Y. Sun, Z. Feng, J. Zhang, Q. Zhou, and J. Luo, “Asynchronous motor imagery detection based on a target guided sub-band filter using wavelet packets,” in *2017 29th Chinese Control And Decision Conference (CCDC)*, May 2017, pp. 4850–4855.
- [184] S. He, Y. Zhou, T. Yu, R. Zhang, Q. Huang, L. Chuai, Madah-Ul-Mustafa, Z. Gu, Z. L. Yu, H. Tan, and Y. Li, “Eeg- and eog-based asynchronous hybrid bci: A system integrating a speller, a web browser, an e-mail client, and a file explorer,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, pp. 1–1, 2019.
- [185] Y. Yu, Y. Liu, J. Jiang, E. Yin, Z. Zhou, and D. Hu, “An asynchronous control paradigm based on sequential motor imagery and its application

References.

- in wheelchair navigation,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 12, pp. 2367–2375, Dec 2018.
- [186] H. An, J. Kim, and S. Lee, “Design of an asynchronous brain-computer interface for control of a virtual avatar,” in *2016 4th International Winter Conference on Brain-Computer Interface (BCI)*, Feb 2016, pp. 1–2.
- [187] Y. Jiang, J. He, D. Li, J. Jin, and Y. Shen, “Signal classification algorithm in motor imagery based on asynchronous brain-computer interface,” in *2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, May 2019, pp. 1–5.
- [188] R. Yousefi, A. Rezazadeh, and T. Chau, “Development of a robust asynchronous brain-switch using erp-based error correction,” *Journal of Neural Engineering*, vol. 16, 09 2019.
- [189] Y. Yu, Z. Zhou, Y. Liu, J. Jiang, E. Yin, N. Zhang, Z. Wang, Y. Liu, X. Wu, and D. Hu, “Self-paced operation of a wheelchair based on a hybrid brain-computer interface combining motor imagery and p300 potential,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2516–2526, Dec 2017.
- [190] L. Wang and X. Wu, “Classification of four-class motor imagery eeg data using spatial filtering,” in *2008 2nd International Conference on Bioinformatics and Biomedical Engineering*, 2008, pp. 2153–2156.
- [191] M. Grosse-Wentrup and M. Buss, “Multiclass common spatial patterns and information theoretic feature extraction,” *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1991–2000, Aug 2008.
- [192] S. M. Christensen, N. S. Holm, and S. Puthusserypady, “An improved five class mi based bci scheme for drone control using filter bank csp,” in *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, Feb 2019, pp. 1–6.
- [193] I. Razzak, M. Blumenstein, and G. Xu, “Multiclass support matrix machines by maximizing the inter-class margin for single trial eeg classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 6, pp. 1117–1127, June 2019.

References.

- [194] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, “Multiclass brain computer interface classification by riemannian geometry,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, April 2012.
- [195] M. Z. A. Faiz and A. A. Al-Hamadani, “Online brain computer interface based five classes eeg to control humanoid robotic hand,” in *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, July 2019, pp. 406–410.
- [196] S. Aliakbaryhosseinabadi, E. N. Kamavuako, N. Jiang, D. Farina, and N. Mrachacz-Kersting, “Classification of movement preparation between attended and distracted self-paced motor tasks,” *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 11, pp. 3060–3071, Nov 2019.
- [197] N. Dagaev, K. Volkova, and A. Ossadtchi, “Latent variable method for automatic adaptation to background states in motor imagery BCI,” *Journal of Neural Engineering*, vol. 15, no. 1, p. 016004, nov 2017. [Online]. Available: <https://doi.org/10.1088/1741-2552/15/1/016004>
- [198] V. Mondini, A. Mangia, and A. Cappello, “Eeg-based bci system using adaptive features extraction and classification procedures,” *Computational Intelligence and Neuroscience*, vol. 2016, pp. 1–14, 08 2016.
- [199] A. Schwarz, J. Brandstetter, J. Pereira, and G. R. Muller-Putz, “Direct comparison of supervised and semi-supervised retraining approaches for co-adaptive bcis,” *Medical & Biological Engineering and Computing*, vol. 57, p. 2347–2357, November 2019.
- [200] S. Saeedi, R. Chavarriaga, and J. d. R. Millán, “Long-term stable control of motor-imagery bci by a locked-in user through adaptive assistance,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 4, pp. 380–391, April 2017.
- [201] S. Perdakis, R. Leeb, and J. d R Millán, “Context-aware adaptive spelling in motor imagery BCI,” *Journal of Neural Engineering*, vol. 13, no. 3, p. 036018, may 2016. [Online]. Available: <https://doi.org/10.1088/1741-2560/13/3/036018>

References.

