Development of Data-Driven Methods for Non-stationary Classification Problems in EEG/MEG based Brain-Computer Interfaces

Pramod Gaur

B.E., PGDIT, M.E.

A Thesis presented for the degree of
Doctor of Philosophy

Intelligent Systems Research Centre
School of Computing, Eng & Intelligent Systems
Ulster University

January 2018

I confirm that the word count of this thesis is less than 100,000 words

Contents

List of Figures

List of Tables

Dedicated to
Prof P. R. Sharma and Mrs Savitri Sharma,
to my wife and little son,
Geetika and Abhinav

Acknowledgements

I would like to thank my supervisors Professor Girijesh Prasad and Prof Hui Wang for their valuable guidance and support throughout the entire research project, without whom this work would not have been possible. I appreciate their tolerance and thanks to them for their time spent on advising me and evaluating my work. I am especially grateful to Prof Prasad for his academic and personal support.

I would like to thank Prof Ram Bilas Pachori, who has given me lots of inspiration and suggestion during my PhD as a good collaborator. I am also indebted to Dr Karl McCreadie and Dr Nazmul Siddique for their invaluable suggestions, and helpful advice as good friends. I would like to extend my gratitude to Prof Liam Maguire for his assessment and advice in my research project through the 100-day and confirmation vivas.

I wish to thank my colleagues and friends with whom I had numerous interesting discussions in the ISRC. Finally, to make this acknowledgement complete, a special thanks to my parents Prof P R Sharma and Mrs Savitri Sharma. I would thank my wife, Geetika Kaushik, who stands beside me and encourages me in all my endeavors and thanks to my dear son, Abhinav for giving happiness, joy and enduring love. Many thanks to my brother Dr Aditya Gaur, his wife Priti Sharma, and my sister Anjali Gaur and brother-in-law Mr Akshay Vashisth, my nephews Vishika and Saanvi, for cheering me up and regular chats to refresh my mind. Also, I thank my in-laws, Shri Raj Kumar Kaushik and Mrs Kamla Sharma, for providing me with their continuous encouragement and support. I would like to dedicate this dissertation to my respectable parents, son and wife.

Abstract

This thesis focuses on the development of adaptive data-driven single channel and multichannel filtering methods for brain-computer interface (BCI) systems. Magnetoencephalography (MEG) and electroencephalogram (EEG) neuroimaging recording techniques are considered to measure neurophysiological activity. The inherent nonstationarity and nonlinearity in MEG/EEG and its multichannel recording nature require a new set of data-driven single and multichannel filtering techniques to estimate more accurately features for enhanced operation of a BCI. Empirical mode decomposition (EMD) and Multivariate EMD (MEMD) are fully data-driven adaptive techniques. These techniques are considered to decompose the nonstationary and nonlinear MEG/EEG signals into a group of components which are highly localised in the time and frequency domain. Also, it is shown that MEMD based filtering can exploit common oscillatory modes within multivariate (multichannel) data. It may be used to more accurately estimate and compare the amplitude information among multiple sources which serves as a key for the feature extraction of a BCI system. These simple filtering techniques are done at the preprocessing stage which helped to reduce the effect of the nonstationarity to a large extent across the sessions for both binary class and multi-class classification problems and identify features which are somewhat invariant against the changes across sessions. Different features such as Hjorth, bandpower, common spatial pattern (CSP), sample entropy and covariance matrix are extracted in the feature extraction stage for comparative evaluation. A novel subject specific MEMD based filtering and covariance matrix as a feature set approach is introduced to classify the multiple classes using Riemannian geometry framework. This approach helped to achieve high kappa value and classification accuracy when evaluated on BCI competition IV dataset 2a. This novel type of filtering can be applied without initial calibration and has the potential to drastically improve the applicability of BCI devices for daily use. Finally, a novel tangent space based transfer learning approach is proposed which utilizes the shared structure across multiple subjects and is an important step towards zero training time for BCI systems.

List of Abbreviations

AFM	Amplitude and Frequency Modulated
AM	Amplitude Modulated
ApEn	Approximate Entropy
BCI	Brain-Computer Interface
BSP	Bispectrum
CSM	Covariate Shift Minimization
CSP	Sommon Spatial Pattern
EEG	Electroencephalogram
EEMD	Ensemble Empirical Mode Decomposition
EMG	Electromyography
EMD	Empirical Mode Decomposition
EOG	Electrooculography
ERD	Event Related Desynchronization
ERS	Event Related Synchronization
ESU	Electrosurgical Unit
FM	Frequency Modulated
GA	Genetic Algorithm
HHS	Hilbert Huang Spectrum
HMM	Hidden Markov Model
IMF	Intrinsic Mode Function
ITR	Information Transfer Rate
KL	Kullback-Leiber
KLIEP	Kullback-Leibler Importance Estimation Procedure

KMM	Kernel Mean Matching
KNN	k-Nearest Neighbour
LDA	Linear Discriminant Analysis
LR	Logistic Regression
LSIF	Least Square Importance Fitting (LSIF)
LVQ	Learning Vector Quantization
MCI	Mild Cognitive Impairments
MEG	Magnetoencephalography
MEMD	Multivariate Empirical Mode Decomposition
MEMD-SI	MEMD based subject independent design
MEMDBF	MEMD based Bandpass Filtering
MIMF	Multivariate Intrinsic Mode Function
MLP	Multilayer Perceptron
MI	Motor Imagery
MRI	Magnetic Resonance Imaging
NN	Neural Network
PSD	Power Spectral Density
PSO	Particle Swarm Optimization
SapEn	Sample Entropy
SMR	Sensorimotor Rhythm
SNR	Signal-to-Noise Ratio
SSA	Stationary Subspace Analysis
SS-MEMDBF	Subject Specific Multivariate Empirical Mode Decomposition
SVM	Support Vector Machines
STFT	Short-Time Fourier Transform
SQUID	Superconducting Quantum Interference Devices
uLSIF	Unconstrained LSIF

Declaration

"I hereby declare that for 2 years following the date on which the thesis is deposited in the Ulster University Doctoral College, the thesis shall remain confidential with access or copying prohibited. Following expiry of this period, I permit

1. the Librarian of the University to allow the thesis to be copied in whole or in part without reference to me on the understanding that such authority applies to the provision of single copies made for study purposes or for inclusion within the stock of another library.

2. the thesis to be made available through the Ulster Institutional Repository and/or EThOS under the terms of the Ulster eTheses Deposit Agreement which I have signed.

IT IS A CONDITION OF USE OF THIS THESIS THAT ANYONE WHO CONSULTS IT MUST RECOGNISE THAT THE COPYRIGHT RESTS WITH THE AUTHOR AND THAT NO QUOTATION FROM THE THESIS AND NO INFORMATION DERIVED FROM IT MAY BE PUBLISHED UNLESS THE SOURCE IS PROPERLY ACKNOWLEDGED."

_____

Pramod Gaur

Ulster University

Chapter 1

Introduction

1.1 Background

The human brain controls the entire functioning of the body parts. It holds responsibility for cognition, memory, perception, emotion, action and attention (Carlson, 2005; Purves et al., 2004). If a person performs any action such as watching television, reading, thinking, the different regions of the brain gets stimulated. This stimulus creates magnetic/electrical signals, which collectively trigger the chemical reactions which allow communication among parts of the body. These magnetic/electrical signals may be observed with various scientific technologies such as magnetoencephalography (MEG), electroencephalogram (EEG), magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), electrocorticography (ECoG) and positron emission tomography (PET). These technologies help to provide a better insight of how the activities are triggered in the human brain when a user is performing specific tasks.

EEG is the most commonly used modality to capture brain signals activity because of non-invasive nature and low set-up prices. It provides good temporal resolution, and usability (Blankertz et al., 2008; Wolpaw et al., 2000). It helps to know the state of mind to some extent an individual is in and may be measured when awake,

sleeping, and whilst anaesthetized because of the difference in electrical potentials for a specific pattern for each of these mind states.

These days EEG is extensively used for the diagnosis and treatment of brain neurodegenerative disorders, abnormalities, and mental disorders. The main challenge is how to extract specific information corresponding to a specific problem. So, we need to have some mechanism for the analysis and classification of recorded EEG signals. These essential steps allow for the development of a system which has the potential to diagnose brain diseases and to provide a better insight into associated cognitive tasks.

MEG is another noninvasive neurophysiological method which may be used to measure the magnetic fields outside the head. This technique gives more precise magnetic fields signal recorded to the femto tesla ($1fT = 10^{-15}$ tesla) unit as compared to EEG but its use has been limited by setup cost of MEG machine. Both these noninvasive recording methods are used to study the brain dynamics. They also provide the temporal changes in the sequences and activation patterns. The main difference lies in the spread of the electric field and magnetic field generated from the same electric dipole in the human brains.

EEG signals provide valuable information for differentiating among various physiological states of the brain. However, the manual interpretation of these signals is a very cumbersome and tedious task. Additionally, diagnosis of these EEG signals requires extensive skill and experience. There is a possibility that this analysis may suffer from inter-observer variability towards decision making and so there is a need to develop new techniques for automated classification systems which can expedite the automatic discrimination among these different brain states using these signals. A popular example may be brain-computer interface (BCI), which is a system which facilitates a means of communication for individuals to communicate with external assistive devices utilizing brain signals such as EEG (Wolpaw et al., 2002). In BCI, the aim is to translate the intent of a user into control command by EEG signals for a neuroprosthetics or a computer application. A popular example

of a BCI modality is motor imagery (MI) based BCI (Pfurtscheller et al., 2006, 1997). The user is expected to imagine the execution of a movement of a particular limb. The rhythmic activity is seen in the sensorimotor cortex of the brain for a specific movement in MI based BCI (Gandhi et al., 2014; Herman et al., 2008). The BCI systems identify these changes in the rhythmic activities and translate them into desired command. One of the major problems in EEG- based BCI systems is the non-stationarity which arises when EEG signals are originating from different sources. In addition, the recorded EEG signals have a low signal-to-noise ratio (SNR)(Nicolas-Alonso and Gomez-Gil, 2012). The low SNR may be due to artifacts resulting from electrooculogram (EOG) or electromyogram (EMG) interference and electrical power lines. To increase the SNR, a useful step would be to remove these distortions or artifacts from raw EEG signals before extracting the features for classification of multiple class MI tasks. Thus, this helps to achieve better feature separability corresponding to different imagined movements (Wolpaw et al., 2002). A block digram showing the basic building blocks of a BCI is shown in figure 1.1. These blocks are discussed in more detail in the next chapter.

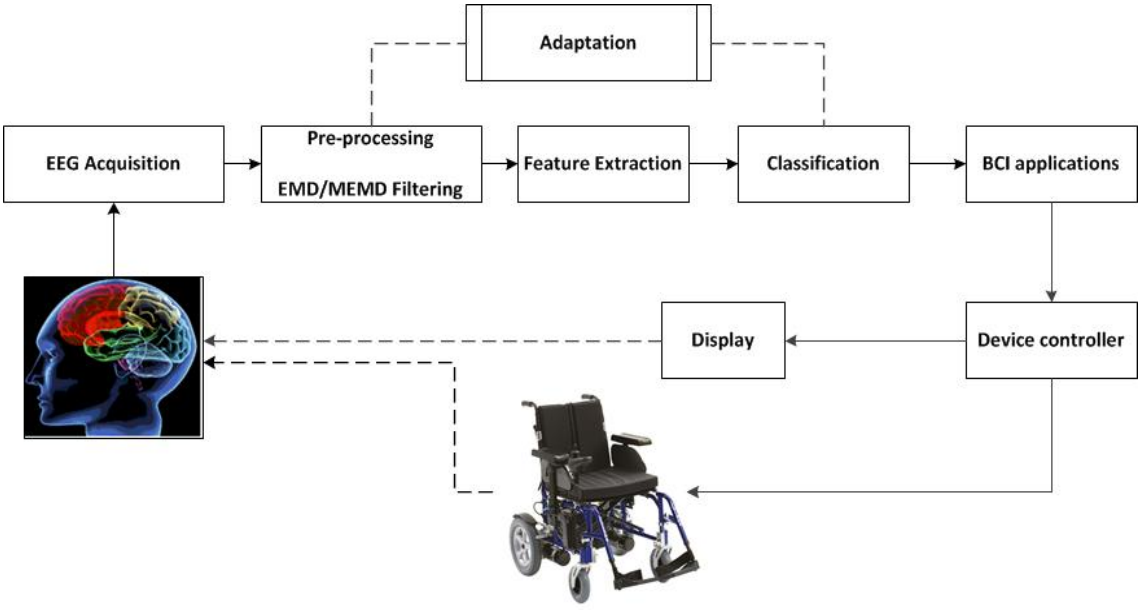


Figure 1.1: Block diagram showing building blocks of a brain-computer interface.

1.2 Aims and Objectives

The work presented in this thesis focuses on the study of MEG/EEG signal processing and classification techniques to design BCI systems as two-class and four-class classification problems. Despite the promising and valuable achievements reported in the literature the BCI field is still a moderately young research field and there is a lot of work which needs to be done to make BCI a mature technology. Among the many possible improvements, this thesis aims to address two main aspects: improving the classification accuracy for two-class and multi-class classification problems of current BCI; designing interpretable single channel and multi-channel preprocessing techniques which enable the handling of inherent non-stationarity in MEG/EEG recording techniques. This should enable a more reliable BCI systems for concrete real-life applications. The BCI research community has highlighted these points as being one of the most important and necessary research topics for the further development of BCI communication systems (Lotte et al., 2007; Wolpaw et al., 2000, 2002).

To accomplish the above aims, the main contributions of this thesis are:

Contribution 1 (C1): Development of a novel single channel filtering technique for handling non-stationarity in the preprocessing stage for classification of two-class MI based EEG signals and four-class multi-direction wrist movement MEG signals.

Contribution 2 (C2): Development of a novel multichannel filtering technique for handling non-stationarity in the preprocessing stage for two-class and four-class classification problems.

Contribution 3 (C3): To provide evidence on the performance of the novel filtering methods by varying the features and state-of-the-art classifiers.

Contribution 4 (C4): Through the development of a novel technique for the automatic identification of subject specific signal characteristics using statistical mea-

sures in both single channel and multichannel filtering techniques.

Contribution 5 (C5): Development of a novel tangent space based transfer learning pipeline for more effective utilisation of spatial information and compare the BCI performance against state-of-the-art classification methods.

1.3 Thesis Structure

The remainder of this thesis is organised as follows:

Chapter 2 introduces a state-of-the-art literature review on the basics of machine learning and BCI. It also discusses the key challenges such as nonstationarity and the low SNR in existing multi-modal recording techniques with a focus on EEG and MEG techniques. Various approaches studied by other research groups to handle the inherent nonstationarity issue have been discussed. The chapter concludes with a detailed discussion on various adaptive decomposition techniques to decompose the data into intrinsic mode functions with a particular focus on single channel empirical mode decomposition and multivariate empirical mode decomposition techniques and the learning strategies that can be implemented to solve binary class and multi-class classification problems.

Chapter 3 discusses the implementation of a significant work in the field of adaptive filtering in EEG/MEG-based BCI systems and is divided into two sections. These sections introduce two new types of filtering techniques and serve as an extension built to the empirical mode decomposition (EMD) technique. These novel filtering techniques can be used to enhance EEG/MEG signals. The EMD method decomposes the EEG signal into a group of intrinsic mode functions (IMFs). These IMFs are considered as a narrow-band, frequency modulated and amplitude modulated signals. The major challenge is to identify the components which provide major contributions to particular neurocognitive tasks. The first study introduces a novel use of the mean frequency measure to identify those IMFs

which have major contributions to mu (8-13 Hz) and beta (16-24 Hz) rhythms. The identified IMFs are combined to provide enhanced EEG signals and the remaining IMFs are discarded. The main aim of the proposed method is to filter EEG signals before feature extraction and classification to enhance the features separability and results in improved classification performance. The features namely, Hjorth and band power features computed from the enhanced EEG signals, have been used as a feature set for classification of the left hand and right hand MIs using a LDA based classification method. Significantly improved performance is obtained when the method is tested on the BCI competition IV datasets. The second study discusses a novel use of the maximum amplitude frequency measure to identify those IMFs which have major contributions to multi-direction wrist movements ($<$ 8Hz) (Tangermann et al., 2012; Waldert et al., 2008b). The identified IMFs are summed to provide enhanced MEG signals. The main aim is to filter MEG signals as a preprocessing step. The sample entropy feature has been computed from the enhanced MEG signals. The feature set has been used for the classification of multi-direction wrist movements. Improved performance is again obtained when the method is evaluated on the BCI competition IV dataset 3.

Chapter 4 contains two sections. The first section introduces a novel filtering technique, namely, MEMD based bandpass filtering (MEMDBF), which serves as an enhancement to multivariate empirical mode decomposition (MEMD) method. The MEMDBF implements multichannel filtering of IMFs based on the mean frequency measure to obtain an enhanced EEG-based BCI. The proposed method helps to handle the inherent non-stationarity and utilises the cross-channel information present in a multi-channel EEG-based BCI. Common spatial pattern (CSP) features have been computed from the filtered EEG signals with the linear discriminant analysis (LDA) used to classify the feature set into left hand and right hand motor imagery (MI). Since the EEG signals are highly subject specific and non-stationary, the second section presents a novel filtering method based on the MEMD using subject independent pooled design BCI (MEMD-SI-BCI) for the classification of MI based EEG signals to achieve an enhanced BCI. The MEMD method

helps to utilize the cross-channel information and enhanced localization properties. It decomposes multichannel EEG signals into a set of multivariate intrinsic mode functions (MIMFs). These MIMFs can be considered narrow-band, amplitude and frequency modulated (AFM) signals. The statistical property, namely, the mean frequency measure of these MIMFs has been used to combine these MIMFs to compute enhanced EEG signals. The CSP feature has been computed from the enhanced EEG signals and has been used as a feature set for classification of left hand and right hand MIs using a LDA based classification method.

Chapter 5 discusses a novel subject specific MEMD based filtering method, namely, SS-MEMDBF to classify the MI based EEG signals into two classes and multiple classes. The MEMD method simultaneously decomposes the multichannel EEG signals into a group of MIMFs. This decomposition enables us to extract the cross-channel information and also localize the specific frequency information. The statistical measure, mean frequency has been used to filter the MIMFs to obtain enhanced EEG signals which better represent motor imagery related brainwave modulations over μ and β rhythms. The sample covariance matrix has been computed and used as a feature set. The feature set has been classified into multiple MI tasks using Riemannian geometry.

Chapter 6 discusses a novel tangent space based transfer learning classification method. In the preprocessing stage, a subject specific MEMD based filtering method, namely, SS-MEMDBF is done and unseen MI based EEG trials are classified using the proposed method into two classes.

Chapter 7 concludes the thesis, wherein conclusion and recommendations are discussed along with concluding summary. The main contributions of the thesis for each chapter with the potential future research directions are also presented in this chapter.

1.4 Publications

This section provides the details of the papers that have been published or submitted to peer-reviewed journals and conferences carried out in this research work.

Journal papers

1. 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, 2018 (**contributes to chapter 5**) **(Impact factor: 3.928)**.

2. P. Gaur, R.B. Pachori, H. Wang, and G. Prasad, Improved Motor Imagery Classification in EEG based BCI using Multivariate EMD based filtering and CSP Features, Expert Systems with Applications (Manuscript number : ESWA-D-18-00786), 2018 (**contributes to chapter 4**).

3. P. Gaur, K. McCreadie, R.B. Pachori, H. Wang, and G. Prasad, Tangent space based transfer learning : Application to MI based EEG-BCI classification, International Journal of Neural Systems (submitted), 2018 (**contributes to chapter 6**).

Book chapter

1. P. Gaur, G. Kaushik, R.B. Pachori, H.Wang, and G. Prasad. Comparison analysis: single and multichannel EMD based filtering with application to BCI. International Conference on Machine Intelligence and Signal Processing, India, 2017 (**contributes to chapters 3 and 5**).

Conference / Abstract paper / Poster

1. Pramod Gaur, Ram Bilas Pachori, Hui Wang, and Girijesh Prasad. An empirical mode decomposition based filtering method for classification of motor-imagery EEG signals for enhancing brain-computer interface. In International

Joint Conference on Neural Networks, pages 1-7, 2015 (**contributes to chapter 3**).

2. Pramod Gaur, Ram Bilas Pachori, Hui Wang, and Girijesh Prasad. A multivariate empirical mode decomposition based filtering for subject independent bci. In 27th Irish Signals and Systems Conference (ISSC), pages 1-7. IEEE, 2016a (**contributes to chapter 4**).

3. Pramod Gaur, Ram Bilas Pachori, Hui Wang, and Girijesh Prasad. Enhanced Motor Imagery Classification in EEG-BCI using Multivariate EMD based Filtering and CSP Features. In Proceedings of the Sixth International Brain-Computer Interface Meeting: BCI Past, Present, and Future, 2016b (**contributes to chapter 4**).

4. P. Gaur, G. Prasad, H.Wang, and R.B. Pachori. An MEG based BCI for classification of multi-direction wrist movements using empirical mode decomposition. In MEG UK 2016, York, UK, 2016c (**contributes to chapter 3**)

5. G. Kaushik, P. Gaur, G. Prasad, H. Wang, and R.B. Pachori. An MEG based multi direction wrist movements analysis using empirical mode decomposition and multivariate empirical mode decomposition. In MEG UK 2017, University of Oxford, UK, 2017 (**contributes to chapter 5**).

6. P. Gaur, J.S. Bornot, G. Prasad, H. Wang, and R.B. Pachori. Decoding of Multi-direction Wrist Movements Using Multivariate Empirical Mode Decomposition. In MEG UK 2017, University of Oxford, UK, 2017 (**contributes to chapter 5**).

Chapter 2

Literature review

2.1 Introduction

A brain-computer interface (BCI) has the potential to positively impact upon the lives of individuals for whom conventional methods of communication or control are ineffective. However, the accuracy of a BCI can be adversely affected by the variability of the brain signals recorded not only from different subjects but also between sessions. Researchers have recently shown increasing interest in this problem of transfer learning which allows for the exploitation of previous data to enhance performance. Although traditional approaches to BCI were reliant upon user adaptation the modern approach to BCI is much more centred on placing the load on the machine to learn. It has already been demonstrated that machine learning/intelligent algorithms and adaptation can enhance the performance of a BCI when applied to any stage of the process whether that is preprocessing, feature extraction or classification. However, due to the inherently non-stationarity of the signals typically used in BCI, traditional machine learning techniques are often sub-optimal leading researchers to address the resultant adaptive learning challenges. Although methods exist such as the Fourier, Hilbert, or wavelet analysis, this chapter examines the suitability of the Empirical Mode Decomposition (EMD) and multivariate EMD (MEMD) methods as an alternative due to their robustness in

both the time and frequency domains. EMD has until recently received relatively little attention from the BCI community because it is highly suitable for non-linear and non-stationarity signals (Huang et al., 1998). But, it suffers from the mode-mixing issue discussed later and only does single channel decomposition which limits its use in real-time BCI applications. A typical real-time BCI system has a relatively high number of channels (16 or 32 or 64 channels) depending on the available recording systems. A decomposition method is required that can decompose all the channels to utilize the cross-channel information. A multivariate extension of EMD (MEMD) (Huang et al., 2003; Park et al., 2013; Rehman and Mandic, 2009; Rilling et al., 2003) has been recently proposed and has gained a lot of attention in the BCI research community. All these decomposition methods give a group of intrinsic mode functions (IMFs). There is a need for the development of an automatic method which can automatically select the IMF based on any cognitive or imagery task. This thesis proposes two filtering techniques based on EMD and MEMD based decomposition which will help to identify the components corresponding to particular cognitive or imagery task. These filtering techniques offer evidence of the effectiveness of this novel technique which addresses the non-stationarity issue in the pre-processing step.