- [202] J. Faller, C. Vidaurre, T. Solis-Escalante, C. Neuper, and R. Scherer, “Autocalibration and recurrent adaptation: Towards a plug and play online erd-bci,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 3, pp. 313–319, May 2012.
- [203] H. Raza, D. Rathee, S.-M. Zhou, H. Cecotti, and G. Prasad, “Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related eeg-based brain-computer interface,” *Neurocomputing*, vol. 343, pp. 154 – 166, 2019, learning in the Presence of Class Imbalance and Concept Drift. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231219301560>
- [204] H. Rong, C. Li, R. Bao, and B. Chen, “Incremental adaptive eeg classification of motor imagery-based bci,” in *2018 International Joint Conference on Neural Networks (IJCNN)*, July 2018, pp. 1–7.
- [205] V. Sharghian, T. Y. Rezaii, A. Farzamnia, and M. A. Tinati, “Online dictionary learning for sparse representation-based classification of motor imagery eeg,” in *2019 27th Iranian Conference on Electrical Engineering (ICEE)*, April 2019, pp. 1793–1797.
- [206] Z. Zhang, R. Foong, K. S. Phua, C. Wang, and K. K. Ang, “Modeling eeg-based motor imagery with session to session online adaptation,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2018, pp. 1988–1991.
- [207] J. Andreu-Perez, F. Cao, H. Hagnas, and G. Yang, “A self-adaptive online brain–machine interface of a humanoid robot through a general type-2 fuzzy inference system,” *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 1, pp. 101–116, Feb 2018.
- [208] K. K. Ang and C. Guan, “Eeg-based strategies to detect motor imagery for control and rehabilitation,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 4, pp. 392–401, April 2017.
- [209] E. Abdalsalam, M. Z. Yusoff, A. Malik, N. Kamel, and D. Mahmoud, “Modulation of sensorimotor rhythms for brain-computer interface using motor imagery with online feedback,” *Signal, Image and Video Processing*, 10 2017.

References.

- [210] R. Ron-Angevin and A. Díaz-Estrella, “Brain–computer interface: Changes in performance using virtual reality techniques,” *Neuroscience Letters*, vol. 449, no. 2, pp. 123–127, 2009. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0304394008015176>
- [211] D. Achancaray, K. Pacheco, E. Carranza, and M. Hayashibe, “Immersive virtual reality feedback in a brain computer interface for upper limb rehabilitation,” in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct 2018, pp. 1006–1010.
- [212] B. Alchalabi and J. Faubert, “A comparison between bci simulation and neurofeedback for forward/backward navigation in virtual reality,” *Computational Intelligence and Neuroscience*, vol. 2019, pp. 1–12, 10 2019.
- [213] J. Asensio-Cubero, J. Gan, and R. Palaniappan, “Multiresolution analysis over graphs for a motor imagery based online bci game,” *Computers in Biology and Medicine*, vol. 68, 11 2015.
- [214] M. Jianjun and B. He, “Exploring training effect in 42 human subjects using a non-invasive sensorimotor rhythm based online bci,” *Frontiers in Human Neuroscience*, vol. 13, 04 2019.
- [215] S. Kim, M. Lee, and S. Lee, “Self-paced training on motor imagery-based bci for minimal calibration time,” in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct 2017, pp. 2297–2301.
- [216] C. Jeunet, B. N’Kaoua, and F. Lotte, “Chapter 1 - advances in user-training for mental-imagery-based bci control: Psychological and cognitive factors and their neural correlates,” in *Brain-Computer Interfaces: Lab Experiments to Real-World Applications*, ser. Progress in Brain Research, D. Coyle, Ed. Elsevier, 2016, vol. 228, pp. 3 – 35. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0079612316300061>
- [217] H. Zhang, Y. Sun, J. Li, F. Wang, and Z. Wang, “Covert verb reading contributes to signal classification of motor imagery in bci,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 1, pp. 45–50, Jan 2018.

References.

- [218] L. Wang, X. Liu, Z. Liang, Z. Yang, and X. Hu, “Analysis and classification of hybrid bci based on motor imagery and speech imagery,” *Measurement*, vol. 147, p. 106842, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0263224119306992>
- [219] Z. Wang, Y. Zhou, L. Chen, B. Gu, S. Liu, M. Xu, H. Qi, F. He, and D. Ming, “A BCI based visual-haptic neurofeedback training improves cortical activations and classification performance during motor imagery,” *Journal of Neural Engineering*, vol. 16, no. 6, p. 066012, oct 2019. [Online]. Available: <https://doi.org/10.1088/1741-2552/16/6/066012>
- [220] S. Liburkina, A. Vasilyev, L. Yakovlev, S. Gordleeva, and A. Kaplan, “Motor imagery based brain computer interface with vibrotactile interaction,” *Zhurnal Vysshei Nervnoi Deyatelnosti Imeni I.P. Pavlova*, vol. 67, pp. 414–429, 01 2017.
- [221] L. Pillette, C. Jeunet, B. Mansencal, R. N’Kambou, B. N’Kaoua, and F. Lotte, “A physical learning companion for mental-imagery bci user training,” *International Journal of Human-Computer Studies*, vol. 136, p. 102380, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1071581919301466>
- [222] F. Škola, S. Tinková, and F. Liarokapis, “Progressive training for motor imagery brain-computer interfaces using gamification and virtual reality embodiment,” *Frontiers in Human Neuroscience*, vol. 13, p. 329, 2019. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnhum.2019.00329>
- [223] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, “The non-invasive berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects,” *NeuroImage*, vol. 37, no. 2, pp. 539 – 550, 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1053811907000535>
- [224] K. P. Thomas, C. Guan, L. C. Tong, and A. P. Vinod, “Discriminative filterbank selection and eeg information fusion for brain computer interface,” in *2009 IEEE International Symposium on Circuits and Systems*, May 2009, pp. 1469–1472.

References.

- [225] L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain computer interfaces, a review,” *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012. [Online]. Available: <http://www.mdpi.com/1424-8220/12/2/1211>
- [226] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, “Filter bank common spatial pattern (fb-csp) in brain-computer interface,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, June 2008, pp. 2390–2397.
- [227] H. Lu, H. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, “Regularized common spatial pattern with aggregation for eeg classification in small-sample setting,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2936–2946, Dec 2010.
- [228] M. Dai, D. Zheng, S. Liu, and P. Zhang, “Transfer kernel common spatial patterns for motor imagery brain-computer interface classification,” *Computational and Mathematical Methods in Medicine*, pp. 49 192–49 208, 2018.
- [229] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct 2010.
- [230] M. Arvaneh, C. Guan, K. K. Ang, and H. C. Quek, “Spatially sparsed common spatial pattern to improve bci performance,” in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 2412–2415.
- [231] Y. Park and W. Chung, “Bci classification using locally generated csp features,” in *2018 6th International Conference on Brain-Computer Interface (BCI)*, Jan 2018, pp. 1–4.
- [232] H. Yang, S. Sakhavi, K. K. Ang, and C. Guan, “On the use of convolutional neural networks and augmented csp features for multi-class motor imagery of eeg signals classification,” in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2015, pp. 2620–2623.

References.

- [233] S. Park and S. Lee, “Small sample setting and frequency band selection problem solving using subband regularized common spatial pattern,” *IEEE Sensors Journal*, vol. 17, no. 10, pp. 2977–2983, May 2017.
- [234] 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*, pp. 1–11, 2018.
- [235] Z. Tang, C. Li, and S. Sun, “Single-trial eeg classification of motor imagery using deep convolutional neural networks,” *Optik*, vol. 130, pp. 11 – 18, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0030402616312980>
- [236] Y. R. Tabar and U. Halici, “A novel deep learning approach for classification of eeg motor imagery signals,” *Journal of Neural Engineering*, vol. 14, no. 1, p. 016003, 2017. [Online]. Available: <http://stacks.iop.org/1741-2552/14/i=1/a=016003>
- [237] M. T. Harandi, M. Salzmann, and R. Hartley, “From manifold to manifold: Geometry-aware dimensionality reduction for spd matrices,” in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 17–32.
- [238] S. Kumar, K. Mamun, and A. Sharma, “Csp-tsm: Optimizing the performance of riemannian tangent space mapping using common spatial pattern for mi-bci,” *Computers in Biology and Medicine*, vol. 91, pp. 231 – 242, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0010482517303517>
- [239] F. Yger, M. Berar, and F. Lotte, “Riemannian approaches in brain-computer interfaces: A review,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1753–1762, Oct 2017.
- [240] M. R. Islam, T. Tanaka, and M. K. I. Molla, “Multiband tangent space mapping and feature selection for classification of EEG during motor imagery,” *Journal of Neural Engineering*, vol. 15, no. 4, pp. 1–14, 2018.

References.