The main aim of the thesis is to develop algorithms which are capable of classifying magnetoencephalography (MEG)/ electroencephalogram (EEG) signals into two-class and four- class. To gain insight of the classification mechanism, this chapter mainly focuses on how EEG/MEG signals are classified into different classes in the context of the BCI. Section 2.2 presents the basic structure of BCI, whereas existing feature extraction techniques studied by various research groups have been discussed in Section 2.3. This chapter also provides a detailed insight of existing classification techniques which are commonly used to classify MEG/EEG signals. Section 2.4 discusses the state-of-the-art feature classification algorithm used in BCI research and Sections 2.6.1 and 2.7 provide some details about EEG and neural oscillations. Section 2.8 discusses the non-stationarity issue persistent in the MEG/EEG signals and the different approaches that have been followed by different research groups. Sections 2.9 and 2.10 provides the details about

adaptive learning and challenges that are involved in the adaptive learning and also discusses the adaptive decomposition techniques and the chapter concludes in the Section 2.11.

2.2 Structure of BCI Design

As discussed in Pfurtscheller et al. (2010), for a device to be defined as a BCI, it must essentially meet the following stated criteria:

1. The device must rely on signals recorded directly from the brain;
2. There must be at least one recordable brain signal that the user can intentionally modulate to effect goal-directed behaviour;
3. It should involve real-time processing;
4. The user must obtain feedback.

Therefore a BCI is defined as a communication system in which messages or commands that individual sends to the external world, do not pass through the brain's normal output pathways of peripheral nerves and muscles (Wolpaw et al., 2000, 2002). In the literature, several different categories of BCI systems can be found. Amongst those a few divergent categories will be considered. Most researchers especially differentiate amongst invasive and non-invasive BCIs and dependent and independent BCIs and synchronous and asynchronous (self-paced) BCIs. BCI systems are useful to those individuals who have motor disabilities because it will help them to improve their quality of life and at the same time, their care cost will also be reduced. The main aim of BCI is to allow severely disabled people to communicate with the external world who are either 'locked-in' or paralyzed by neurological neuromuscular disorders, such as spinal cord injury, and brain stem stroke (Nicolas-Alonso and Gomez-Gil, 2012). Generally, a BCI can be treated as an artificial intelligence based system that identifies a certain set of patterns in the

EEG brain signals through a set of successive states. The modern BCI system is a closed-loop process comprising of six steps which are stated and shown in Fig. 1.1 as follow: (i) Measurement of brain activity, (ii) Pre-processing or enhancing the signal to make it suitable for further processing, (iii) Feature extraction, (iv) Feature selection, (v) Feature translation and (iv) Feedback. A brief summary of each step is now discussed below:

1. **Data acquisition**: Measuring brain activities plays a crucial role in BCI communications. The electrical signals are measured using a variety of electrodes based on the mental tasks performed. The EEG is the most widely used in BCI research (Wolpaw et al., 2000, 2002).

2. **Pre-processing**: The recorded EEG signals may have poor signal-to-noise ratio so it is always an essential step to remove the noise and artifacts which do not correspond to the acquired predefined mental tasks. Machine learning and signal processing may be used to remove the ocular artifacts, muscular artifacts or utility frequency (60 Hz or 50 Hz) and to improve the signal-to-noise ratio (SNR). Good pre-processing leads to enhanced signal quality which may result in better feature separability and help achieve higher classification performance.

3. **Feature extraction**: It is a difficult problem to classify the raw signals. In this step, features are identified which can be used to correctly discriminate the predefined related mental tasks. These can be extracted from the spatial domain, time domain, frequency domain or a combination thereof.

4. **Feature dimensionality reduction**: In BCI systems, the high dimensional feature classification problem is a commonly faced issue. In this step this problem is taken care of by first identifying the best features or the combination of features (feature selection) and then projecting the identified features from a higher dimensional space to a lower dimensional space which is more separable for the classification task (dimensionality reduction).

5. **Feature translation/ Classification (pattern matching)**: In this step, the extracted feature vectors from the signals are assigned a class. Every class corresponds to a predefined mental task. For instance, left hand and right hand could be identified as two classes for Motor Imagery (MI). The applied methods could be supervised, unsupervised, linear or non-linear.

6. **Operating Protocol**: This provides details about the following questions (i) whether the control/communication is discrete or continuous, (ii) how and when the classification will start and end, and (iii) how the feedback will be provided to the user.

2.3 Feature Extraction

In BCI design, the main aim of EEG signal processing is to translate the raw EEG signal recorded from the electrode into the imagined mental state of the user. A pattern recognition approach is utilized to achieve this translation. This is usually accomplished by two main steps: (i) feature extraction and, (ii) classification. During this step, if a feature or a combination of features extracted does not provide a better separability, this may lead to poor classification accuracy, thus it plays a crucial role in BCI applications. With respect to the design of the BCI applications, the critical properties of the EEG signals must be accounted for. These properties are:

- EEG signals have poor SNR and may contains outliers.

- Most often feature vectors are of high dimensionality because several features are extracted from several channels and across several time segments before they are combined to form a single feature vector.

- BCI features may be highly non-stationary since the characteristics of EEG signals change rapidly with time and across sessions.

The features can be extracted from EEG signals using three important sources of information as discussed below:

- **Temporal (time) information**: These features describe how the EEG signal changes with time. In practice this means using the EEG signals values in different time windows or at different time points.

- **Spectral (frequency) information**: These features mainly describe how the power changes in a given frequency bands. In practice, this means that the power computed in aforementioned frequency bands will be used as features.

- **Spatial (space) information**: These features describe the location (spatially) from where the EEG signals originate. In practice, it allows us to focus on some specific channels, or select specific EEG channels. Table 2.1 summarizes the feature extraction methods which have commonly been used in the literature.

2.4 Feature Classification

In addition to BCI, correctly classified EEG signals are widely used in the diagnosis of brain disorders or diseases and help to provide a better understanding of various cognitive processes. The recorded EEG signal contains a large amount of data and hence it is very important to extract appropriate features from the recorded EEG data, and do the classification based on the extracted features. In general, the available classifiers could be categorised based on a range of commonly used properties, such as generative-discriminative, and static-dynamic.

Generative - discriminative: In generative classifiers, in order to categorise a feature vector, firstly the likelihood of each class is computed and then the most likely one is chosen. e.g., Bayes quadratic. On the contrary, in order to correctly categorise a feature vector, a discriminative classifier only determines a criterion of discriminating the class membership or the classes. e.g., support vector machine

Table 2.1: Summary of feature extraction methods.

Time	Signal amplitude	It is most simple and efficient temporal information parameter. It is defined as the time course of the EEG signal amplitude.	
	Autoregressive parameters	It assumes that the signal measured at a time is computed as weighted sum of the values of this signal at previous time steps. Other variants are: multivariate AR parameters, Adaptive AR (AAR) parameters.	Looney et al. (2008)
	Hjorth parameters	It describes the temporal characteristics of a signal using three measures namely, the activity, the complexity and the mobility.	Gandhi et al. (2011, 2014)
Frequency	Band power features	It is computed by first filtering through a band-pass filter a signal in a specific frequency range, followed by squaring the filtered signal and finally averaging the values obtained over a given time widow.	Gandhi et al. (2014), Gao et al. (2013), Herman et al. (2008), Brodu et al. (2011)
	Power spectral density features	Notably used for motor imagery classification.	Shahid and Prasad (2011)
Space	CSP	Spatial filter designed for 2-class problems. Multiclass extensions exist.Good result for synchronous BCIs. Less effective for asynchronous BCIs. Its performance is affected by the spatial resolution. Some electrode,locations offer more discriminative information for some specific brain activities than others. Improved versions of CSP: RCSP, CSSP, CSSSP.	Coyle (2009), Von Bünau et al. (2010), Bamdadian et al. (2013), Raza et al. (2014), Lu et al. (2009)
Time-frequency	Wavelets	It is capable of providing the time and frequency information simultaneously. It gives variable resolution. Suitable for non-stationary signals. Provides both frequency and temporal information.	Gilles (2013)
	STFT	It provides fixed resolution at all times.	Park et al. (2013)

Lotte et al. (2007).

Static - Dynamic: Static classifiers, in which a time-invariant feature vector is classified so the temporal information is not considered during classification. e.g., multilayer perceptron (MLP). Dynamic classifiers are often used for classifying a combination of multiple feature vectors, and they account for temporal information. e.g., hidden markov model (HMM). For BCI system design, researchers have reported several types of classifiers such as linear classifiers, non-linear Bayesian classifiers, neural networks, nearest neighbour classifiers and an ensemble of classifiers (Lotte et al., 2007; Norani et al., 2010).

1. Linear classifiers fall under the category of discriminant algorithms. They are the most prominently used in the BCI applications, the two types are support vector machines (SVM) and linear discriminant analysis (LDA).

2. In BCI applications, the MLP is the most extensively used Neural Network (NN). It is sensitive to overtraining, especially when the EEG data is non-stationary and noisy (Bashashati et al., 2007; Lotte et al., 2007; Norani et al., 2010). Further possible types of NNs commonly used in BCIs are the learning vector quantization (LVQ) neural network and the Gaussian classifier.

3. In non-linear Bayesian classifiers, the two types commonly used in BCI applications are HMM and Bayes quadratic (Norani et al., 2010). These classifiers produce non-linear decision boundaries. Since they are generative classifiers, they reject the uncertain samples proficiently as compared to discriminative classifiers (Lotte et al., 2007).

4. Nearest neighbour classifiers are often used in BCI applications, e.g., mahalanobis distance and k-nearest neighbour (KNN). Often though this has not shown a promising performance with high dimensional feature vectors (Lotte et al., 2007).

5. A combination of classifiers (ensemble classifier) can also be used to reduce the variance which results in an increase in classification accuracy. Bagging

(Bootstrap aggregator), voting, boosting, and stacking are the classifier combination strategies typically used in BCI systems (Lotte et al., 2007). It is commonly reported that an ensemble classier outperforms a single classifier under certain conditions (Lotte et al., 2007). Adaptive weighted ensemble classifier has also been studied using a combination of classifiers (Liyanage et al., 2013).

2.5 Performance Metrics

In pattern classification, a confusion matrix contains information about the estimated and actual classifications of the trained classifier and is used to evaluate the performance of the classifier.

where, the meaning of each entry in the above confusion matrix are discussed below:

- TP is the number of correctly classified left hand motor imagery.
- FN is the number of incorrectly identified left hand motor imagery.
- FP is the number of incorrectly classified right hand motor imagery.
- TN is the number of correctly identified right hand motor imagery.

Accuracy: It is a statistical measure of how well the binary classifier correctly identifies a condition. It is a measure of both true positives and true negatives from the total number of cases examined.

$$\text{Accuracy} = \frac{TP + TN}{TN + FP + TP + FN} \times 100 \tag{2.5.1}$$

Kappa Value: This measures the agreement between two raters who each classify N items into C mutually exclusive categories.

$$\text{Kappa} = \frac{p_o - p_e}{(1 - p_e)} \tag{2.5.2}$$

where p_o is the relative observed agreement among raters, and p_e is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category.

2.6 Brain Signals

There are different types of non-invasive neurophysiological methods studied in the literature but EEG and MEG recording techniques have been considered to study brain signals in this thesis. Both of these methods have been used to study the temporal changes in the activation patterns and the brain dynamics. In the next subsections, the fundamental of these neurophysiological methods has been discussed.

2.6.1 Electroencephalography

Electroencephalography (EEG) is a practical non-invasive technique for measuring the electrical brain activity on the scalp using the electrodes (Grosse et al., 2002) and is typically used in BCI experiments. The recording is usually done using Ag-AgCl electrodes and the range for signal value is 0.5 - 100 microvolts. Following are the advantages of the EEG: (i) excellent time resolution (ii) ease of use, and (iii) low cost (Wolpaw et al., 2002). There are some challenges pertaining to the recorded EEG signal, one of which is poor SNR - the possible two sources of EEG noise could be: (a) external environment source such as lighting, power lines and a large number of electronic gadgets such as mobile phones, computers, etc. (b) physiological artifacts such as (electromyogram (EMG), muscle), (electrooculogram (EOG), eye), and (ECG, heart). There are other techniques available to record the electrical brain activity, the difference lies in the placement of the electrodes (Wolpaw et al.,

2002).

- Electrocorticography (ECoG) is a technique in which the electrodes are placed inside the skull (cortical surface) to record the electrical activity.
- Local field potentials (LFP) is a technique in which electrodes are placed inside the brain.

These methods provide better frequency bandwidth and spatial resolution. At the same time, they are difficult to use and invasive in nature as compared to EEG signals (Wolpaw et al., 2002). Although in this thesis, the focus is to use EEG signals but same technique can be applied to MEG data. The majority of the analysis later is performed on EEG data.

EEG applications

Aside from the application to BCI, EEG has been explored in a variety of other applications such as:

1. Epilepsy/ Epileptic seizure - The EEG signals of epileptic patients contain distinctive discharges of waves and spikes. Thus, it is used for diagnosing, classifying and monitoring epilepsy (Bajaj and Pachori, 2012, 2013; Sharma and Pachori, 2015).
2. Sleep disorders - EEG has been studied for classification of sleep disorders (Sharma et al., 2017).
3. Human Emotions - It is being explored for classification of human emotions like happy, sad, neutral, and fear (Ang et al., 2017).

2.6.2 Magnetoencephalography

Magnetoencephalography (MEG) is a practical non-invasive neuroimaging technique for measuring the minuscule changes in the magnetic fields produced generated by the electric current. It uses an array of sensors known as superconducting quantum interference devices (SQUID) positioned over the scalp. These SQUID can pick up the tiny magnetic fields in the order of femto tesla ($1fT = 10^{-15}$ tesla) corresponding to electrical activity with the brain. Typical Elekta MEG machine can record 306 channels including 204 gradiometers and 102 magnetometers. This technique is becoming popular these days and used in the following studies as reported in the literature: 1) Precise source localisation by combining MEG and MRI for epileptic patients. 2) Early stage diagnosis of Mild cognitive impairments (MCI) in elderly people. 3) Early stage diagnosis in children suffering from Autism. There are disadvantages associated with this technique including set-up cost and maintenance cost to keep the MEG system working. Also, portability is a major issue with this system. There are some challenges like other recording technique pertaining to the recorded MEG signal, prominently poor signal-to-noise ratio (SNR) - the possible two sources of MEG noise could be: (a) external environment generating from lighting, power lines. (b) A large number of electronic gadgets such as mobile phones, any metallic things, metallic screws in dentures etc can degrade signal quality to a large extent by inducing noise, etc. (b) physiological artefacts such as (EOG, eye), (EMG, muscle), and (ECG, heart).

2.7 Neural Oscillations

It can be useful to understand the mechanisms or neural oscillations involved in the generation of the EEG giving an indication of the underlying cognitive state (Pfurtscheller et al., 1997) which plays a crucial role in BCI research. The information in the brain is propagated by sending short electrical pulses typically known as spikes. A signal of an oscillatory nature is obtained when spikes are

superimposed from a group of neurons. The oscillations obtained are called neural oscillations and are further divided into numerous frequency bands which are fully described in Table 2.2.

Table 2.2: Summary of neural oscillations.

Frequency Band	Frequency Range (Hz)	Description
Delta	0.5 - 4	Being the slowest brain rhythm.
Theta	4 - 8	Its occurrence is not frequent in adult humans.
Alpha	8 - 13	It is the predominant wave during wakefulness.
Mu	8 - 13	Their frequency range is same as alpha. Present over the motor cortex region.
Beta	13 - 30	Mostly represent the states of attention and alertness.
Gamma	>30	They are associated with information processing.

Sensorimotor rhythm (SMR) contains valuable information about the decoded signal, and the activity is usually measured over the sensorimotor cortex as a neural rhythmic activity. The event in which there is a decrease in SMR amplitude over the cortical activity is known as event related desynchronization (ERD). On the other hand, if there is an increase in motor cortical activity then it is known as event related synchronization (ERS). Often bandpower is measured in the mu and beta frequency bands relative to the pre-stimulus baseline period. If there is significant decrement or increment in bandpower, it indicates the presence of ERD/ERS. The corresponding cognitive process is identified based on the activity in the cortical region. For example, ERD observed over the right primary motor cortex indicates that there may be a plan to move or an actual movement in the left hand.

Adaptive BCI to handle non-stationarity: In BCI research, extensive study of adaptive methods is reported in the literature. As per the current trend, the adaptation may be possible at several stages of a BCI system such as preprocessing, feature extraction, and/or classification. It has been identified that adaptation may help to overcome the adverse effect of non-stationarity (cf. 2.8, 2.8.1) in EEG signals (Schlögl et al., 2009). Specifically, there are two types of non-stationarities identified: (i) short-term changes and (ii) long-term changes. The short-term changes correlate

to different mental activities like mental arithmetic, hand movements etc. On the other side, long-term changes are correlated to changes in the recording conditions, fatigue and effects of feedback training.

2.8 What is Non-stationarity

An important question, and a key focus of this thesis, is whether or not it is possible to find out if the given time-series is non-stationary. In order to address this concern, two properties are identified in a labelled time-series where the distributions are different. For instance, in the case of multivariate normal distribution, computing the first two moments of the data, (i.e. mean and covariance) and checking how they vary with time will identify the non-stationarity. It is always important to check a reasonable significance level for the underlying distribution. To check the difference between the two probability distributions we can do statistical hypothesis testing which is defined as a method of statistical inference. It is used to make the decision between a null hypothesis H_0 and the alternative hypothesis H_1. In a given hypothesis, sample information must be summarized by test statistics. The critical region is evaluated based on a given level of significance (α). The null hypothesis H_0 is rejected if the test sample is present in the critical region.

2.8.1 Approaches to handle non-stationarity

The traditional machine learning algorithms typically presume the stationary nature of the data, which often leads to deteriorated performance because of the inherent non-stationarity in EEG-based BCI (Krauledat et al., 2007; Shenoy et al., 2006). Several possible causes of changes in signal properties during inter-session and intra-session could be fatigue, change in impedance or placements of the electrodes. When the training and testing data are recorded on different days this could also be one possible reason which might lead to performance deflation (Krauledat et al., 2007). In the last few years, researchers have proposed

several algorithms to mitigate the effect of the inherent non-stationarity present in EEG-based BCI applications. They are broadly classified into two categories: (i) The algorithms that ameliorate the model to become invariant and robust against the changes (Krauledat et al., 2007; Lotte and Guan, 2011; Samek et al., 2012; Tomioka and Müller, 2010; Von Bünau et al., 2009, 2010). The focus of most of these algorithms was to extract the invariant features by regularizing the common spatial patterns algorithm (Krauledat et al., 2007; Lotte and Guan, 2011; Samek et al., 2012). Moreover, researchers have improved the performance by identifying and then extracting the stationary segment from the EEG signal, then applied the common spatial pattern (CSP) algorithm named as the stationary subspace analysis (SSA) algorithm (Samek et al., 2012; Von Bünau et al., 2009, 2010). (ii) The algorithms which make the model adapt to the changes (Li et al., 2010; Sugiyama et al., 2007; Vidaurre et al., 2007, 2011, 2008). Researchers have shown that by doing bias adaptation, a simple adaptation process, BCI performance can be significantly enhanced (Shenoy et al., 2006; Vidaurre et al., 2011). In some existing works, a few researchers have primarily focused on adapting the feature space (Chen et al., 2010; Li et al., 2009; Sun and Zhang, 2006) and some proposed techniques focused on adapting the classifier space (Schlögl et al., 2009; Shenoy et al., 2006). In general, one can mitigate the effect of inherent non-stationarity in EEG signals by (i) projecting to stationary subspaces (Von Bünau et al., 2009), (ii) constructing invariant features (Wojcikiewicz et al., 2011), (iii) tracking non-stationarity (Schlögl et al., 2009; Vidaurre and Blankertz, 2010) or by (iv) modelling non-stationarity and using adaptive cross- validations schemes (Sugiyama et al., 2007). Recently researchers have proposed EEG data space adaptation which reduces the session-to-session non-stationarity in EEG-based BCI applications. They have proposed both supervised and unsupervised versions (Arvaneh et al., 2013). The key idea is "*to compute a linear transformation that maps the EEG data from the evaluation session to the training session, such that the distribution difference between these sessions is minimized*."

Covariate shift adaptation: In classical supervised learning, it is presumed that input data points in the training and testing phases should follow the same prob-

ability distribution, but in real-world applications this constraint often fails. For instance, in non-stationary environments, the input data points in the training and testing phases have different probability distribution (Sugiyama, 2012). Thus, the covariate shift is defined as "*the situation where the training input points and test input points follow different probability distributions, but the conditional distributions of output values given input points are unchanged*" (Sugiyama et al., 2007). In covariate shift, when the test points are situated outside the training samples, it might lead to the extrapolation problem because only training points are used for learning the function. In a recently reported algorithm, the term "*importance*" plays a very crucial role in covariate shift adaptation (Sugiyama et al., 2008). It is the ratio of the test and training probability density functions which is bounded $\frac{P_{tst}(x)}{P_{tr}(x)} < \infty$ for all x. However, evaluating the density estimation for high dimensional data is supposed to be hard from the computational complexity point of view (Sugiyama et al., 2008). As stated in Vapnik's principle "*avoid solving more difficult intermediate problem when solving a target problem*" (Moreno-Torres et al., 2012; Quionero-Candela et al., 2009). Thereafter, researchers developed new techniques called direct importance estimation techniques which include kernel mean matching (KMM), logistic regression (LR), Kullback-Leibler importance estimation procedure (KLIEP) (Kanamori et al., 2009; Sugiyama et al., 2008), least square importance fitting (LSIF), and unconstrained LSIF (uLSIF) (Kanamori et al., 2009). Each of the above methods has advantages and disadvantages associated with them in terms of model selection, optimization and density estimation. In online learning, in order to improve the system's performance, an unsupervised adaptation method called covariate shift minimization (CSM) has been presented (Satti et al., 2010). In this method, the feature set distribution is examined to find the covariate shift amongst the feature distributions of the training data and the test data (Satti et al., 2010). In order to tackle the non-stationarity present in EEG signals, an unsupervised adaptive classifier is used by Vidaurre et al. (2011). It can be applied to diversified fields in BCI applications because label information is not mandatory. The three types of adaptation methods are as follow: "*(i) supervised adaptive LDA (ii) unsupervised adaptive LDA I: common mean changes (iii) unsupervised adaptive LDA II: common mean*

changes and covariance changes" (Vidaurre et al., 2011).

As can be seen from the previous discussion there are several performance deteriorating factors in EEG-based BCI such as non-stationarity, low SNR etc. The EEG signals can have low SNR due to electrical power line and other artifacts resulting from EMG or EOG interferences. The removal of these artifacts or distortions from EEG signals before extracting features for classification of MI tasks is a useful step in order to increase SNR (Nicolas-Alonso and Gomez-Gil, 2012).

Several attempts have been made to understand the dynamics of EEG signals by exploring different frequency bands, such as μ (8-13 Hz) and β (13-25 Hz) rhythms. It is a fact that the responses and topographies obtained from the beta (β) rhythm is distinct as compared to the mu (μ) rhythm corresponding to limb movements. It has been shown during limb movements, that there is normally an increase in the oscillatory power of the beta rhythm observed in the ipsilateral sensorimotor cortex and simultaneously there is a decrease in oscillatory power of the mu rhythm observed in the contralateral sensorimotor cortex (Gandhi et al., 2014; Herman et al., 2008). The BCI systems identify these changes to provide some meaningful command. One of the major issues in BCI systems is the intrinsic non-stationarity present in the EEG signals which happens when these signals originate from different sources. To increase SNR, the most useful step would be to enhance the EEG signals by eliminating these distortions or artifacts from the raw EEG signals in the preprocessing stage. This step will help to obtain better separability in the feature set corresponding to multiple MI tasks (Wolpaw et al., 2002).

An extension method based on CSP has been studied to handle the adverse results of intervention from noisy EEG signals (Lemm et al., 2005). A Bayesian learning method has been implemented for spatial filtering in (Zhang et al., 2013) for handling EEG signals with extremely low SNR. The methods built on the self-organizing fuzzy neural network (SOFNN) and the neural network (NN) concept have also been proposed to attain better feature separation for MI tasks in MI based

BCI (Coyle, 2009; Coyle et al., 2005, 2009). Recently, a filtering technique based on the quantum neural network has been proposed before the feature extraction step in Gandhi et al. (2014, 2015) to gain better separation between classes.

An empirical mode decomposition (EMD) technique is also well suited for analysis of non-stationary and non-linear signals (Sharma and Pachori, 2015; Sharma et al., 2015b). This method is data dependent and adaptive in nature. It gives a group of intrinsic mode functions (IMFs). These IMFs are considered as narrow-band amplitude and frequency modulated (AFM) signals. Univariate EMD suffers from the problem of mode-mixing wherein similar frequencies occur in different IMFs (Park et al., 2013). To overcome this issue, a multi-channel version namely, multivariate EMD (MEMD) has been investigated to show the comparison with univariate EMD to classify different MI EEG signals considering all the IMFs for use in BCI (Davies and James, 2013, 2014; Park et al., 2013, 2014). The MEMD allows a high localization of information pertaining to specific frequency bands. It decomposes the raw EEG signal into a finite set of frequency modulated (FM) and amplitude modulated (AM) components known as multivariate IMFs (MIMFs) (Park et al., 2013). It also provides the same number of IMFs for all the data channels in the time domain.

In BCI research community, there are different variants of CSP algorithm studied and used by several groups (Ang et al., 2012; Zhang et al., 2013) to extract more separable spatial patterns as features. In this thesis, sample covariance matrix is exploited as feature set, as the sample covariance matrix contains the spatial information present in EEG signal. The main objective is to devise a unique step by combining the spatial filtering and the classification. However, sample covariance matrices structure needs to be handled carefully in Riemannian manifold. In this respect, a rich framework is facilitated by Riemannian geometry (Barachant et al., 2012) to handle these matrices.

This thesis seeks to address the inherent non-stationarity in the EEG by enhancing the EEG/MEG signals using the two proposed filtering techniques for single chan-

nel and multi-channel EEG/MEG signals in the preprocessing step as discussed in the later chapters.

2.9 Adaptive Learning Challenges

There are some major issues which draw the attention of the BCI researchers, namely, non-stationarity (cf. 2.8, 2.8.1) in intra and inter session recordings, the bias-variance trade-off and the curse-of-dimensionality. In machine learning research community, last two problems are very common and a lot of work has been done by researchers to address these challenges. These days, non-stationarity is attracting lot of attention from different researcher groups. Indeed, the focus of this thesis is also on handling non-stationarity in BCI systems.

- **Non-stationarity**: Non-stationarity is often seen in brain signals between inter-session transfers, also known as a covariate shift and can cause the classifier performance to deteriorate with the time (Mohammadi et al., 2013; Satti et al., 2010; Sugiyama et al., 2007). In BCI, several researchers have proposed various non-stationarity adaptation techniques for handling inter-session non-stationarity in EEG signals. Those techniques have been covered in greater depth in section 2.8 and 2.8.1.

- **The curse-of-dimensionality**: If the training set is small and the dimensionality of the feature vector is high, it is known as the curse-of-dimensionality. This is a major issue in BCI systems. In order to improve the system performance, the system needs to undergo training very often which is not favourable for most of the subjects. Indeed, retraining the system very frequently is not a good indication because it involves subject time and extra effort, and at the same time professional supervision to ensure that training happened under suitable conditions.

- **Bias-variance trade-off**: The variance shows the sensitivity to the input train-

ing dataset used. The bias describes the difference between the best mapping and the estimated mapping, and is highly dependent on the method chosen such as linear, quadratic etc. EEG-based BCI tends to have inherent non-stationarities, so some mechanism is needed to keep the variance very low. Invariably, EEG signals in session-to-session transfer suffer from high variance and high bias (Lotte et al., 2007). Generally, the EEG signals in multiple sessions may also suffer from both high variance and high bias. Hence the challenge is to have low variance and low bias to gain a better classification accuracy.

In practical real-world situations, still there are open research challenges which deal with EEG signal processing because of the noise introduced during recording, inherent non- stationarity, signal complexity and the amount of data available during the training phase (Lotte, 2014). In EEG-based BCI, the biggest challenge is to handle the effect of the non-stationarity caused by the inter-session transfers. In order to reduce the effect of non-stationarities, the approaches discussed in section 2.8 and 2.8.1 use either an adaptation method or by selecting the non-stationarity generating process and then handling it by taking suitable corrective action. In this respect there is a need to handle the adverse effect of the non-stationarity present in the real-world environment. This leads to a novel research program to be explored for an adaptive learning model for an evolving system in a non-stationary environment. In an attempt to take full advantage of present methods to build an adaptive learning model, different facets of non-stationary environment such as adaptive decomposition, dataset shifts and adaptive learning must be investigated. For this reason, the adaptive decomposition technique and adaptive classifiers are vital in achieving this goal. The adaptive decomposition is highly suitable for non-stationary signals because it provides a set of IMFs which can be considered as narrow-band amplitude and frequency modulated (AFM) signals for analysis.

2.10 Adaptive Decomposition

There are many types of decomposition techniques available in the literature such as empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), and multivariate empirical mode decomposition (MEMD). We have only considered single channel decomposition (EMD) and multivariate extension of this decomposition. These approaches have been selected based on the nature of the available EEG/MEG dataset.

2.10.1 Empirical Mode Decomposition

In this section, the background of the EMD algorithm is discussed and its ability to work at the level of instantaneous frequency and will demonstrate an example of how EMD decomposition works on EEG/MEG data.

Background: Data analysis is an indispensable part of both practical engineering and pure science. When attempting to develop a numerical model to solve a real-world data problem, there may be several issues in the data to estimate the parameters such as:

- nonstationarity of the data
- nonlinearity of the data
- short length of the trial

In most applications spectrum analysis is commonly used. Additionally, Fourier spectral analysis is also widely implemented due to both the calculation speed and its simplicity but there are certain restrictions when applied on real-world data because it is intended for periodic and stationary data, and linear systems.

The stationarity of a time-series X(t) in the strict sense is defined if the joint distributions

of

$$[X(t_1), X(t_2), ..., X(t_n)] \; and \; [X(t_1+\tau), X(t_2+\tau), ..., X(t_n+\tau)] \tag{2.10.3}$$

are the same for all t_i and τ.

The definition of stationarity in the wide sense is

$$E(|X(t)^2|) < \infty, E(X(t)) = m, C(X(t_1), X(t_2)) = C(X(t_1+\tau), X(t_2+\tau)) = C(t_1-t_2) \tag{2.10.4}$$

where $E()$ denotes the expected value operator and $C()$ gives the covariance operator. Due to the limitation of trial length, limited datasets can satisfy this stationarity condition in real-world.

Several methods have been proposed to handle the non-stationarity present in neurophysiological data, for example, "spectrogram" and "wavelet analysis". The spectrogram serves as a special case of Fourier spectral analysis which works on a short time segment of datasets. By moving the window over the entire time span, multiple sets of frequency spectrums are obtained. These spectrums are combined in a time-frequency distribution.

Most of the analysis depends on the traditional Fourier analysis but it may not be the best solution unless the neurophysiological data are stationary in every window. This sounds very unrealistic in a real-world signal.

Secondly, the wavelet analysis is a linear analysis technique. It facilitates a uniform resolution for all the scales, which relies on the size of the basic wavelet function (Morse wavelet). The basic definition of wavelet analysis is given as

$$Wf(c, d; X, \psi) = |c|^{-1/2} \int_{-\infty}^{\infty} X(t)\psi^*(\frac{t-d}{c})dt \tag{2.10.5}$$

where ψ^* gives the basic wavelet function, c denotes the dilation factor and d provides the translation of the origin.

The most commonly used wavelet in the wavelet family is the Morse wavelet[1] which has the problem of leakage, and which is generated by the limitation of the basic wavelet function length. Due to this problem, it is difficult to define the energy-frequency-time distribution quantitatively.

In this chapter, the single channel empirical mode decomposition (EMD) method will be discussed as an alternative to the conventional Fourier and wavelet analysis. The EMD is a fully data-driven operation for obtaining a highly localised time-frequency estimation for a nonlinear and nonstationary signal (Huang et al., 1998), by decomposing it into a finite set of AM/FM components, intrinsic mode functions (IMFs). Also, multivariate version of EMD will be also studied to utilise the cross-channel information present across channels and to achieve highly localised time-frequency estimation across channels.

Table 2.3: Comparison of EMD with other state-of-the-art methods.

	Fourier	**Wavelet**	**Hilbert (EMD)**
Basis	Apriori	Apriori	Adaptive
Presentation	Frequency- energy	Time-frequency-energy	Time-frequency-energy
Frequency	Convolution: global uncertainty	Convolution: regional uncertainty	Differential: local certainty
Feature extraction	No	Continuous : yes, discrete : no	Yes
Non-stationarity	No	Yes	Yes
Theoretical base	Theory complete	Theory complete	Empirical

Algorithm

In order to qualify as an IMF, it must satisfy two mandatory conditions: (i) the number of extrema and the number of zero crossings must differ by at most one or be the same, (ii) at any point, the mean value of the envelopes defined by the local maxima and the local minima is zero. The EMD algorithm (Flandrin et al., 2004; Huang et al., 1998) for a signal $y(t)$ can be summarized by the following sifting process:

[1]The Morse wavelet is the most commonly used wavelet transform, and had been used for the performance comparison with EMD in (Huang et al., 1998).

(1) Assume $i_1(t) = y(t)$

(2) Determine the extrema (maxima and minima) from the $i_1(t)$.

(3) Compute the upper envelope $E_{\text{max}}(t)$ and lower envelope $E_{\text{min}}(t)$ by interpolating the maxima and minima using cubic spline interpolation respectively.

(4) Compute the local mean of $E_{\text{max}}(t)$ and $E_{\text{min}}(t)$ as:

$$\text{mean}(t) = \frac{E_{\text{max}}(t) + E_{\text{min}}(t)}{2} \tag{2.10.6}$$

(5) Substract mean(t) from the original signal $y(t)$ as:

$$i_1(t) = i_1(t) - \text{mean}(t) \tag{2.10.7}$$

(6) Check whether the $i_1(t)$ is an IMF by applying the aforementioned two basic conditions of IMF.

(7) Repeat steps $(2) - (6)$, until an IMF $i_1(t)$ is determined.

Once the first IMF is obtained, define the $\text{M}_1(t) = i_1(t)$ which tends to have the smallest temporal scale in the signal $y(t)$. In order to determine the remaining IMFs, the residual signal $R_1(t) = y(t) - \text{M}_1(t)$ can be treated as a new signal. The above mentioned sifting process is then repeated until the final residual obtained becomes monotonic function from which no further IMFs can be obtained. After obtaining all IMFs, the original signal $y(t)$ can be expressed as a sum of these IMFs and final residual (Huang et al., 1998):

$$y(t) = \sum_{q=1}^{N} \text{M}_q(t) + R_N(t) \tag{2.10.8}$$

where N is the number of extracted IMFs and $R_N(t)$ is final residual. The signal $y(t)$ can also be approximated as sum of amplitude and frequency modulated (AFM) sinusoids (Huang et al., 1998):

$$y(t) \approx \sum_{q=1}^{N} a_q(t) \cos[\phi_q(t)] \tag{2.10.9}$$

where $a_q(t)$ is amplitude envelope and $\phi_q(t)$ is instantaneous phase of q^{th} IMF of the signal $y(t)$.

However, EMD suffers from the mode-mixing problem. Also, the cross-channel information present across the channels is not utilized because it does decomposition of one channel at a time. To handle these issues, a lot of other variants have been proposed. To overcome the mode-mixing problem, EEMD technique was proposed in (Wu and Huang, 2009). Unfortunately, EEMD is a time-consuming method and may add noise to the original signal. In addition, it may not be suitable for real-time implementation of the proposed algorithm for this thesis with more number of channels. Further, Rehman and Mandic have proposed a multivariate version of the EMD method utilizing cross-channel information called MEMD (Park et al., 2013). It is not only suitable for dealing with multichannel signals but also solves the problem of mode-mixing by adding white Gaussian noise to different channels.

Fig. 2.1 displays all of the obtained IMFs of an EEG signals for a single channel. It should be noted that the first IMF, IMF_1, shows the fastest oscillation whereas IMF_7 gives the slowest oscillation of the EEG signal. Fig. 2.2 displays the magnitude of the fast Fourier transform for all the IMFs obtained using EMD method. A single trial EEG signal of subject A01 from BCI competition IV dataset 2A has been considered for demonstration purpose.

To investigate the power resolution of the Hilbert Huang spectrum (HHS), the frequency shift sin waves of 8 Hz and 13 Hz are considered shown in Fig. 2.3(a), where the frequency of the sin wave is changed from 8Hz to 13Hz at 4 seconds. The sampling rate is 250Hz. The signal HHS is compared with conventional time-frequency analysis methods, short-time Fourier transform (STFT), Morse wavelet transform. It is evident from the Figures 2.3(b), 2.4(a), and 2.4(b) shows the shifting pattern of frequency components are well estimated using all these three methods. The EMD method has enabled to achieve highly localised frequency modulated signal and the transition of frequency change is very smooth. This method doesn't

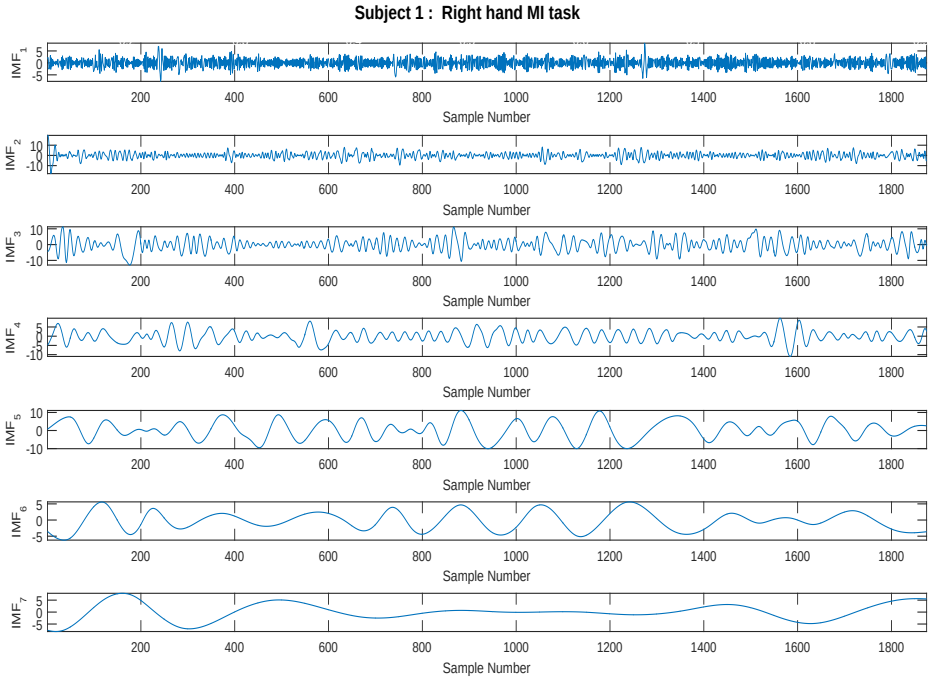


Figure 2.1: The EEG signal of subject A01T for the right hand movement and its first seven IMFs.

require any apriori knowledge about the nature of the neurophysiological signal hence, it is adaptive in nature. Comparison of different techniques has been shown in Table 2.3. The other methods spread over a wide range as compared with HHS which provided more localised time-frequency components.

In the next section, more details about the MEMD will be discussed in a greater detail.

2.10.2 Multivariate Empirical Mode Decomposition

As just described, the EMD is a data-driven technique to decompose a signal into a finite set of band limited basis functions called IMFs (Huang et al., 1998). The MEMD was recently developed, where instead of computing the local mean using the average of upper and lower envelopes like conventional EMD, the multiple $n-$dimensional envelopes are generated by projecting the signal along every di-

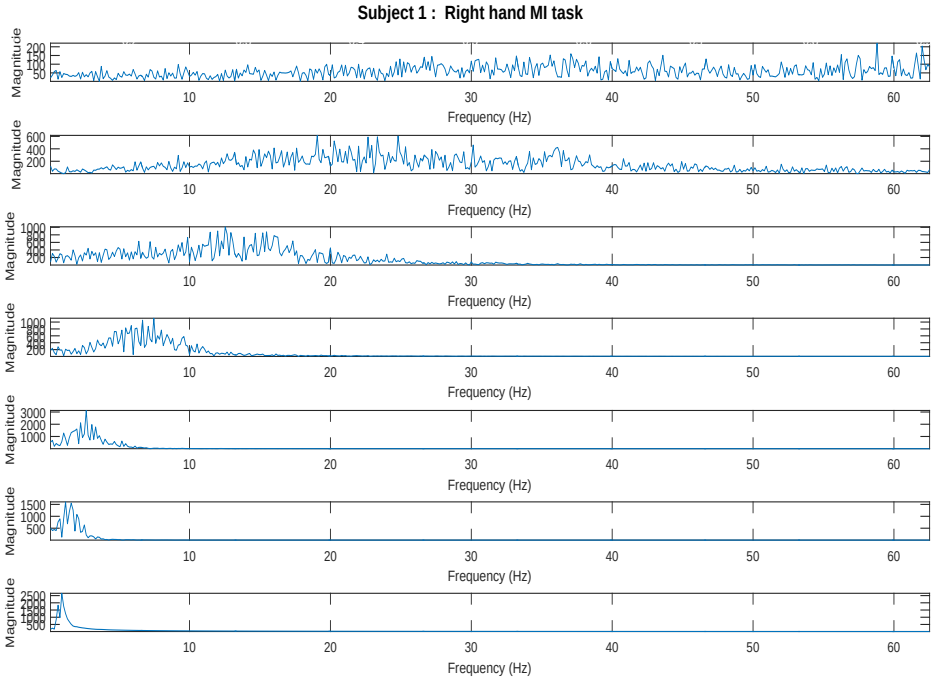
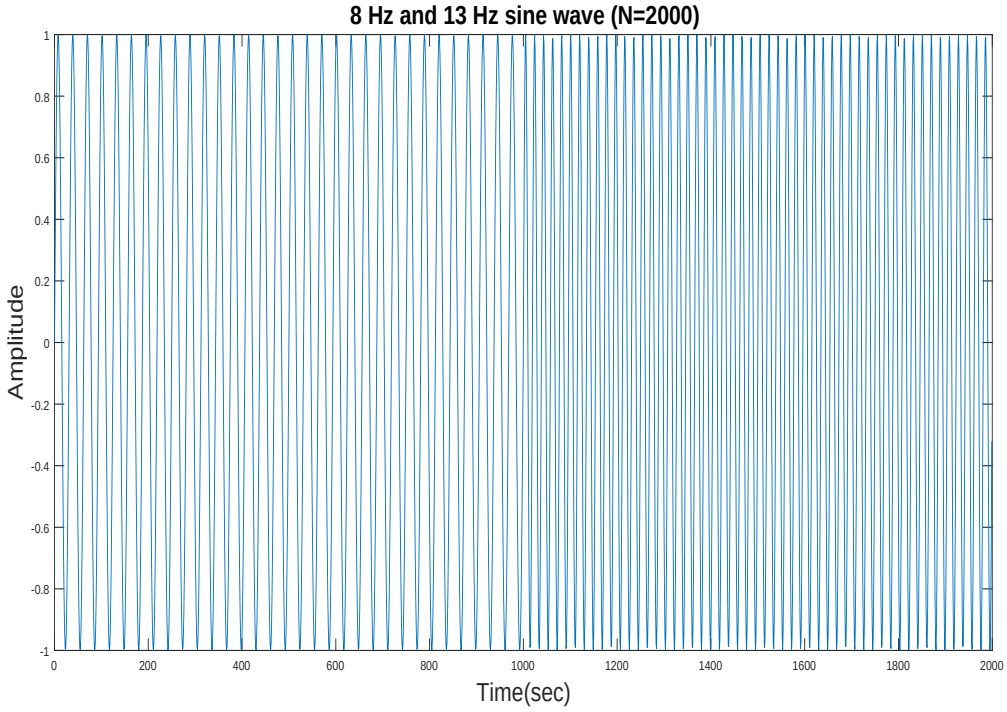


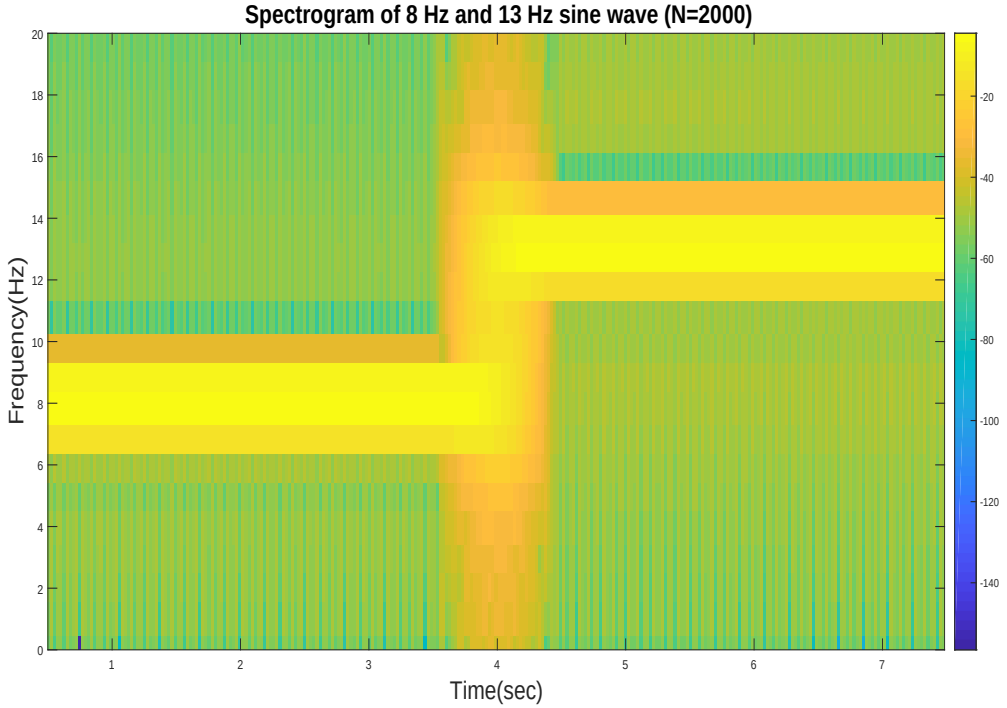
Figure 2.2: The magnitude of the FFT of subject A01T for the right hand movement and its first seven IMFs.

rections in $n-$variate spaces. These projections are averaged to obtain the local mean. As discussed previously in section 2.9, EEG signals tend to have a low SNR and may suffer from interference from EMG, EOG, or electrosurgical units (ESU) (Pfurtscheller et al., 1997). The EEG signals of interest corresponding to μ and β rhythms may contain noise which can cause erroneous results. Hence, a method is required that does not undermine the original signal and can filter out noise. In 1998, Huang et al. proposed EMD which decomposes the original signal into a finite set of band limited basis functions which are known as IMFs (Huang et al., 1998), given by Equation 2.10.8.