- [241] Z. Chebbi and M. Moakher, “Means of hermitian positive-definite matrices based on the log-determinant α -divergence function,” *Linear Algebra and its Applications*, vol. 436, no. 7, pp. 1872 – 1889, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S002437951100783X>
- [242] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, “Riemannian geometry applied to bci classification,” in *Latent Variable Analysis and Signal Separation*, V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, and E. Vincent, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 629–636.
- [243] J. Olias, R. Martín-Clemente, M. A. Sarmiento-Vega, and S. Cruces, “Eeg signal processing in mi-bci applications with improved covariance matrix estimators,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 895–904, 2019.
- [244] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, “Optimizing spatial filters for robust eeg single-trial analysis,” *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.
- [245] X. Li and H. Wang, “Smooth spatial filter for common spatial patterns,” in *Neural Information Processing*, M. Lee, A. Hirose, Z.-G. Hou, and R. M. Kil, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 315–322.
- [246] B. Blankertz, K. . Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. R. Millan, M. Schroder, and N. Birbaumer, “The bci competition iii: validating alternative approaches to actual bci problems,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, June 2006.
- [247] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. Miller, G. Mueller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, “Review of the bci competition iv,” *Frontiers in Neuroscience*, vol. 6, p. 55, 2012. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnins.2012.00055>

References.

- [248] 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*, vol. 58, no. 6, pp. 1865–1873, June 2011.
- [249] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, “Eeg-based discrimination between imagination of right and left hand movement,” *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 6, pp. 642 – 651, 1997. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0013469497000801>
- [250] F. Lotte and C. Guan, “Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms.” *Ieee Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.
- [251] J. Sim and C. C. Wright, “The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements,” *Physical Therapy*, vol. 85, no. 3, pp. 257–268, 03 2005. [Online]. Available: <https://doi.org/10.1093/ptj/85.3.257>
- [252] K. Belwafi, O. Romain, S. Gannouni, F. Ghaffari, R. Djemal, and B. Ouni, “An embedded implementation based on adaptive filter bank for brain–computer interface systems,” *Journal of Neuroscience Methods*, vol. 305, pp. 1 – 16, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S016502701830116X>
- [253] S. Selim, M. Tantawi, H. Shedeed, and A. Badr, “Reducing execution time for real-time motor imagery based bci systems,” in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016*, A. E. Hassanien, K. Shaalan, T. Gaber, A. T. Azar, and M. F. Tolba, Eds. Cham: Springer International Publishing, 2017, pp. 555–565.
- [254] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, “A multi-class eeg-based bci classification using multivariate empirical mode decomposition based filtering and riemannian geometry,” *Expert Systems with Applications*, vol. 95, pp. 201 – 211, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417417307492>
- [255] H. Raza, H. Cecotti, Y. Li, and G. Prasad, “Adaptive learning with covariate shift-detection for motor imagery-based brain–computer

References.

- interface,” *Soft Computing*, vol. 20, no. 8, pp. 3085–3096, Aug 2016. [Online]. Available: <https://doi.org/10.1007/s00500-015-1937-5>
- [256] S. Kumar, A. Sharma, and T. Tsunoda, “An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information,” *BMC Bioinformatics*, vol. 18, pp. 125–137, 2017.
- [257] Y. Ren and Y. Wu, “Convolutional deep belief networks for feature extraction of eeg signal,” in *2014 International Joint Conference on Neural Networks (IJCNN)*, July 2014, pp. 2850–2853.
- [258] S. Kumar, A. Sharma, K. Mamun, and T. Tsunoda, “A deep learning approach for motor imagery eeg signal classification,” in *2016 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, Dec 2016, pp. 34–39.
- [259] H. He and D. Wu, “Transfer learning enhanced common spatial pattern filtering for brain computer interfaces (bcis): Overview and a new approach,” in *Neural Information Processing*, D. Liu, S. Xie, Y. Li, D. Zhao, and E.-S. M. El-Alfy, Eds. Cham: Springer International Publishing, 2017, pp. 811–821.
- [260] H. Kang, Y. Nam, and S. Choi, “Composite common spatial pattern for subject-to-subject transfer,” *IEEE Signal Processing Letters*, vol. 16, no. 8, pp. 683–686, Aug 2009.
- [261] L. Van Der Maaten, “Accelerating t-sne using tree-based algorithms,” *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, Jan. 2014. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2627435.2697068>
- [262] G. Pfurtscheller and C. Neuper, “Motor imagery and direct brain-computer communication,” *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, July 2001.
- [263] Q. She, K. Chen, Y. Ma, T. Nguyen, and Y. Zhang, “Sparse representation-based extreme learning machine for motor imagery eeg classification,” *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1–9, 10 2018.

References.

- [264] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted boltzmann machines," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 6, pp. 566–576, June 2017.
- [265] A. Singh, A. A. Hussain, S. Lal, and H. W. Guesgen, "A comprehensive review on critical issues and possible solutions of motor imagery based electroencephalography brain-computer interface," *Sensors*, vol. 21, no. 6, 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/6/2173>
- [266] Q. Barthélemy, L. Mayaud, D. Ojeda, and M. Congedo, "The riemannian potato field: A tool for online signal quality index of eeg," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 2, pp. 244–255, 2019.