Later, in 2013 they proposed a noise-assisted MEMD (N-A MEMD) method (ur Rehman et al., 2013), which is not only suitable for dealing with multichannel signals, but also solves the problem of mode-mixing by adding white Gaussian noise to different channels. In the computation of N-A MEMD, the mean $M(t)$ is calculated by means of the multivariate envelope curves, expressed as follows (ur Rehman et al.,

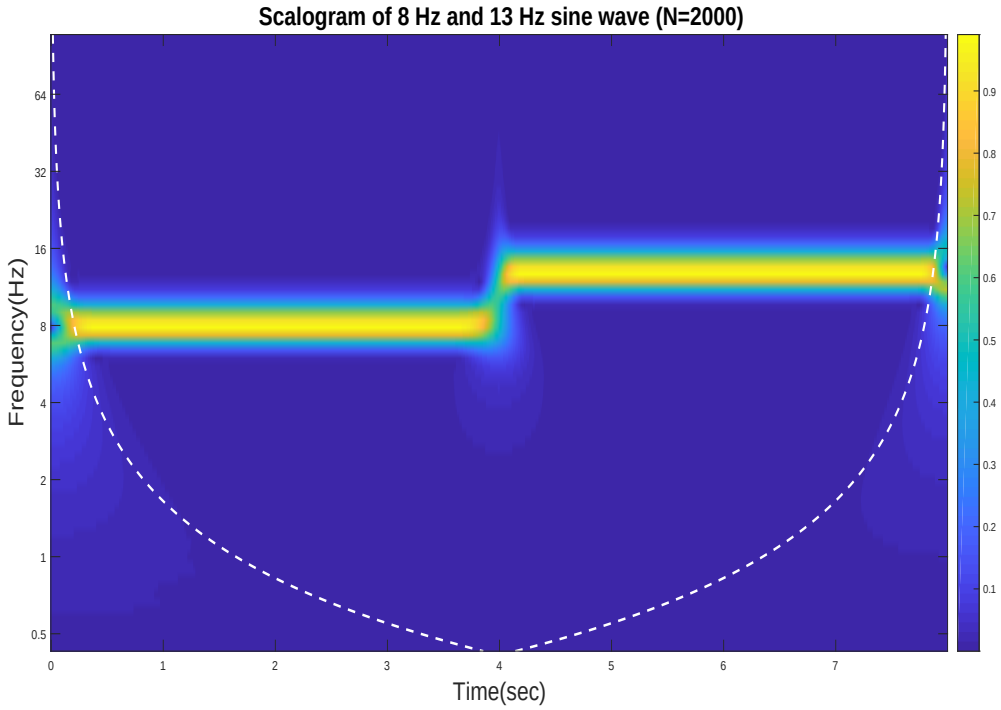


(a)

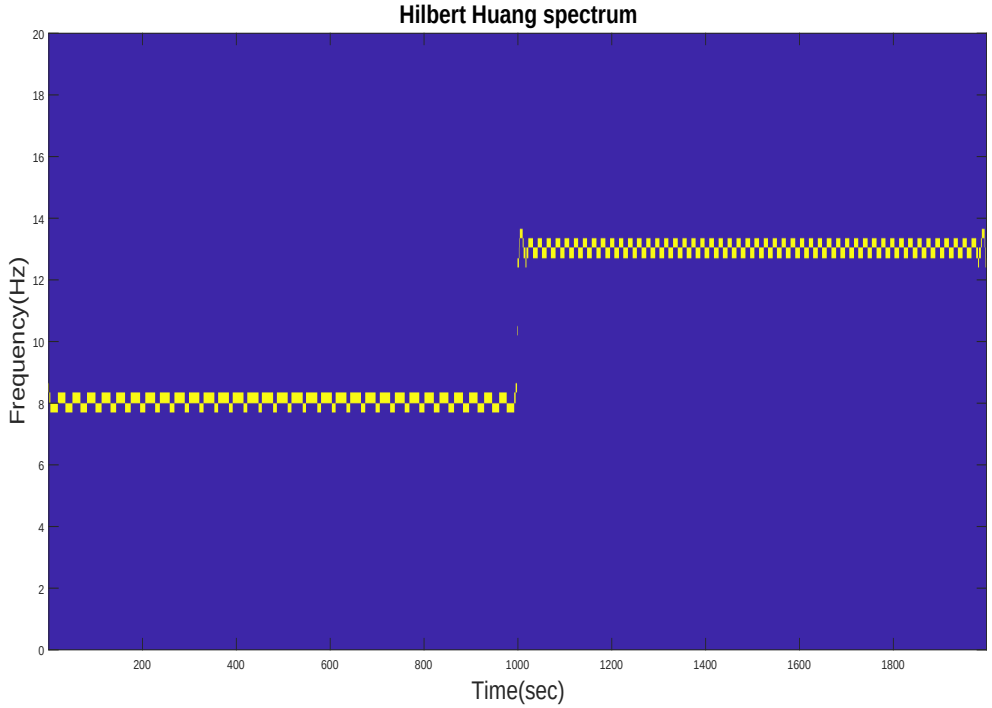


(b)

Figure 2.3: Time-frequency representations of frequency shift in sin waves of 8 Hz and 13 Hz. (a) Signal in time domain, (b) spectrogram produced for sin waves of 8 Hz and 13 Hz.



(a)



(b)

Figure 2.4: Time-frequency representations of frequency shift in sin waves of 8 Hz and 13 Hz. (a) Morse wavelet spectrum, (b) Hilbert Huang spectrum (EMD). Please note the localized time-frequency representation obtained using EMD.

2013):

$$Y(t) = \frac{1}{m}\sum_{j=1}^{m} e^{\theta_j}(t) \tag{2.10.10}$$

where $e^{\theta_j}(t)$ are the multivariate envelope curves for whole set of direction vectors, here m gives a set of direction vectors and j is length of the vectors (Park et al., 2013). Then, the candidate IMF $R(t)$ by $R(t) = X(t) - Y(t)$ is computed. If the candidate IMF satisfies the stoppage criterion, the candidate IMF becomes the multivariate IMF. If not, the input $X(t)$ will equal the remainder $R(t)$ and the remainder is computed again. The whole process is repeated until all of the multivariate IMFs are extracted. Regarding the stoppage criterion, this is similar to the original EMD proposed (Huang et al., 1998) using decomposing signal until the signal becomes monotonic or no more IMFs can be derived (Huang et al., 1998).

Fig. 2.5 displays all of the obtained IMFs of an EEG signals for multiple channels namely, FCz, C3, and Cz. As mentioned earlier, the first IMF, IMF_1, shows the fastest oscillation whereas IMF_9 gives the slowest oscillation of the EEG signal across all the three channels. Only three channels have been considered for demonstration purpose of the provided twenty-two channels.

The nonstationary and nonlinear EEG signals are well decomposed using EMD and MEMD methods. The seven IMFs are obtained adaptively using EMD method without any assumption about basis functions like Fourier analysis. The same rule applies for the MEMD decomposition as well.

In this thesis, filtering techniques built on EMD and MEMD will be used to enhance the EEG/MEG signals to increase feature separability as discussed in the later contribution chapters.

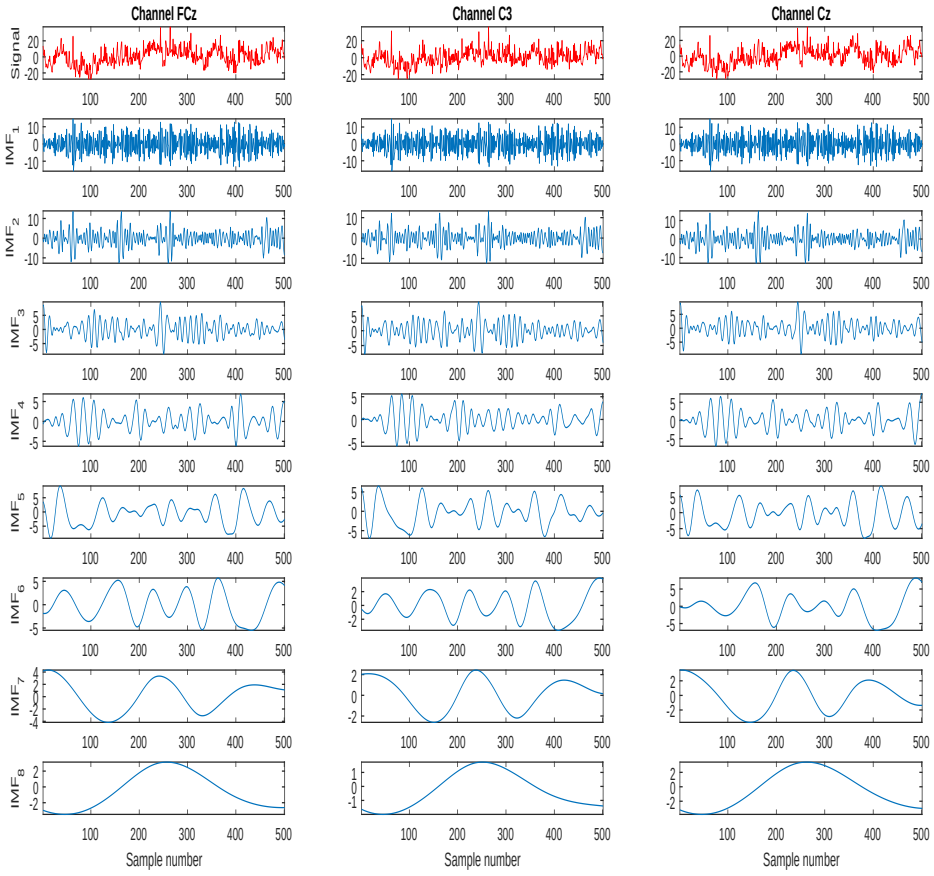


Figure 2.5: The EEG signals of subject A01T for the foot movement and its first nine IMFs.

2.11 Conclusion

The brain-computer interface offers those with debilitating conditions the possibility of communication or control. However, the inherent non-stationarity present in the signals typically used in BCI remains an ongoing challenge for researchers. This chapter has examined not only modern transfer learning techniques, but also existing methods to tackle the non-stationarity challenge by examining the Fourier, Hilbert, and wavelet analyses. This chapter also addresses why there is a need for the development of a method which can automatically select the IMF based on the any cognitive or imagery task. In the next few chapters, the discussion will focus on how the EEG/MEG can be enhanced using these decomposition methods and how to efficiently utilize different components to enhance the EEG/MEG signals and its application to two-class and four-class MI based BCI and four-class wrist movement classification problems. Chapter 3 primarily focuses on contribution C1, where the novel single channel EMD based filtering technique will be used to enhance the EEG/MEG signals and how the different statistical measures may be utilized to identify the components for a particular task.

Chapter 3

Empirical Mode Decomposition based Filtering

3.1 Introduction

This chapter seeks to address contribution C1 of this thesis by developing a novel single channel empirical mode decomposition (EMD) filtering technique for handling non-stationarity in the pre-processing stage as discussed in the previous chapter. This chapter is divided into two sections with the first section describing a study which implements this EMD based filtering method for enhancing performance of a two-class motor imagery based brain-computer interface (BCI) using electroencephalography (EEG) data. The second section extends the study and also addresses C1 by applying the techniques introduced in the first section and implements this novel filtering method on magnetoencephalography (MEG) data for classification of multi-direction wrist movements.

The results from both studies provide evidence for the effectiveness of EMD for improving the classification accuracy when applied, not only to the widely used EEG-based BCI competition datasets 2A and 2B, but also when classifying motor imagery (MI) from BCI competition 3 MEG dataset. These two studies together

demonstrate the effectiveness of the EMD method when applied to both EEG and MEG data for enhancing the performance of a BCI.

3.2 Study 1: EMD 2-class EEG MI

3.2.1 Methods

As discussed and mentioned previously (cf. 2.10.1), the EMD method breaks EEG signals into a set of intrinsic mode functions (IMFs) (Huang et al., 1998) which can be considered narrow-band, Amplitude and Frequency Modulated (AFM) signals. A novel single channel filtering method is built as an extension to the EMD (Huang et al., 1998; Jia et al., 2011) for enhancement of EEG signals before extracting features for classification of left and right hand MIs. For the first time in the literature, a novel use of mean frequency is proposed to first automatically identify these IMFs and further sum up the identified IMFs to obtain enhanced EEG signals corresponding to the mu and beta rhythms. The Hjorth and band power features are then computed from the enhanced EEG signals. These features are then classified into left and right hand MIs using an linear discriminant analysis (LDA) classifier. The proposed methodology consists of three major steps namely, EMD based EEG signal enhancement, feature extraction, and LDA classifier. A block diagram of the proposed method is shown in Fig. 3.1.

3.2.2 Dataset

BCI competition IV dataset 2A description

This dataset consists of EEG signals performing four different MI tasks: movements of the left hand, right hand, feet, and tongue from nine healthy subjects. The dataset contains two sessions, one for training and one for evaluation. The sessions were recorded on different days for each of the subjects. Each session was recorded with

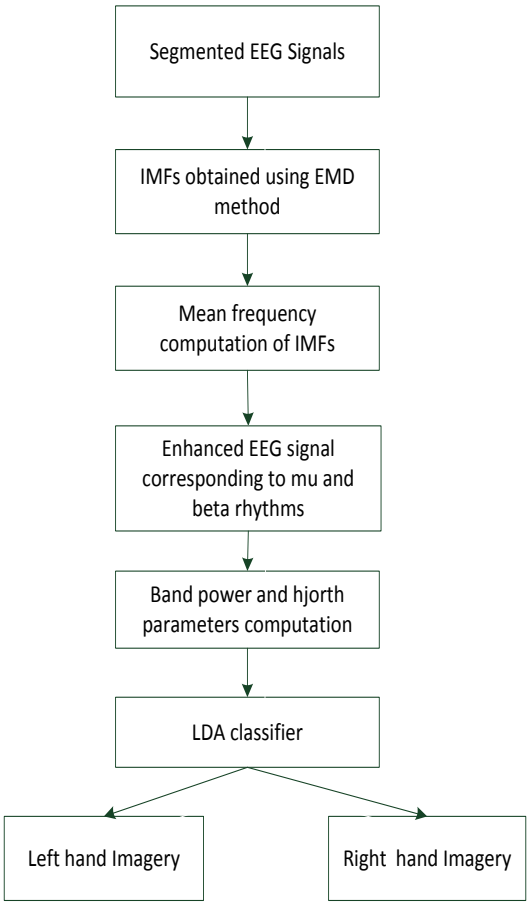


Figure 3.1: Block diagram of the proposed method.

22 EEG channels and 3 monopolar electrooculography (EOG) channels (with left mastoid serving as reference) and includes 288 trials of data (72 for each of the four MI tasks) as shown in Table 3.1. The EEG signals were bandpass filtered between 0.5 Hz and 100 Hz and sampled at the sampling rate of 250 Hz. An additional 50 Hz notch filter has been applied to suppress line noise. Refer to (Brunner et al., 2008) for further details on the BCI competition IV dataset 2A.

BCI competition IV dataset 2B description

This dataset consists of EEG signals performing two different MI tasks: movements

Table 3.1: Datasets description.

BCI competition IV	Dataset 2A EEG	Dataset 2B EEG
Subjects	9 Subjects [A01-A09]	9 Subjects [B01-B09]
Channels	Total channels: 25 EEG channels: 22 (monopolar) EOG channels: 3 (monopolar)	3 bipolar EEG channels C3, Cz, and C4 3 monopolar EOG channels
Recorded sessions	Session - I (288 Trials) Session - II (288 Trials)	Session - I (120 Trials) Session - II (120 Trials) Session - III (160 Trials) Session - IV (160 Trials) Session - V (160 Trials)
Classes	Left hand (class 1) Right hand (class 2) Foot (class 3) Tongue (class 4)	Left hand (class 1) Right hand (class 2)
Sampling frequency	250 Hz	250 Hz
Trial length	7.5 seconds	8 seconds

of the left hand and right hand from nine healthy subjects. The dataset contains five sessions, with the first two sessions containing training data without feedback, and the last three sessions with feedback. Each session was recorded with three bipolar EEG channels and three monopolar EOG channels (with left mastoid serving as a reference) and includes 160 trials of data (80 for each of the two MI tasks) as shown in Table 3.1. The EEG signals were bandpass filtered between 0.5 Hz and 100 Hz and sampled at the sampling rate of 250 Hz. An additional 50 Hz notch filter has been applied to suppress line noise. Refer to (Leeb et al., 2008) for further details on the BCI competition IV dataset 2B.

3.2.3 Empirical Mode Decomposition

The EMD method (cf. 2.10.1) automatically decomposes a signal $y(t)$ into a finite set of IMFs $M_q(t)$, which can be considered band limited and symmetric functions (Huang et al., 1998). Features defined based on the symmetric nature of IMFs have been explored for classification of epileptic seizure related EEG signals (Bajaj and Pachori, 2012; Pachori, 2008; Sharma and Pachori, 2015). For biomedical signals like EEG, it has been shown that better localization of time-varying frequency components of μ and β rhythms during MI can be obtained using the EMD method as compared to the short-time Fourier transform (STFT) and wavelet transform based methods (Davies and James, 2013; Park et al., 2013).

However, the extracted IMFs are narrow-band components of the signal. In order to obtain an enhanced EEG signal corresponding to μ rhythm (8-13 Hz) and β rhythm (14-24 Hz), from the original EEG signal, the selection of an optimal number of IMFs is required. This selection of IMFs is done based on the mean frequency computation from these IMFs so as to obtain an enhanced EEG signal corresponding to μ and β rhythms. Normally, the number of IMFs selected are between two and four depending on the nature of physiological signal. No normalization has been done on the selected IMFs, in fact they were summed after being selected. The mean frequency of each IMF is computed as the sum of a product of IMF spectrum power and the frequency divided by the total sum of IMF power spectrum in the frequency domain (Pachori, 2008; Phinyomark et al., 2012). The mathematical expression of mean frequency is denoted as

$$MF_{IMF} = \frac{\sum_{b=1}^{n} P_b f_b}{\sum_{b=1}^{n} P_b} \tag{3.2.1}$$

where n denotes the length of frequency bin, and P_b gives the power spectrum at the frequency bin b. f_b represents the frequency value at the frequency bin b. These computed mean frequencies represent centroids of the IMF in the spectrum in frequency domain. The enhanced EEG signals are obtained from the summing

of these IMFs whose mean frequencies belong to frequency bands μ and β bands. In order to cover both the bands (μ and β), frequencies in the range of 6-26 Hz have been considered for selection of IMFs of EEG signals. Therefore, left hand and right hand MI EEG signals are decomposed using this EMD method (cf. 2.10.1).

3.2.4 Feature Extraction

BCI feature extraction approaches are many and varied and include techniques such as power spectral density (PSD) (Herman et al., 2008), band power, Hjorth parameters (Hjorth, 1970), and bispectrum (BSP) (Shahid and Prasad, 2011). However, researchers in the BCI community focus mainly on frequency domain features at the signal processing stage. The most commonly used features in BCI applications for classification of left hand, right hand, both feet and tongue MI EEG signals are band power and Hjorth features (Gandhi et al., 2014; Park et al., 2011; Wolpaw et al., 2002). These feature combinations have been collectively extracted from the enhanced EEG signals to classify MI EEG signals. The band power features are computed as the square of the amplitude of the signal over a small time window of 1 second in this study. Typically, the band powers of the two frequency bands associated with the μ and β rhythms are computed for classification of EEG signals corresponding to left hand, right hand, both feet and tongue MI tasks. The frequency ranges (8-12 Hz) and (16-24 Hz) have been selected corresponding to frequency bands μ and β respectively (Bamdadian et al., 2013; Gaur et al., 2015; Shahid and Prasad, 2011; Shahid et al., 2010).

The first Hjorth parameter, we consider is activity, which is the measure of the average power of the signal (variance of the signal). Mathematically, it can be expressed as (Hjorth, 1970):

$$\text{Activity} = \sum_{i=1}^{N_s} \frac{[y(i) - mean]^2}{N_s} \tag{3.2.2}$$

where N_s is the number of samples in the window. The second Hjorth parameter, we consider is mobility, which is an estimate of the mean frequency. It can be

defined as follows (Hjorth, 1970):

$$\text{Mobility} = \sqrt{\frac{\text{var}(y')}{\text{var}(y)}} \tag{3.2.3}$$

The third Hjorth parameter, we consider is complexity, which is an estimate of the bandwidth of the signal. It can be defined as follows (Hjorth, 1970):

$$\text{Complexity} = \sqrt{\frac{\text{Mobility}(y')}{\text{Mobility}(y)}} \tag{3.2.4}$$

where, y is the signal and y' is the first derivative of the signal, and μ is the mean of the signal in the computation sampling window. The aforementioned Hjorth features have been computed from a 1 second window of EEG signals from the three channels namely C3, C4 and/or Cz, respectively. All these three Hjorth parameters are used as input features for LDA.

These bands are selected for computing features as they are more reactive during a cued MI task (Raza et al., 2014) in the form of event related desynchronization (ERD) and event related synchronization (ERS) over the sensorimotor cortex (Pfurtscheller and Neuper, 2001; Pfurtscheller et al., 1997). The combined features based on band powers and Hjorth parameters have also been used as a final feature set for classification of multiple class MI based EEG signals.

3.2.5 Linear Discriminant Analysis

Generally, it is a very tedious task to classify the extracted features for classification of EEG signals in BCI applications. The demanding task is to find the optimum combination of the features which can reduce classification errors and can provide better feature separability. An LDA classifier has been applied which is most commonly implemented in EEG-based BCI applications. The LDA classifier tries to reduce the dimensionality and simultaneously protects most of the class discrimination information. Suppose, we have two classes of data, denoted by cls_1 and cls_2. Then, we classify the n-dimensional sample points $x = \{x_1, x_2, x_3, \ldots, x_n\}$, where m_1 samples belongs to class cls_1, and m_2 samples belongs to class cls_2. The

main goal is to enact a hyperplane $y = w^t x$ from the set of all possible lines. The selected line may maximize the discrimination between two classes. To obtain a good projection vector, the distance between the two classes needs to be measured. The mean vector of each class in x-space and y-space is represented by following equations (Lotte et al., 2007):

$$\upsilon_i = \frac{1}{N_i} \sum_{x \in w_i} x, \tag{3.2.5}$$

$$\text{and } \vartheta_i = \frac{1}{N_i} \sum_{y \in w_i} y = \frac{1}{N_i} \sum_{y \in w_i} w^t x = w^t \upsilon_i \tag{3.2.6}$$

The objective function is expressed as the distance between the two projected means. It can be defined as follows (Gaur et al., 2015; Lotte et al., 2007; Vidaurre et al., 2011):

$$J(w) = |\vartheta_1 - \vartheta_2| = |w^t(\upsilon_1 - \upsilon_2)| \tag{3.2.7}$$

However, the distance measured between these projected means may not always be a good measure as the standard deviation between classes has not been considered. In order to overcome this restriction, an enhancement of LDA has been proposed known as Fisher's LDA classifier. It determines a decision boundary or most likely a hyperplane in the feature space to classify the features in to distinct classes. It finds out the separation boundary between two given distributions in terms of the ratio of two group variances as given below (Lotte et al., 2007; Vidaurre et al., 2011):

$$J(w) = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}} = \frac{w^t(\upsilon_1 - \upsilon_2)^2}{w^t S_1 w + w^t S_2 w} \tag{3.2.8}$$

where υ_1, υ_2 are the mean of the classes and S_1, S_2 are the variances of the feature distributions between two classes w_1, w_2 respectively. The maximum separation between the two classes can be shown by (3.2.9) (Lotte et al., 2007):

$$w^* = \frac{\upsilon_1 - \upsilon_2}{(S_1 + S_2)} \tag{3.2.9}$$

The w^* is weight vector which provides optimum direction of projection of the data. In Fisher's LDA, the decision boundary uses following equation to classify the feature vector x as (Lotte et al., 2007):

$$y = xw^t + b \tag{3.2.10}$$

where b is the bias or threshold. Thus, by choosing a threshold b, it can be used to classify a new feature if $y(x_{new}) \geq b$ or $y(x_{new}) \leq b$ to one of the classes based on the sign of the b.

3.2.6 Results and Discussion

In order to evaluate the performance of the proposed method, the BCI competition IV dataset 2B (Leeb et al., 2008) has been used. The dataset contains EEG signals from nine healthy subjects, denoted by B01-B09, where each contained five sessions. The EEG signals of all nine subjects have been used to study the effectiveness of the proposed method. To evaluate the method, the data was selected from C3 and C4 channels related to the sensorimotor areas. There are a different number of trials in each session, e.g., 60, 70 or 80 as discussed in Table 3.1. Each trial involved a paradigm period of 8 second (Leeb et al., 2008). In the training phase, a single session namely ^03T has been used. For the evaluation phase, we have used two sessions namely, ^04E and ^05E to compute the accuracy of classifying left and right MI EEG signals. These sessions are selected in order to maintain consistency and provide comparison with other filtering techniques (Gandhi et al., 2014). It should be noted that, the ^ in the session name denotes the subject number in the range B01 to B09. We have also used BCI competition IV dataset A (Brunner et al., 2008) for evaluating the performance of the proposed method. The dataset contains EEG recordings from the nine healthy subjects, namely (A01-A09). The EEG signals of all the nine subjects have been considered to evaluate the performance of the proposed method. For each of the nine subjects, the data recorded over two sessions are provided, e.g., A01T and A01E (Brunner et al., 2008). EEG signals were used from only two channels namely C3, and C4 respectively. For each of the subjects,

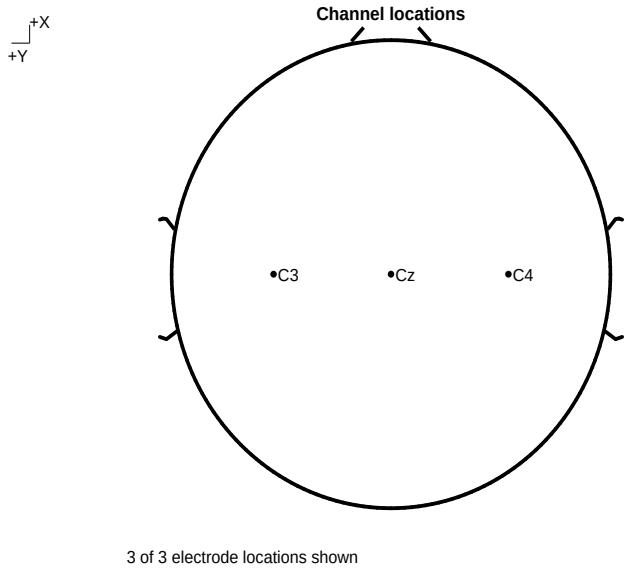


Figure 3.2: Channel locations for channels C3, C4 and Cz.

in order to compute the classification accuracy (in %), the LDA classifier has been trained with 100% data from the B0β03T and tested/evaluated on the 100% data for each of the sessions, B0β04E and B0β05E, where β is the subject number. Since the MI task starts at 3 second, the LDA classifier was trained and tested with the features corresponding to EEG signals from 3 second to 8 second time-interval of the MI paradigm. During the training session, we have performed a 5-fold cross-validation to determine the window size for the best possible classification accuracy by the LDA classifier for classification of the left and right hand MI based EEG signals. The window size denotes the sample points considered from which the features will be extracted and rest of the sample points will be discarded, for example, 1 second window size means 250 sample points will be selected for feature extraction because the sampling frequency is 250 Hz.

To explain the working of the EMD method (cf. 2.10.1), two single trial EEG signals are considered from the dataset B0103T to obtain IMFs (Fig. 3.4 and Fig. 3.5). The left MI EEG signal and its nine IMFs are shown in Fig. 3.4. Similarly, the Fig. 3.5 shows the right hand MI EEG signal and its nine IMFs.

The mean frequency was computed for each of the IMFs of the EEG signals corre-

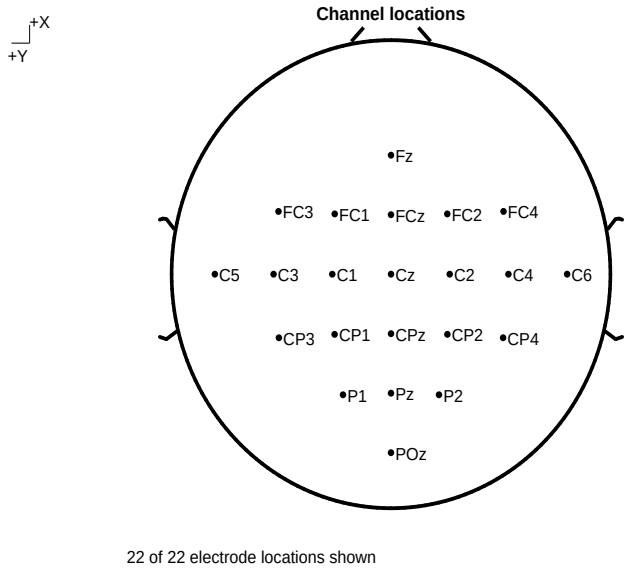


Figure 3.3: Channel locations for all twenty-two channels.

sponding to left and right hand MI tasks. In order to obtain enhanced EEG signals corresponding to left and right hand task, the IMFs whose mean frequencies fall in the range 6-24 Hz were selected. It should be noted that this frequency range covers the μ band (8-13 Hz) and β band (18-24 Hz). These frequency-bands are very important for detection of MI EEG signals (Wolpaw et al., 2000). The features namely, Hjorth parameters and band powers are then computed for the enhanced EEG signals obtained using the selected IMFs. In our study three frequency-bands namely 8-12Hz, 18-22 Hz, and 16-24 Hz were taken for extracting the band powers features and the frequency band 6-24 Hz for the Hjorth features. The extracted features have been given as input features to the LDA classifier for classification of left and right hand MI EEG signals. Table 3.2 shows the maximum classification accuracy for BCI competition IV dataset 2B with EMD based filtering, which provides information on both the enhanced EEG signals and with the raw EEG signals, for the nine subjects denoted by B01-B09 across three sessions 03T, 04E, and 05E. It should be noted that only channels C3 and C4, and Cz are considered for computing the results. Manual channel selection was done in the training session and the same set of channels are used in the evaluation session. The preliminary analysis reveals that only channels C3 and C4 provide better classification accuracy for all

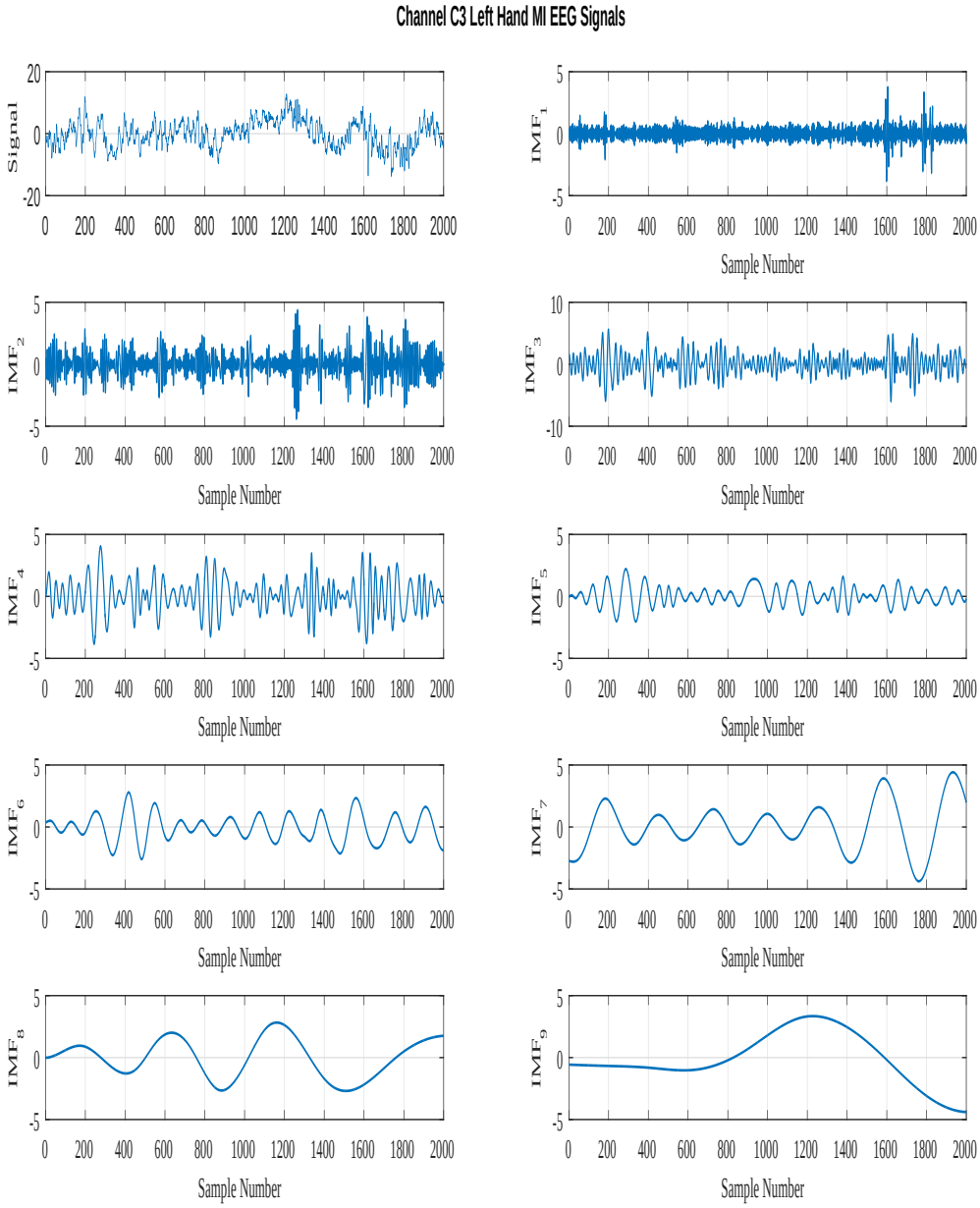


Figure 3.4: The EEG signal corresponding to channel C3 of the trial 10 of B0403T for the left hand movement and its first nine IMFs. Y-axis represents the amplitude in the time domain for all each of the IMFs.

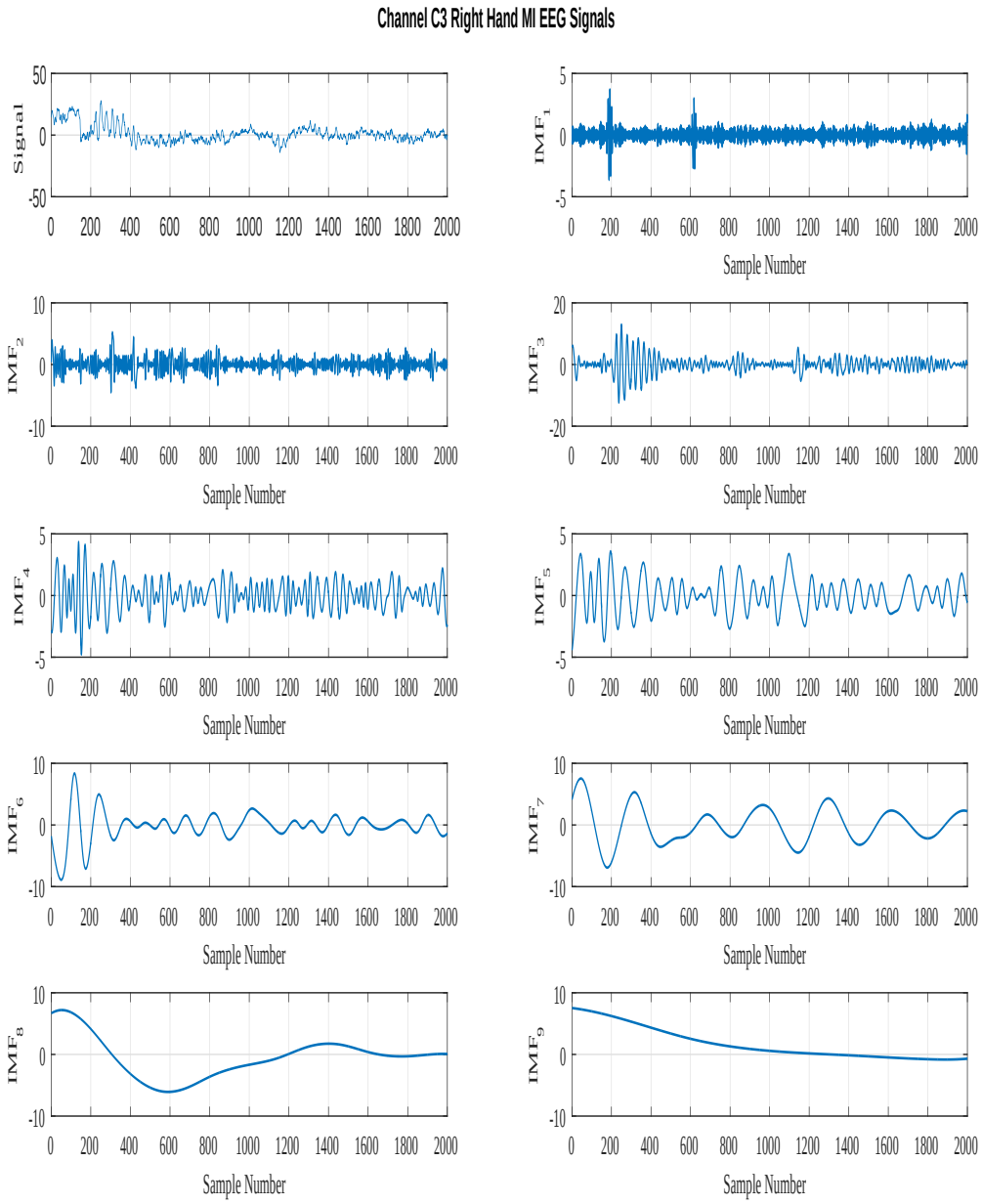


Figure 3.5: The EEG signal corresponding to channel C3 of the trial 10 of B0403T for the right hand movement and its first nine IMFs.

Table 3.2: Maximum classification accuracies of the proposed method based on EMD and without EMD studied on BCI competition IV dataset 2B.

Subject	Accuracy with EMD based filtering(%)			Accuracy with raw EEG(%)			Channel
	Training (03T)	Evaluation (04E)	Evaluation (05E)	Training (03T)	Evaluation (04E)	Evaluation (05E)	
B01	80	70	55.63	75	60	50	C3 C4
B02	77.5	64.17	70	63.13	58.33	55	C3 C4
B03	98.75	98.75	100	55.63	50	55.63	C3 C4
B04	90.63	95.63	81.25	95.63	94.38	78.13	C3 C4
B05	90	98.13	97.5	73.75	80.63	77.5	C3 C4
B06	93.75	89.38	91.88	62.5	64.38	73.75	C3 C4 Cz
B07	80.63	68.75	75.63	68.75	61.25	60.63	C3 C4
B08	82.5	85	90.63	82.5	82.5	89.38	C3 C4
B09	85.63	86.25	84.38	79.38	80.63	75.63	C3 C4
Average	**86.6**	**84**	**82.99**	**72.92**	**72.23**	**68.4**	
Std	**7.13**	**13.27**	**14.2**	**12.12**	**14.66**	**13.42**	

subjects except for the subject B06. Subject B06 provides higher classification accuracy when channel Cz is also selected along with the channels C3 and C4. There is inter-subject variability because of non-stationarity across the subject performing the same MI task. Some of the subjects perform well while other subjects are not good at performing MI task. After applying the EMD based filtering, the group average of the maximum classification accuracy for all subjects across the three sessions improved by 10.54% ($p < 0.001$). In the training session 03T, the results clearly showed the average of the maximum classification accuracy was enhanced by 9.65% ($p < 0.05$) when comparing the EMD based filtering with the raw EEG signals using the same combination of band powers and Hjorth features. In the evaluation sessions, the average improvement with the EMD based filtering is > 9% for 04E and in the case 05E, the average classification accuracy improved by > 12% ($p < 0.01$). Indeed, as shown in Table 3.2, the ten sessions including training and evaluation provided an improvement of > 10% ($p < 0.001$) classification accuracy with EMD based filtering method across subjects.

Table 3.3: Maximum classification accuracies of the proposed method based on EMD and without EMD studied on BCI competition IV dataset 2A.

Subject	Accuracy with EMD based filtering (%)		Accuracy with raw EEG (%)		Channel
	Training	Evaluation	Training	Evaluation	
A01	78.6	66.7	52.9	56.3	C3,C4,Cz
A02	68.6	63.9	56.4	54.9	C3
A03	89.3	77.8	62.9	63.2	C3,C4,Cz
A04	72.1	63.2	48.6	55.6	C3,C4
A05	75	72.2	51.4	50	C3
A06	64.3	70.1	57.1	54.2	C3,Cz
A07	78.6	64.6	51.4	50.7	C3,C4
A08	71.4	76.4	56.4	61.1	C3,Cz
A09	77.9	77.1	66.4	74.3	C4
Average	**75.1**	**70.2**	**55.9**	**57.8**	
Std	**7.20**	**5.91**	**5.74**	**7.51**	
***p*-value**	**0.001**	**0.001**			

Table 3.3 shows the average classification accuracy as well as the maximum of the classification accuracy obtained for BCI competition IV dataset 2A for the classification of left and right hand MI EEG signals. Channels C3, C4 and Cz were selected to compute the classification accuracy to provide a fair comparison with other filtering technique (Gandhi et al., 2014). In order to compute the classification accuracy in the training session, a five-fold cross-validation mechanism has been applied. To compute the classification accuracy in the evaluation session, a model with 100% data from the training session is created, then each of the single trials of evaluation session is assigned a class.

Different channel combinations were selected manually in the training session using a five-fold cross-validation scheme for each subject based on the optimal accuracies as reported in the Table 3.3. The same set of selected channels were used for computing the classification accuracy in the evaluation session. After applying the EMD based filtering, the group average of the maximum classification accuracy across the two sessions was improved by 15.90% ($p < 0.001$). The

average of the maximum classification accuracy for training stage showed highly significant improvement of $>$ 19% (p < 0.001). However, in the evaluation session, it showed an improvement of $>$ 12% (p < 0.001) in the classification accuracy when compared with the raw EEG signals using the same set of band powers and Hjorth features. The six sessions including training and evaluation showed highly significant improvement of > 20% (p < 0.001), and eight sessions showed improvement in the range of 10% (p < 0.001) to 20% (p < 0.001) across all the nine subjects. The p-value across session-wise for all the nine subjects has been calculated using the repeated measures analysis of variance using the ranova command in MATLAB which calculates a repeated measures analysis of variance.

Fig. 3.6 and Fig. 3.7 show the difference between EMD based filtering and without EMD based filtering results. The performance improvement across all the nine subjects are illustrated with the bar graphs. In Fig. 3.6, the performance improvement for the BCI dataset 2B have been shown for the training and the evaluation sessions respectively. The bar graphs show an improvement of accuracy when the EMD based filtering method is applied when compared to raw EEG signals across 3 sessions (i.e., training 03T, evaluation 04E and evaluation 05E) and for 9 subjects.

With the proposed EMD based filtering, seven out of nine subjects have shown improvement in the classification accuracy as shown in the fig. 3.6 for the training session. In the evaluation session 04E, there is an improvement in classification accuracy for all the nine subjects using EMD based filtering. In the evaluation session 05E, eight subjects have shown improvement in the classification accuracy using the EMD based filtering.

For the BCI dataset 2A, the classification accuracy has been reported for the training and the evaluation session, respectively as shown in Fig. 3.7. The bar graphs show an improvement of accuracy when the EMD based filtering method is applied when compared to raw EEG signals across 2 sessions (i.e., training T, evaluation E). Classification accuracy have been significantly improved across the nine subjects

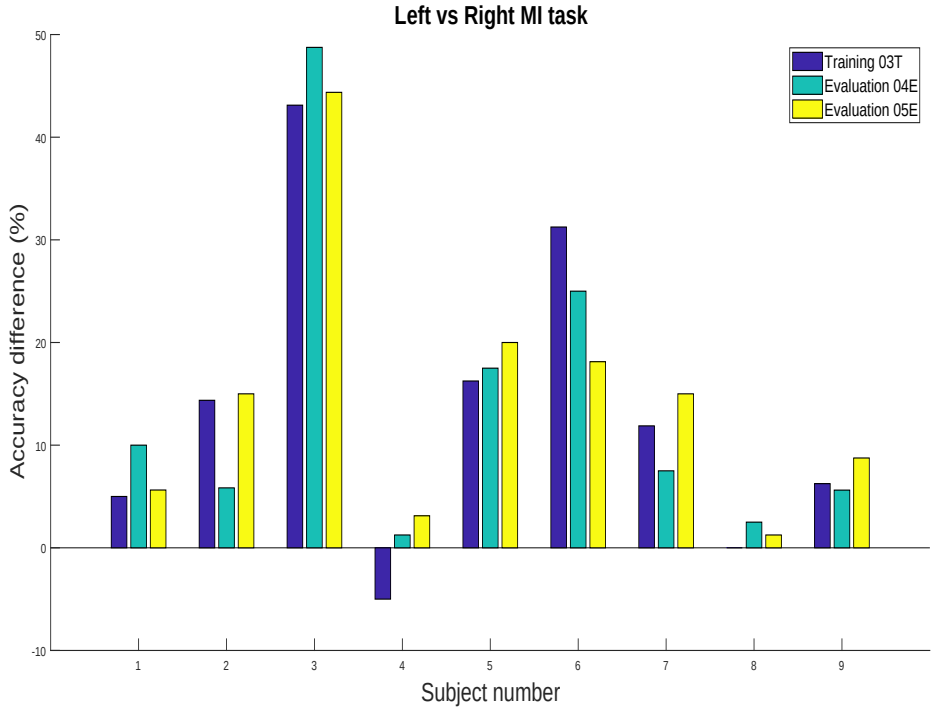


Figure 3.6: The bar graphs show accuracy difference between EMD based filtering and without EMD based filtering for 9 subjects in BCI competition IV dataset 2B.

using the proposed EMD based filtering. In both, the training session and the evaluation session, results obtained from all the nine subjects have shown performance improvement > 7% ($p < 0.001$) in the classification accuracy.

The proposed EMD based filtering has shown a performance improvement in MI based BCI when compared to quantum neural network filtering (Gandhi et al., 2014) as reported in Table 3.4. The EMD based filtering done at the preprocessing stage has thus helped to achieve an average classification accuracy of 70.22 % (p=0.0391) whilst quantum neural network filtering (Gandhi et al., 2014) reported an average classification accuracy of 66.59%. Seven of the nine subjects have shown improvement in the classification accuracy of the evaluation session. The same set of features and classification method has been done to provide a fair comparison. The p-value has been computed with Wilcoxon signed rank test method.

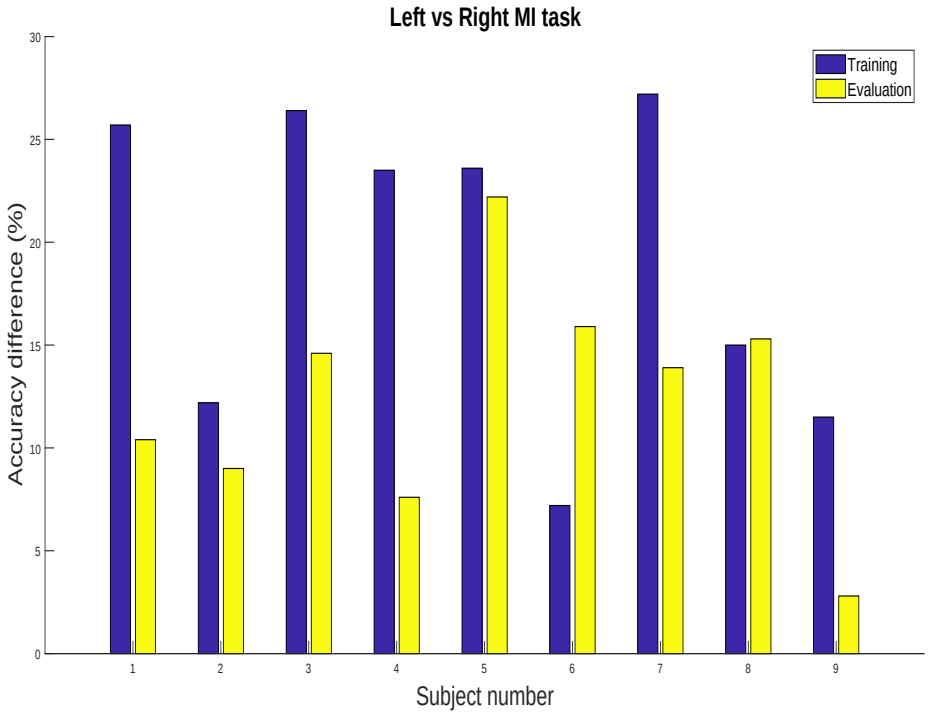


Figure 3.7: The bar graph shows accuracy difference between EMD based filtering and without EMD based filtering for 9 subjects in BCI competition IV dataset 2A.

Table 3.4: Comparison of classification accuracy with other filtering technique in the evaluation session of BCI competition IV dataset 2A.

Subject	EMD based filtering	Quantum neural network filtering Gandhi et al. (2014)
A01	66.7	61.11
A02	63.9	61.11
A03	77.8	79.17
A04	63.2	60.42
A05	72.2	71.53
A06	70.1	61.11
A07	64.6	58.33
A08	76.4	67.36
A09	77.1	79.17
Average	70.2	66.59
$p-$value		0.0391

3.3 Study 2: EMD 4-class MEG

Having presented evidence of the effectiveness of EMD when applied to EEG data, the same technique is now applied to MEG data for the classification of multi-direction wrist movements for enhancing BCI and also addresses C1 of this thesis.

3.3.1 Methods

Similar to the previous EEG study, and as mentioned previously (cf 2.10.1) the EMD method breaks magnetoencephalography (MEG) signals into a set of IMFs (Huang et al., 1998) which can be considered narrow-band, AFM signals. The maximum amplitude frequency measure of these IMFs has been used to combine these IMFs in order to obtain enhanced MEG signals which have major contributions from low frequency band (<8 Hz) (Waldert et al., 2008b). The maximum amplitude frequency is defined as the frequency component in power spectrum where amplitude value is maximum in time domain. The BCI competition IV dataset 3 contains MEG signals for four classes, namely, right, forward, left and backward wrist movements. The signals from 10 channels above the motor areas have been used for the study. Significantly improved performance is obtained when the method is tested on this dataset, which demonstrates the effectiveness of the proposed method not only on EEG data but also for MEG data.

3.3.2 Dataset

This dataset contains MEG signals when two healthy subjects performed four wrist movements in four different directions: right, forward, left and backward. The dataset contains two sessions where the first session contains training data, and the other session is used for evaluation. Each session was recorded with ten MEG channels located above the motor areas as shown in Fig. 3.10. Each training

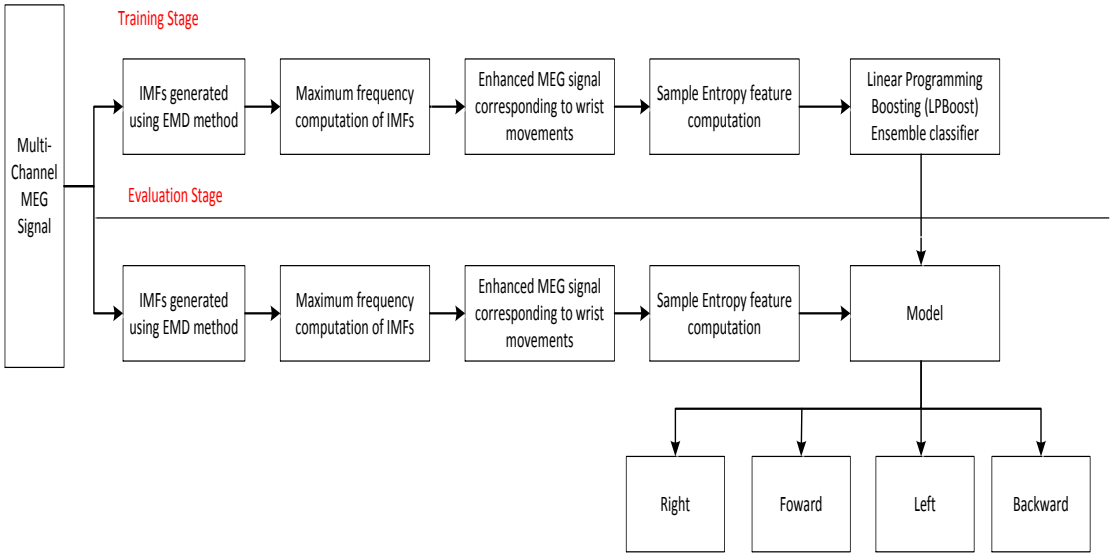


Figure 3.8: Block diagram of the proposed method.

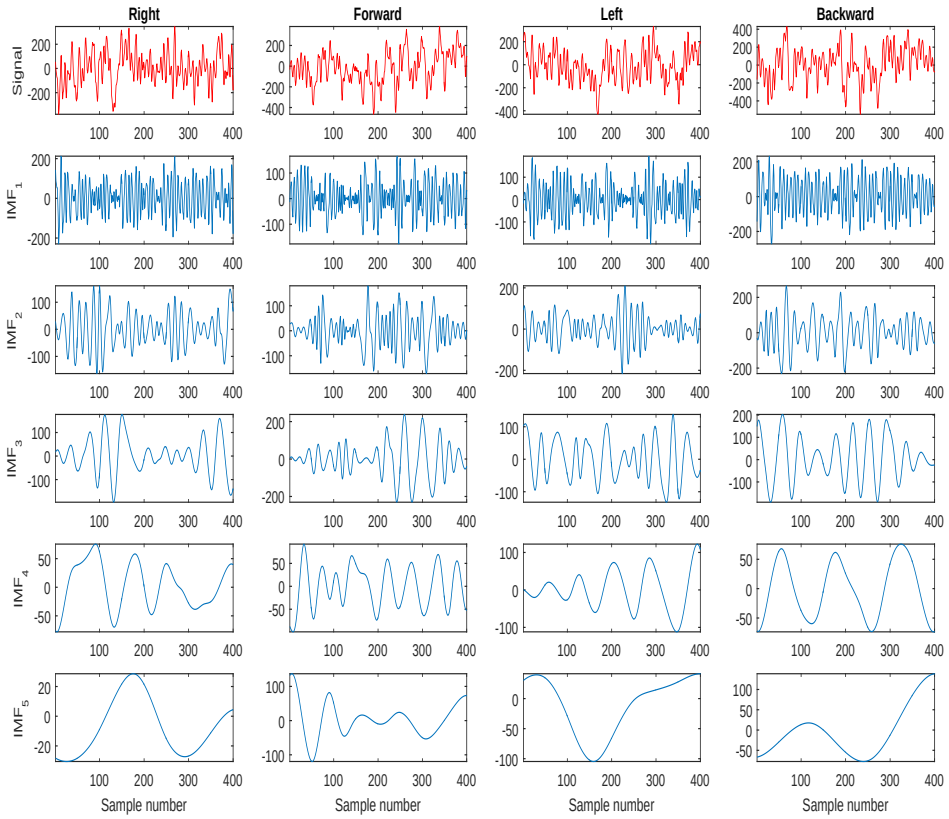


Figure 3.9: The EEG signal corresponding to channel C3 of the trial 10 of B0403T for the right hand movement and its first nine IMFs.

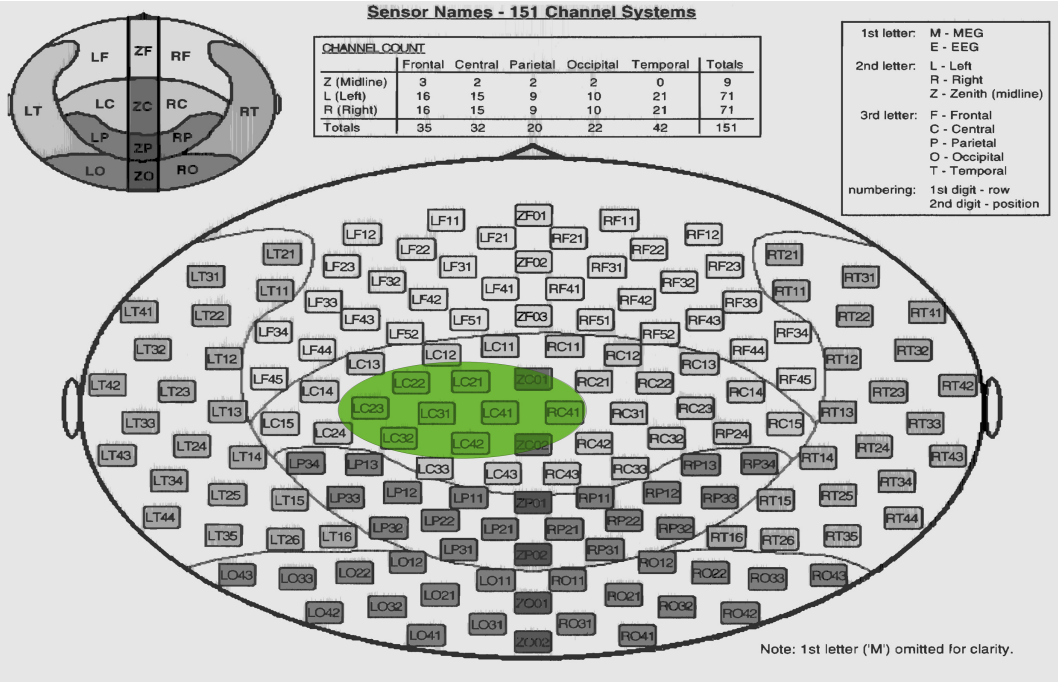


Figure 3.10: MLC21, MLC22, MLC23, MLC32, MLC31, MLC41, MLC42, MZC01, MZC02, and MRC41 channels are used for the present work. These channels are highlighted in green colour.

session includes 160 trials of training data (40 for each of the four wrist movements) as shown in Table 3.5. The MEG signals were bandpass filtered between 0.5 Hz and 100 Hz and re-sampled to 400 Hz from 625 Hz (Waldert et al., 2008a). In the evaluation session, subject 1 contains 74 trials and subject 2 contains 73 trials as shown in Table 3.5. For more details on the BCI competition IV dataset 3 please refer to Waldert et al. (2008a).

3.3.3 Feature Extraction

Sample Entropy (SapEn) (Richman and Moorman, 2000) is a modified version of the Approximate Entropy (ApEn) and is used as a complexity measure of time series. It prevents the bias caused by the use of the self matches in the computation of ApEn and improves performance. Furthermore, SapEn is independent of the long record length and improves the relative consistency (Richman and Moorman, 2000). Moreover, the SapEn algorithm is simpler than the ApEn algorithm and the

Table 3.5: Dataset description.

BCI competition IV	Dataset 3 MEG
Subjects	2 Subjects [S01-S02]
Channels	Ten channels
Recorded sessions	Subject 1 Session 1 (160 Trials) Subject 1 Session 2 (74 Trials) Subject 2 Session 1 (160 Trials) Subject 2 Session 2 (73 Trials)
Classes	Right (class 1) Forward (class 2) Left (class 3) Backward (class 4)
Sampling frequency	400 Hz
Trial length	1 sec

computation time of SapEn is nearly half of the ApEn (Richman and Moorman, 2000). Sample entropy has shown promising results in classification of focal and non-focal seizure EEG classification problems (Sharma et al., 2015a,b).

The SapEn was computed as a feature in order to classify MEG signals for four classes, namely, right, forward, left and backward. The SapEn of the signal is defined as the negative natural logarithm of the conditional probability that two sequences similar for m points remain similar at the next point, where self matches are not included in calculating the probability. Thus, a lower value of SapEn of the signal also indicates more self-similarity in the time-series. In order to calculate the SapEn of the signal, tolerance parameter (r) and embedding dimension (m) must be specified (Sharma et al., 2015b; Song et al., 2010).

Formally, the algorithm for computing SapEn of the signal $x[n]$ consists of the following steps (Wang et al., 2012):

(1) Consider a signal (EEG signal or IMF) $y[n]$ of length N, this signal can be represented by the sequence as, $\{y[1], y[2], ..., y[N]\}$.

(2) Form m dimension vectors consecutively, starting with the i-th point of the signal sequence in the step (1),

$$Y_m[i] = [y[i], y[i+1], ..., y[i+m-1]], i = 1, 2, ..., N-m+1, \tag{3.3.11}$$

(3) Define the distance $d(Y_m[i], Y_m[j])$ between two vectors $Y_m[i]$ and $Y_m[j]$ as the absolute maximum difference between their scalar components:

$$d(Y_m[i], Y_m[j]) = \max_{k=0,1,...,m-1}(|y[i+k] - y[j+k]|), i \neq j \tag{3.3.12}$$

(4) For a given tolerance parameter r, for every i-th value, compute the distance $d(Y_m[i], Y_m[j])$ from (3.3.12). Count the number of distances which are less than or equal to r, denoted as PQ_i. Then compute the ratio of this number to

$N - m - 1$, denoted as $PQ_i^m(r)$,

$$PQ_i^m(r) = \frac{1}{N-m-1} PQ_i \tag{3.3.13}$$

then,

$$PQ^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} PQ_i^m(r) \tag{3.3.14}$$

(5) Increase the dimension m to $m+1$, and then repeat the steps (1) to (4), and compute $PQ_i^{m+1}(r)$.

(6) The SapEn of the signal can be expressed as by (Sharma et al., 2015b):

$$SampEn(N, m, r) = -ln \frac{PQ^{m+1}(r)}{PQ^m(r)} \tag{3.3.15}$$

3.3.4 Results and Discussion

During the training phase, a single session namely s1 has been used to compute a training model. Session s2 has been used for the evaluating the model In the evaluation phase, one session s2 for computing the classification accuracy to classify the MEG signals into right, forward, left and backward wrist movements in both subjects. Each movement has 40 trials each, giving a total of 160 trials in the training session for each subject and evaluation session has 74 trials for subject S01 and 73 trials for subject S02. The maximum amplitude frequency measure is computed for each of the IMFs of the MEG signals corresponding to right, forward, left and backward wrist movements. In order to obtain enhanced MEG signals corresponding to multi-directional wrist movements, the IMFs whose maximum frequencies fall in the range 0.1 - 8 Hz are selected. However, this frequency range covers low bandpass filtering range (< 8 Hz) (Waldert et al., 2008b), although the frequency range from 58 to 70 Hz (high gamma band) was also investigated but the classification accuracy obtained was low as compared to the frequency range 0.1-8 Hz. These frequency bands are crucial for identification of multi-direction wrist movement based MEG signals but only frequency range 0.1-8 Hz was considered for this study.

Table 3.6: Classification accuracies of the proposed method based on EMD and without EMD studied on BCI competition IV dataset 3.

Subject	Proposed method	Winner1	Winner2	Winner3	Winner 4
S01	40.54	**59.5**	31.1	16.2	23.0
S02	**39.72**	34.3	19.2	31.5	17.8
Average	40.13	46.90	25.15	23.85	20.4

Table 3.6 shows the classification accuracy for BCI competition IV dataset 3 with EMD based filtering providing the enhanced MEG signals. It also shows the comparison with BCI competition winners (Hajipour Sardouie and Shamsollahi, 2012) for two subjects. All ten channels as shown in Figure 3.10 are considered for calculating the results. Table 3.6 displays the classification accuracy calculated with the proposed methodology and compared with other similar works. The proposed EMD based filtering has shown a performance improvement in multi-directional wrist movements BCI classification when compared to results reported by others except the BCI competition winner. The EMD based filtering done at the preprocessing stage has thus helped to achieve an average of two subjects classification accuracy of 40.13%. The results obtained have been calculated with a one versus rest mechanism to classify the multi-direction wrist movements MEG signals into multiple classes. With the results, we are placed at the second position in the table.

The top competition winner extracted a combination of statistical features, frequency domain features and wavelet coefficients as a feature set. They also generated two artificial bipolar channels giving a total of twelve channels. Furthermore, they performed feature reduction using a genetic algorithm and then, classification with a combination of LDA and SVM using a linear kernel. The average classification accuracy obtained across the two subjects was 46.90% with a standard deviation of 17.81 (Hajipour Sardouie and Shamsollahi, 2012). The runner-up applied a low-pass filter between 0.5-8 Hz. Then segmented the time segment and extracted the time feature from the segments. They then took the first three

components of the abs and first five PCA components of the angle of 128 FFT for each channel and corresponding samples. Further, they classified the feature set with an LDA classifier after performing the feature reduction step. They have reported classification accuracy of 25.15%. The third group computed the feature set which consists of wavelet coefficient, temporal and statistical parameters. They thereafter applied a genetic algorithm and PCA for feature reduction respectively and used linear SVM for feature classification and reported 23.9% as classification accuracy. The fourth group applied low-pass filtering between 0.5 and 8Hz and then segmented the time-series into small segments of 0.5 seconds. They extracted the first five PCA components of the angle and the first three angles of 128 FFT for each segment and each channel. They classified the feature set using LDA classifier and reported a classification accuracy of 20.4% for two subjects.

3.4 Conclusion

In this chapter the first contribution of this thesis has been met (C1) by exploring a novel single channel empirical mode decomposition (EMD) based filtering method for enhancing the performance of a motor imagery based BCI. In the first 2-class EEG-based study, a combination of IMFs whose mean frequencies fall in the frequency range of μ and β rhythms has provided a significant improvement (> 10%, 2-class) in accuracy when classifying left and right hand MI signals when compared to those using raw EEG. The second 4-class MEG-based study has also met the first thesis contribution (C1) by exploring single channel EMD based filtering method for enhancing performance when classifying wrist movements in BCI. The proposed method identifies a combination of IMFs whose maximum frequency falls in the low frequency band (<8 Hz) and provided comparable results using sample entropy features to classify multi directional wrist movement signals when compared to the BCI competition winners (40.13%, 4-class). These two studies together demonstrate the effectiveness of the EMD method when applied to both EEG and MEG data for enhancing the performance of a BCI.

The next chapter seeks to address the second (C2) and third (C3) contributions of the thesis by investigating the effectiveness of an extension to the EMD method known as multivariate empirical mode decomposition (MEMD) based filtering (cf. 2.10.2) when applied to multichannel EEG data in a motor imagery based BCI paradigm.

Chapter 4

Multivariate Empirical Mode Decomposition based Filtering

4.1 Introduction

The previous chapter presented two studies which together demonstrated the effectiveness of the novel single channel empirical mode decomposition (EMD) based filtering method when applied to both electroencephalogram (EEG) and magnetoencephalography (MEG) data for enhancing the performance of a brain-computer interface (BCI). This chapter seeks to address the second (C2) and third (C3) contributions of this thesis by investigating the effectiveness of an extension to the EMD method known as multivariate empirical mode decomposition (MEMD)(cf. 2.10.2) and proposing a novel multichannel EMD based filtering when applied to multichannel EEG data in a motor imagery based BCI paradigm. This chapter again describes two separate studies both of which explore the application of this MEMD based filtering method along with common spatial pattern (CSP) features for enhancing the performance of a two-class motor imagery (MI) based BCI. The first study studies MEMD based filtering with CSP features where a separate training model is created for each subject. The second study extends the multi-channel MEMD based filtering method put forward in the first study,

by presenting a novel multi-channel MEMD filtering based subject independent (MEMD-SI) BCI by training the system on MI EEG data from all subjects.

4.2 Study 1 : MEMD based filtering 2-class EEG MI

4.2.1 Methods

As discussed previously (cf. 2.10.1) and demonstrated in the previous chapter the EMD method is highly suitable for analysis of non-stationary signals. Whereas single channel EMD suffers from the problem of mode-mixing, wherein similar frequencies occur in different IMFs (Park et al., 2013), the multi-channel version, namely, multivariate EMD (MEMD), seeks to address this problem (Park et al., 2013)(cf. 2.10.1).

In this study, a new MEMD based bandpass filtering (MEMDBF) is proposed by selecting the multivariate intrinsic mode functions (MIMFs) which contributes to μ and β rhythms observed over the central region of the brain when the subjects plan or execute hand movements. A block diagram representation of the proposed MEMD based bandpass filtering methodology is shown in Figure 4.1. Further, the candidate MIMFs are selected based on the mean frequency measure calculation corresponding to μ and β rhythms. The enhanced MIMF is obtained by summation of all the candidate MIMFs. The features extracted from the enhanced MIMFs are used for classification of left hand and right hand MI tasks as compared with our previously proposed filtering technique (Gaur et al., 2015) based on EMD which is restricted to the decomposition of EEG signals on one channel at a time in BCI. This MEMDBF filtering is done ahead of any feature extraction and classification steps. Its goal is to provide better feature separability, leading to reduced error rates and high task classification accuracy in an MI based BCI. Spatial filters are applied which maximize the variance in one class and minimize it in the other class, additionally, CSP features are computed (cf. 4.2.4) using the first and last three

pairs of spatial filters from the enhanced EEG signals and a linear discriminant analysis (LDA) classifier (cf. 3.2.5) is used to classify the feature sets into left and right hand MIs.

4.2.2 Dataset

The MEMD based filtering technique has been investigated on the BCI competition IV dataset 2A (Brunner et al., 2008) which has already been described in the first study of the previous chapter (cf. 3.2.2).The time interval selection for MI classification is a key factor that helps to reduce the error rates. In this work, CSP features are extracted from the enhanced EEG signals from fifteen channels as shown in Figure 4.2. Features are extracted from EEG signals between 0.5 sec and 2.5 sec after onset of the visual cue in the training step, keeping it same as the competition winner (Ang et al., 2012). In this work, the number of channels selected were increased from three channels (Gandhi et al., 2014) to fifteen channels (i.e., FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2 and CP4) over the motor cortex region for the analysis because the most activity is seen for left hand and right hand motor imagery over the ipsilateral sensorimotor cortex and contralateral sensorimotor cortex (Gandhi et al., 2014; Herman et al., 2008; Wolpaw et al., 2000). Refer to (Brunner et al., 2008) for further details on the BCI competition IV dataset 2A.

4.2.3 Multivariate Empirical Mode Decomposition (MEMD)

The left hand and right hand MI EEG signals are decomposed using MEMD method. A brief description of MEMD method is provided in Section 2.10.2 of Chapter 2. The IMFs obtained with the MEMD method of left hand and right hand MI EEG signals are shown in Figure 4.3 and Figure 4.4 respectively.

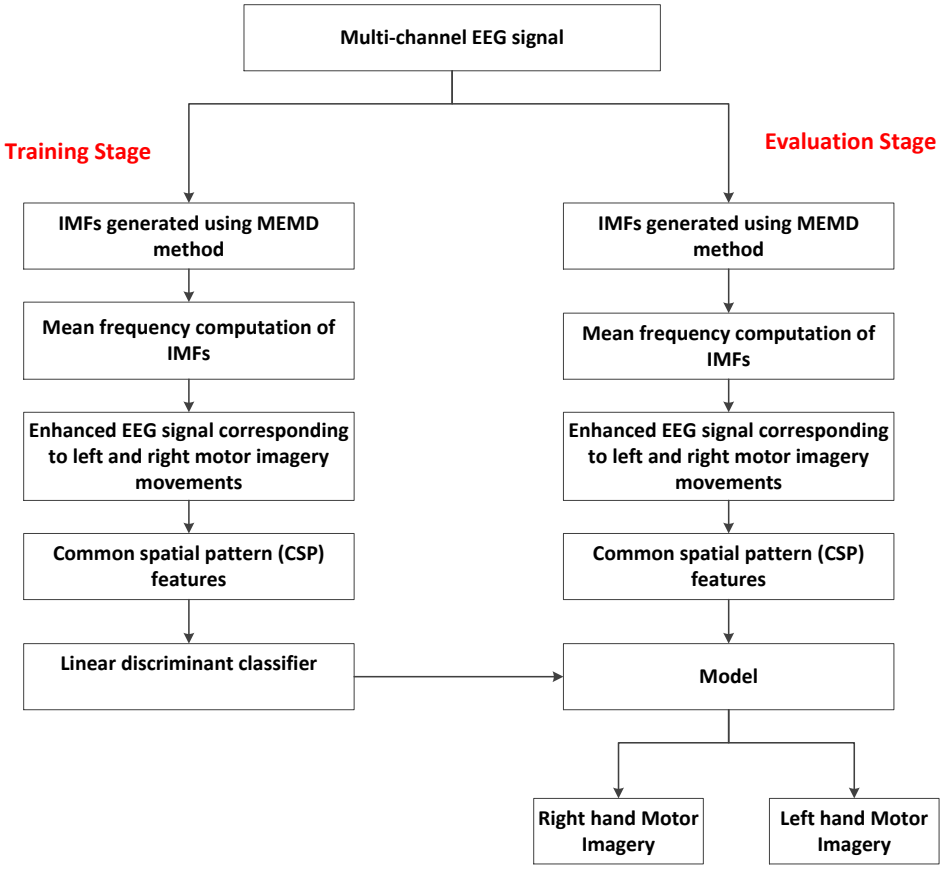


Figure 4.1: Block diagram of the proposed methodology.

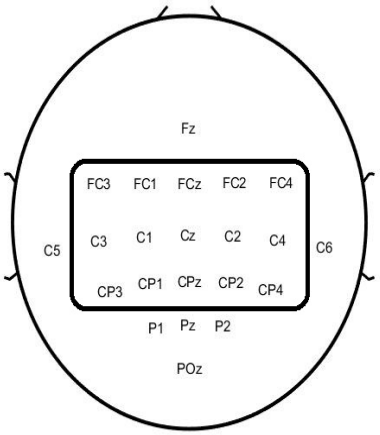


Figure 4.2: Channels used for the present study.

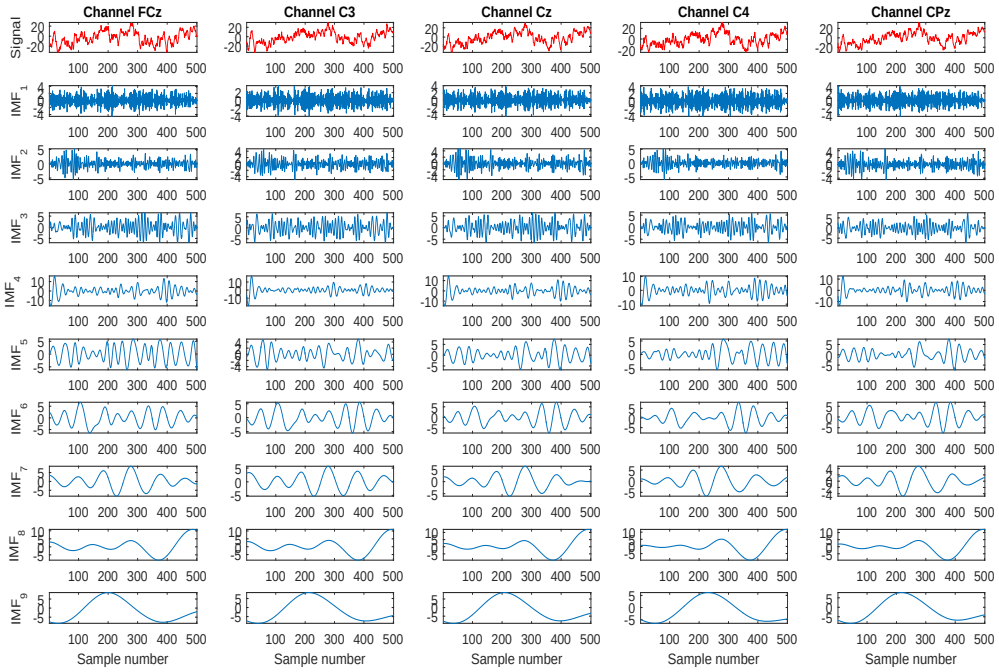


Figure 4.3: The EEG signal corresponding to channel C3 of the trial 1 of A08T for the left hand movement and its first nine IMFs.

4.2.4 Common Spatial Pattern (CSP)

In feature extraction stage, a widely used feature corresponding to MI based BCI has been used which uses the CSP algorithm from fifteen channels as shown in Figure 4.2. The CSP algorithm may be understood as a method which generates weight maps of the selected channels for EEG signals. The weight maps provide the importance of EEG signal content of the channels to separate the two conditions present in the data (Blankertz et al., 2008).

The weight maps are spatial filters which are then projected onto data. With the projection of these spatial filters, the data is altered in a way that the ratio of the variance for EEG amplitudes between the provided two conditions is maximized. Therefore, the variance of the filtered EEG signal may serve as a discriminative feature for a two-class classification problem. The recorded EEG scalp potentials may have very poor spatial resolution because of volume conduction. With the

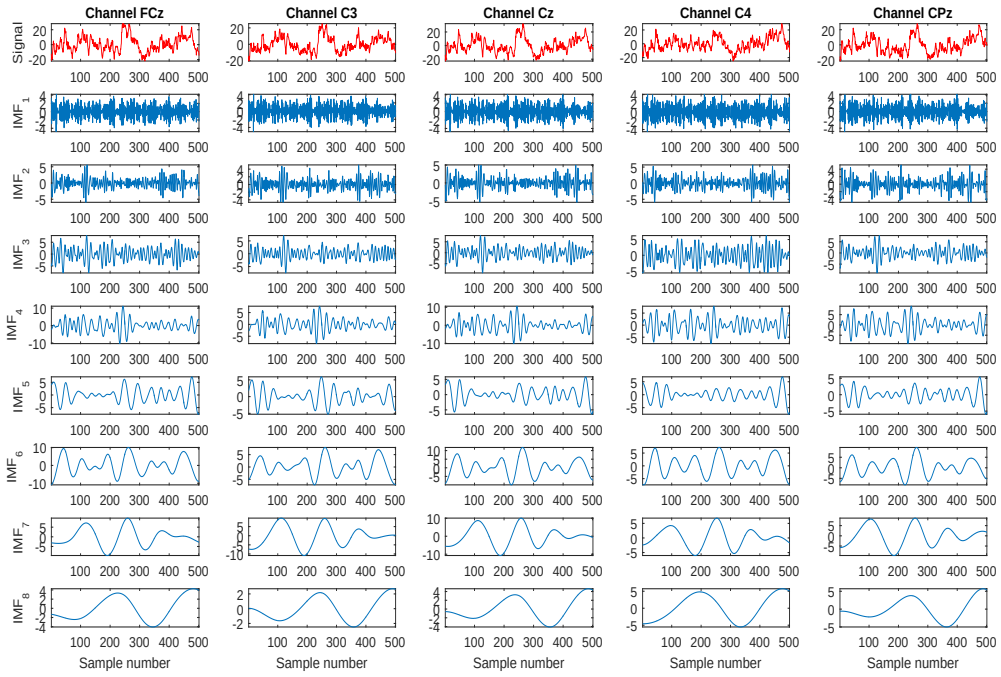


Figure 4.4: The EEG signal corresponding to channel C3 of the trial 2 of A08T for the right hand movement and its first nine IMFs.

poor spatial resolution, the classification of EEG signals becomes more tough especially if other sources produce more strong signals and the signal of interest is weak in the specific frequency range (Blankertz et al., 2008).

As mentioned in the introduction(cf. 2.1), the CSP algorithm has shown promising results in computing spatial filters for detecting event related desyncronization / event related synchronization (ERD/ERS) (Ang et al., 2012; Blankertz et al., 2008). It is a trial specific supervised decomposition of signals which is parameterized by a projection matrix $PM \in \Re^{Chn \times Chn}$ where Chn denotes the number of channels selected. In EEG signal sensor space, PM gives the projection of a single trial $Tr \in \Re^{Chn \times t}$ to $C \in \Re^{Chn \times t}$ in the surrogate sensor space, which is represented as:

$$C = (PM)^T \times Tr \tag{4.2.1}$$

where C gives $Chn \times t$ EEG measurement data selected from a single trial, and t provides the number of sample points per channel. The spatial filters are denoted

by the rows of PM. The spatially filtered signal C provided in (4.2.1), maximizes the ratio of the variances of the two classes. A CSP analysis is employed to obtain an efficient discrimination between two different conditions which are described by ERD/S mechanisms. However, the variances concerning to a small subset of spatial filters are usually selected. The first M and last M rows of C i.e., C_e , $e \in \{1, 2, ..., 2M\}$ given in (4.2.1) are used. In this study, we have considered $M = 5$ spatial filters. Please refer to (Blankertz et al., 2008) for more details.

4.2.5 Classification

In this study, an LDA classifier is applied (discussed earlier in section cf. 3.2.5) which is most commonly used for EEG signals in BCI applications. The LDA classifier tries to reduce the dimensionality and simultaneously protects most of the class discrimination information. Also, the effectiveness of the CSP features to classify the left hand and right hand MI EEG signals is evaluated using the LDA classifier (Pfurtscheller et al., 1997).

4.2.6 Study1: Results and Discussion

For the computation of classification accuracy (in %) for each subject, 100% of A0ST data has been considered for training the classifier model using an LDA classifier. Then, it is evaluated on 100% data A0SE for each evaluation session, where S represents the subject number. In the MI paradigm, the MI task begins at 2 second; the training session and evaluation session features have been extracted from the 2.5 to 4.5 seconds time interval similar to the competition winner (Ang et al., 2012).

During the training session, a five-fold cross-validation has been applied to classify the EEG signals into LHMI and RHMI tasks. To demonstrate the decomposition dynamics of the MEMD technique, single trial EEG signals per class are considered from the subject A08's training session data A08T. Figure 4.3 demonstrates the

LHMI tasks raw EEG signals and its obtained IMFs whilst Figure 4.4 gives the RHMI EEG signal and its obtained IMFs.

The statistical mean frequency measure has been computed for each of the IMFs obtained pertaining to LHMI and RHMI tasks. To get enhanced EEG signals pertaining to these MI tasks, the IMFs are first identified based on mean frequencies which lie in the frequency range $4 - 24$ Hz (Gaur et al., 2015). This frequency range comprises the mu (μ) band and beta (β) band. These bands play a critical role in the identification of MI EEG signals (Gandhi et al., 2014, 2015; Gaur et al., 2015).

Figure 4.5 and Figure 4.6 display the feature distribution of four features. The box plot in Figure 4.5 represents the four features using the Kruskal-Wallis test with the MEMDBF method. To show the effect of the proposed filtering method on the available CSP features, feature separability is evaluated using the Wilcoxon test method for the LHMI and RHMI tasks. The CSP features are arranged in decreasing order of class separability based on the p-value computed using Wilcoxin test method. The CSP features extracted from the raw EEG signals were not significant and after applying the proposed filtering these features are significant as discussed later. Although, this procedure was done for all the fifteen channels but for illustration, only four features were shown in the Figure 4.6. The proposed preprocessing method has thus helped achieve statistically **significant improvement in feature separability (p < 0.005)** in the training session for the LHMI and RHMI tasks. Figure 4.6 displays the same four features from the raw EEG signals giving p-values of 0.0522, 0.9109, 0.1136, and 0.0475. The p-values reveal the fact that the three features (with p-value 0.0522, 0.9109, and 0.1136) are not significantly different in their feature distribution for the LHMI and RHMI tasks. However, with the proposed pipeline, the p-values show a statistically significant difference in feature distribution for all four features. Although all the subjects were considered for the study but subject A01 was used for illustration purpose. The non-parametric Wilcoxon test is used for ranking the four features.

Table 4.1 shows the classification accuracy for the BCI competition IV dataset 2A

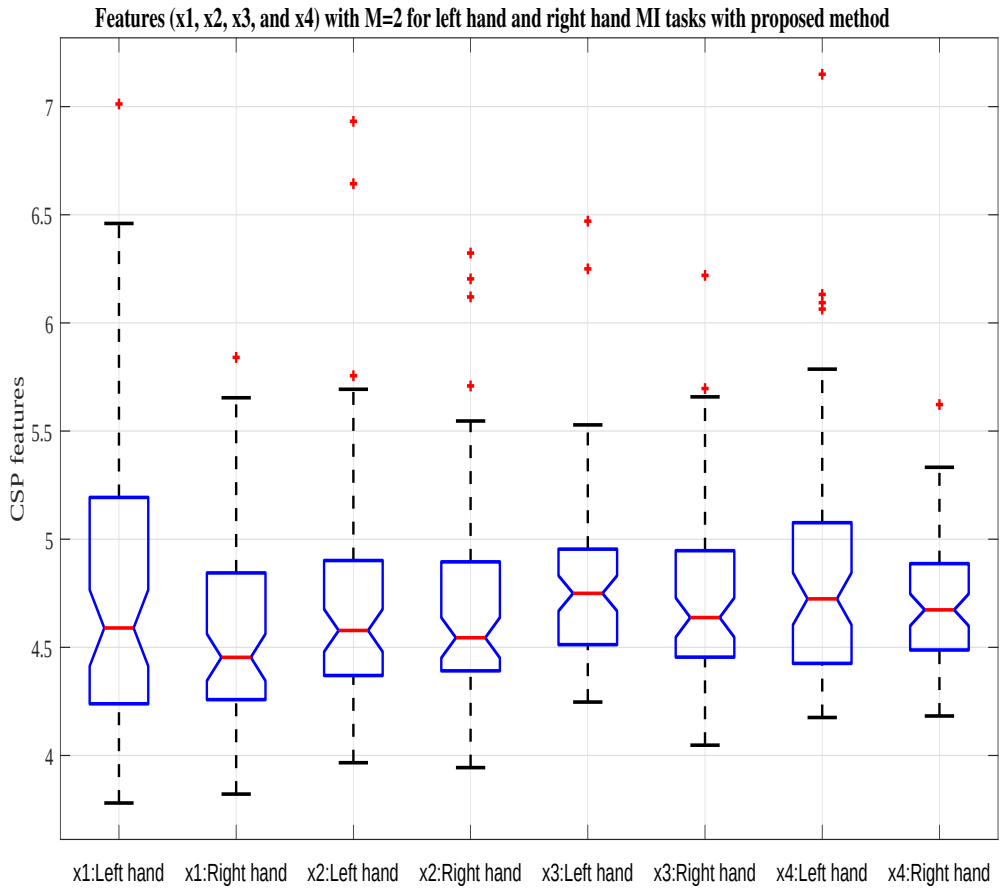


Figure 4.5: The box plot displays the calculated four features using MEMDBF in the training session for left hand and right hand MI tasks are statistically significant features in terms of separability with p-values < 0.005.

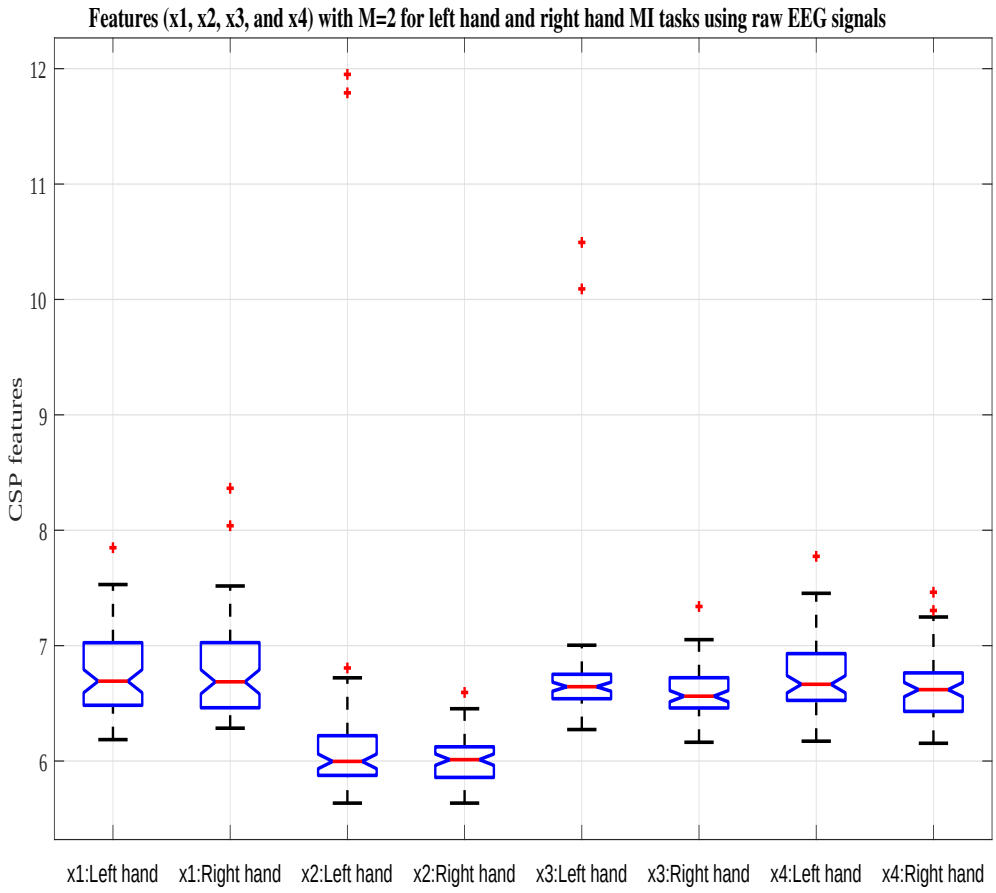


Figure 4.6: The box plot reveals that the same four features from the raw EEG signals are not statistically significant in terms of separability.

obtained using LDA classifier with raw EEG signal and the preprocessed EEG signals with the MEMDBF method. This method provides enhanced EEG signals for all A01-A09 subjects as compared to raw EEG signals, across two sessions namely, the training and evaluation sessions. With the enhanced EEG signals using the MEMDBF method, the group average of **classification accuracy improved by 6.37%** across all subjects considering both training and evaluation sessions. The results computed in the training session clearly depict that the average of the **classification accuracy improved by 2.84%** (p **= 0.3594**) with standard deviation of 12.98 with MEMDBF-CSP method compared with the raw EEG signals considering the same features. Notably, **eight of the nine subjects have improved** in classification accuracy in the evaluation session and also the group average of **classification accuracy across all nine subjects has improved by** $>$ **9.91% (**p **= 0.0078).**

Table 4.1: Classification accuracies (in %) obtained with the MEMDBF method and raw EEG signals by LDA classifier evaluated on BCI competition IV dataset 2A.

Subject	Accuracy with MEMDBF-CSP			Accuracy with raw EEG	
	Training	Evaluation1	Evaluation2	Training	Evaluation1
A01	91.67	90.78	90.81	72.27	69.44
A02	59.05	58.45	61.33	63.21	50
A03	93.27	93.43	95.69	91.65	90.28
A04	73.76	74.14	74.2	71.58	59.03
A05	68.05	60.74	64.44	67.92	50
A06	76.34	68.52	68.44	67.99	54.86
A07	79.57	80	76.52	86.18	65.28
A08	94.43	97.01	96.27	95.19	97.92
A09	94.29	94.62	94.62	88.86	91.67
Average	81.16	79.74	80.26	78.32	69.83
p**-value**	0.3594	0.0078			

With the MEMDBF method, the difference between accuracies obtained in the training session and evaluation session have been very minimal ($<$ 3 %). As discussed, the training session accuracies have been computed using a five-fold cross-validation mechanism. In the evaluation session, there are two different ways in which classification accuracies have been reported in the columns Eval-

uation1 and Evaluation2. Column Evaluation1 accuracies have been computed by creating a model with 100% of the training session data. Column Evaluation2 accuracies have been calculated by using a five-fold cross-validation mechanism on the evaluation session data. The difference between columns Evaluation1 and Evaluation2 help to show the effect of the non-stationarity across the sessions. Two different classification accuracies have been computed to compare the results and verify to what extent non-stationarity and non-linearity in the EEG data have been accounted. Column Evaluation1 reported an average classification accuracy of 79.74 while Evaluation2 reported an average classification accuracy of 80.26. Now, it is clearly evident that the average classification accuracy computed for a particular session is very minimal ($< 1\%$) using two different approaches across nine subjects. Thus, the proposed filtering has helped to handle the adverse effect of the non-stationarity to a larger extent. In the column Evaluation1, Subjects A01 and A02 have a difference of $< 1\%$ in terms of classification accuracy across training and evaluation sessions. Subjects A03, A04, A08, and A09 have obtained greater classification accuracy in the evaluation session as compared to the training session accuracies. Thus, **the results clearly show the proposed pipeline has helped to counteract the inherent intersession non-stationarity present in the EEG signals**. The difference in the group average of classification accuracies across columns Evaluation1 and Evaluation2 may be accounted using adaptive techniques/ transfer learning mechanisms (cf. 2.8.1).

Table 4.2 presents the comparison of classification accuracy values calculated with the MEMDBF-CSP method and other comparable works in the literature. The MEMDBF-CSP has shown comparable performance with one approach reported in (Lotte and Guan, 2011) and substantial improvement when compared with other research works reported in (Gandhi et al., 2014; Raza et al., 2015). The superior average classification accuracy has been achieved across nine subjects in comparison to results reported by four most recent advanced methods. The method-1 reported average classification accuracy 66.59 % ($p = 0.0273$) (Gandhi et al., 2014), method-2 reported 78.01% ($p = 0.2031$) (Lotte and Guan, 2011), method-3 computed 73.84% ($p = 0.0078$) (Raza et al., 2015) and method-4 reported 74.92%(p

= 0.0391) (Raza et al., 2015). The Wilcoxon signed rank test has been used to compute the p-values. These methods investigated the same two-class classification problem in order to classify the left and right hands MI tasks. The method-1 studied quantum neural network filtering Gandhi et al. (2014) and extracted the band power and Hjorth features collectively from three channels (C3, C4, and Cz), then classified the feature set by LDA classifier. The method-2 extracted the CSP features from all twenty-two channels on bandpass-filtered EEG data between 8-30 Hz. Thereafter, the features set was calculated by taking the log variance of three pairs (m=3) of selected filters. Finally, they classified the feature set by LDA classifier (Lotte and Guan, 2011). The method-2 considered all twenty-two channels to compute the classification accuracy while comparable results are obtained by using only fifteen channels. Method-3 and method-4 extracted the CSP feature and detected the covariate shift using ten channels and then applied adaptive learning and transductive learning to adapt the shift (Raza, 2016).

Table 4.2: Comparison of classification accuracies (%) obtained with the MEMDBF method and other state-of-the-art methods evaluated on BCI competition IV dataset 2A.

Subject	MEMDBF-CSP	Method-1	Method-2	Method-3	Method-4
A01	**90.78**	61.11	88.89	90.28	90.28
A02	58.45	**61.11**	51.39	54.17	57.64
A03	93.43	79.17	**96.53**	93.75	95.14
A04	**74.14**	60.42	70.14	64.58	65.97
A05	60.74	**71.53**	54.86	57.64	61.11
A06	68.52	61.11	**71.53**	65.28	65.28
A07	80	58.33	**81.25**	62.5	61.11
A08	**97.01**	67.36	93.75	90.97	91.67
A09	**94.62**	79.17	93.75	85.42	86.11
Average	79.74	66.59	78.01	73.84	74.92
***p*-value**		0.0273	0.2031	0.0078	0.0391

These methods investigated the same two-class classification problem to classify the LHMI and RHMI tasks. Method-1 studied quantum neural network filtering (Gandhi et al., 2014) and extracted the band power and Hjorth features collectively, then classified the feature set by an LDA classifier. Method-2 extracted the CSP

features on bandpass filtered EEG data between 8-30 Hz. Thereafter, the features set was calculated by taking the log variance of three pairs of selected filters. Finally, they classified the feature set using an LDA classifier (Lotte and Guan, 2011). Method-2 considered all twenty-two channels to compute the classification accuracy while comparable results are obtained with the MEMDBF-CSP method using only fifteen channels. Method-3 and method-4 extracted CSP features and detected the covariate shift and then applied adaptive learning and transductive learning to adapt to the covariate shifts (Raza et al., 2015). MEMDBF-CSP demonstrates a tangible improvement in classification accuracy for four of the nine subjects as marked in boldface in Table 4.2.

4.3 Study 2: Subject Independent MEMDBF

4.3.1 Methods

As discussed previously (cf. 2.8.1), one of the most challenging task is to classify motion intentions since the recorded EEG signals have inherent non-stationarities which are due to changes in the signal properties over time within a session as well as across sessions. Also, another limitation is long calibration time, which is limiting the use of BCI in patients and healthy people. EEG signals are highly subject specific and there exists a lot of non-stationarity (user variability) across sessions and subjects as well. Every time there is a need to collect numerous training data trials for machine learning methods particularly used in BCI paradigm. Thus it becomes difficult to achieve a stable operation of BCI. To this end, a novel filtering method based on the MEMD using subject independent pooled design BCI (MEMD-SI-BCI) for classification of MI based EEG signals is proposed to achieve enhanced BCI. A block diagram of the proposed method is shown in Figure 4.7.

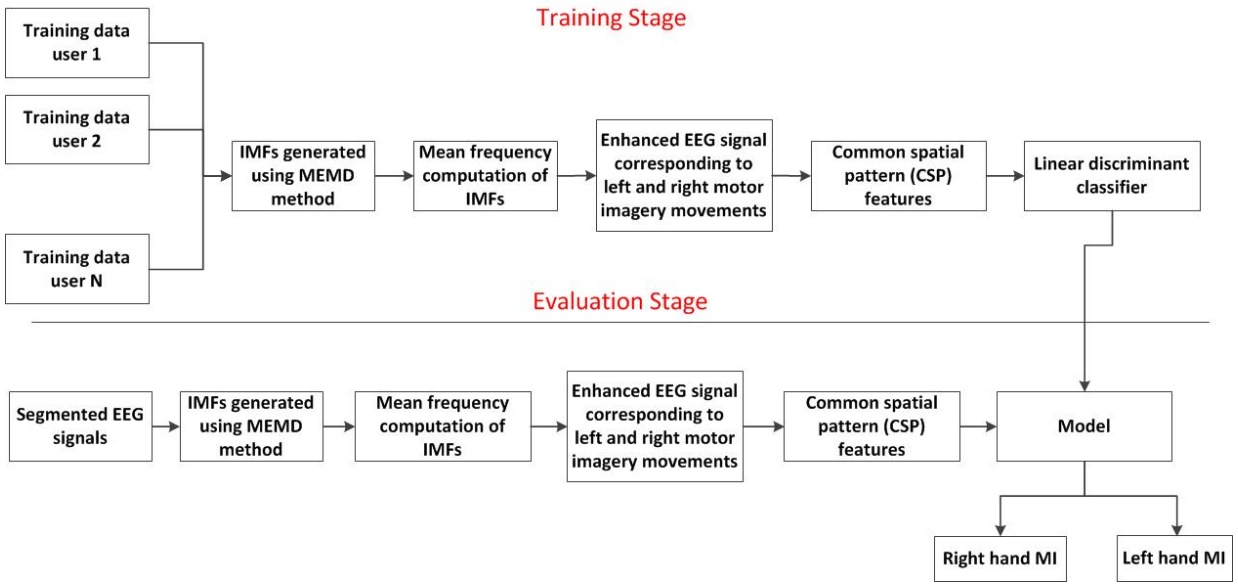


Figure 4.7: Block diagram of the MEMD-SI-BCI method.

4.3.2 Dataset

The MEMD-SI-BCI method has been investigated on the BCI competition IV dataset 2A (Brunner et al., 2008). For more information about the dataset, refer to cf. 3.2.2.

4.3.3 Multivariate Empirical Mode Decomposition (MEMD)

The left hand and right hand MI EEG signals using the MEMD-SI-BCI method are decomposed using MEMD method. More details of MEMD method are provided in Section 2.10.2 of Chapter 2. The original EEG signal and its computed IMFs with the MEMD method of left hand and right hand MI EEG signals are shown in Figure 4.3 and Figure 4.4 respectively.

4.3.4 Common Spatial Pattern (CSP)

The CSP algorithm was again used as described previously in section 4.2.4 of this chapter. As reported in the literature, there is no fixed algorithm to select the spatial

filters. The number of spatial filters (m) has been selected randomly such as (Lotte and Guan, 2011) considered m=3 spatial filters whereas m=1 spatial filters are selected by (Raza et al., 2015). Here, m is taken as the first and last columns of the CSP matrix. In this study, $m = 4$ and $m = 5$ spatial filters are considered. All the combinations taking m=1,2,3,...,7 was taken into consideration and finally, $m = 4$ and $m = 5$ spatial filter were identified which was providing better classification accuracy in a training session with a five-fold cross-validation scheme. These selected spatial filters from the training CSP matrix were used to compute the CSP features in the evaluation session. Further, these extracted CSP based features an input features for the LDA classifier. Although this study considered fixed set of fifteen channels from the provided twenty-two channels for all nine subjects but to counteract the inter-subject non-stationarity, subject specific could have been done.

4.3.5 Linear Discriminant Analysis

The classification of left hand and right hand MI EEG signals is performed using LDA classifier (cf. 3.2.5).

4.3.6 Study2 : Results and Discussion

The MEMD-SI-BCI based filtering has been evaluated on publicly available BCI competition IV dataset 2A (Brunner et al., 2008). In this study, EEG signals recorded are considered from **fifteen channels** (Figure 4.2) related to the sensorimotor areas. Although the actual data were recorded from twenty bipolar EEG channels and three Electrooculography (EOG) channels with a sampling frequency of 250 Hz (cf. 3.2.2). There are seventy two trials provided in each session. Each trial involved a paradigm period of 7.5 second (Brunner et al., 2008). In the training stage, a single session namely $\hat{}T$ has been used. For the evaluation phase, one session namely, $\hat{}E$ has been used for computing the accuracy in the classification of left and right

MI EEG signals. It should be noted that, as in earlier studies, the ^ in the session name denotes the subject number which again ranges from A01 to A09.

Table 4.3: Classification accuracies using the proposed method based on MEMD and without MEMD studied on BCI competition IV dataset 2A.

Subject	Accuracy with MEMD-SI-BCI				Accuracy with SI-BCI			
	Training	Evaluation	Training	Evaluation	Training	Evaluation	Training	Evaluation
	$m = 4$		$m = 5$		$m = 4$		$m = 5$	
A01	70.14	91.67	72.16	92.36	68.2	69.44	68.2	68.06
A02	70.21	55.56	71.07	58.33	67.91	49.31	67.9	52.78
A03	69.22	90.97	67.45	91.67	68.36	70.83	68.14	71.53
A04	70.6	62.5	70.68	63.89	68.21	59.72	67.91	58.33
A05	68.91	61.11	71.45	59.03	68.29	49.31	68.52	49.31
A06	70.37	68.06	71.29	67.36	68.44	55.56	68.68	55.56
A07	70.29	61.11	70.37	60.42	68.3	50.69	67.98	51.39
A08	70.52	96.53	70.67	96.53	68.29	91.67	69.14	91.67
A09	65.44	65.97	67.82	66.67	67.9	54.86	68.36	56.94
Average	69.52	72.61	70.33	72.92	68.21	61.27	68.31	61.73
Std	1.64	15.79	1.62	15.81	0.19	13.95	0.41	13.45
p-value	0.034	0.001	0.005	0.001				

In order to compute the classification accuracy (in %) in the training stage, a 5-fold cross-validation has been applied. For the evaluation session, the model has been trained with an LDA classifier using 100% training data from the nine subjects and evaluated on the 100% data for the session, A0βE, where β denotes the subject number. The classification model has been used to classify sequentially the evaluation session data of all the nine subjects. Since the MI task starts at 3 seconds, the features are extracted in both of the training and evaluated sessions corresponding to EEG signals from 0.5 second to 2.5 second time-interval after the start of the MI paradigm. The classification accuracy is computed using an LDA classifier for two-class classification of the left and right hand MI EEG signals in both training and evaluation sessions for each of the subjects.

In order to explain the working of the MEMD method, two single trial EEG signals are considered of fifteen channels from the dataset A01T to obtain IMFs but for

illustration the plot for five channels is shown. The left MI EEG signal and its first nine IMFs are shown in Figure 4.3. Similarly, the Figure 4.4 depicts the right hand MI EEG signal from the five channels and its obtained IMFs.

The statistical measure, namely, mean frequency has again been calculated for each multivariate IMFs of the EEG signals from the selected fifteen channels in the motor cortex region corresponding to left and right hand MI tasks. To achieve enhanced EEG signals corresponding to left and right hand MI tasks, the IMFs whose mean frequencies fall in the range 6-24 Hz were selected. This frequency range takes into account the μ band (8-13 Hz) and low β band (18-24 Hz) which have considerable importance in classifying left and right hand MI EEG signals as mentioned in (cf. 4.2.4). The CSP feature is then computed for the enhanced EEG signals obtained using the selected IMFs. In our study the results obtained are reported using two spatial filters $m = 4$ and $m = 5$ where m denotes the first m and the last m columns of spatial filter matrix. Then, the extracted feature has been fed as an input feature to the LDA classifier for classification of left and right hand MI EEG signals.

Table 4.3 presents the classification accuracy with MEMD based filtering-SI-BCI (MEMD-SI-BCI) and with the raw SI-BCI method for BCI competition IV dataset 2A. This method has provided enhanced EEG signals using subject independent MEMD-BCI for the each of nine subjects in both training T and evaluation E sessions respectively. In this study, only fifteen channels corresponding to motor cortex have been considered of the provided twenty two channels for obtaining the results. Comparing the MEMD-SI-BCI results, it is clear that the new method presented in this paper **provides a significant improvement in classification accuracy in the evaluation sessions of all nine subjects** when compared with the raw SI-BCI results. It has shown improvement $> 11\%$ ($p < 0.001$) in the evaluation session with $m = 4$ and $> 11\%$ ($p < 0.001$) in evaluation session with $m = 5$. In the training session, since a generalized model is created for all the subjects, there is a slight improvement in the classification accuracy. All nine sessions have shown improvement in the evaluation stage with $m = 4$. A total of seven out of the nine

sessions have shown a **highly significant improvement** $>$ **10%** and the other two sessions have shown an improvement $>$ 2%. On the other hand, with $m = 5$, two of the subjects have shown **significant improvement of** $>$ **20%** and a total of seven sessions have shown improvement in the range of $>$ 4% and $<$ 24%. To conclude, with $m = 4$ and $m = 5$, an improvement is achieved in all of the eighteen sessions of nine subjects. The p-values for all the nine subjects was calculated using the anova2 command available in MATLAB which calculates 2-way analysis of variance. The non-parametric Kruskal-Wallis test was also done on the evaluation session and p-value obtained for column evaluation with m=4 gives 0.08 and 0.05 for m=5.

4.4 Conclusion

This chapter sought to address the second (C2) and three (C3) contributions of the thesis by investigating the effectiveness of an extension to the EMD method known as multivariate empirical mode decomposition (MEMD) (cf. 2.10.2) based filtering when applied to multichannel EEG data in a motor imagery based BCI paradigm.

In the first study the classification accuracy obtained from the MEMD based filtering has shown significant improvement during both the training and the evaluation stages across multiple sessions. Additionally, when compared against both raw EEG signals and a quantum neural network based EEG filtering method (Gandhi et al., 2014), MEMD based filtering in conjunction with CSP features has shown superior performance in terms of classification accuracy.

In the second study, which explored an application of MEMD based subject independent design (MEMD-SI), significant improvements were achieved in terms of classification accuracy of left and right hand MI EEG signals when compared with raw EEG signals.

This chapter explored an application of this MEMD based filtering method with common spatial pattern (CSP) features for enhancing performance of two-class MI based BCI. The key idea is that at the preprocessing stage, MEMD based filtering removes inherent non-stationarity present in the EEG signals to some extent whilst also filtering artifacts and noise. The enhanced EEG signals have zero mean and there is no complexity introduced at the feature extraction or the classification stages. A highly significant performance has been obtained in MI based BCI simply by enhancing the EEG signals at the pre-processing stage. A selection of multiple IMFs whose mean frequencies fall in the frequency range of μ and β rhythms have helped to improve the classification accuracy of left hand and right hand MI EEG signals as compared to raw EEG signals. The next two chapters will address the last two contributions of this thesis (C4 and C5) by varying the features and different classification techniques, and proposing a novel tangent space based transfer learning classification method.

Chapter 5

Subject Specific Multivariate Empirical Mode Decomposition based Filtering

5.1 Introduction

The previous chapter investigated the effectiveness of the proposed multivariate empirical mode decomposition (MEMD) based filtering when applied to multi-channel electroencephalogram (EEG) data in a motor imagery (MI) based brain-computer interface (BCI) paradigm. Additionally, it presented an MEMD based subject independent design (MEMD-SI), wherein significant improvements were achieved in terms of classification accuracy of left and right hand MI EEG signals. This chapter will address the fourth contribution for this thesis (C4) by introducing a subject specific multivariate empirical mode decomposition (SS-MEMD) filtering method which seeks to improve the performance by further improving the classification accuracy and by proposing a novel tangent space based transfer learning classification method. The chapter concludes by comparing the MEMD technique to other similar works using Riemannian geometry demonstrating a substantial

performance improvement in multi-class MI based BCI classification.

The aims of this chapter are as follow:

1. To study the intra- and inter-subject non-stationarity present in the EEG signals;

2. To identify whether the same frequency or different frequency components are involved in motor imagery activity when EEG signals are recorded from the same cortical areas across the subjects;

3. To automatically identify the subject specific MIMFs based on mean frequency measure;

4. To present the results in terms of classification accuracy and Kappa value when single trials are classified.

5.2 Methods

A new subject specific MEMD based filtering method for classification of multi-class MI tasks is proposed by extending the previously described filtering technique (Gaur et al., 2015) based on EMD which is restricted to two-class MI tasks in BCI. The enhanced EEG signals have been obtained corresponding to μ and β rhythms before extracting features in order to classify right hand, left hand, foot and tongue MI tasks. The capability of the existing MEMD method (Davies and James, 2013, 2014; Park et al., 2013, 2014) has been utilized to decompose the signal into a set of narrow-band MIMFs and subject specific MIMFs. These have been selected based on a statistical measure namely, mean frequency, corresponding to the μ and β rhythms of each subject. Thereafter, the summation of selected MIMFs is performed to attain enhanced EEG signals corresponding to a particular subject. A block diagram of the proposed subject specific MEMD based filtering methodology is shown in Figure 5.1.

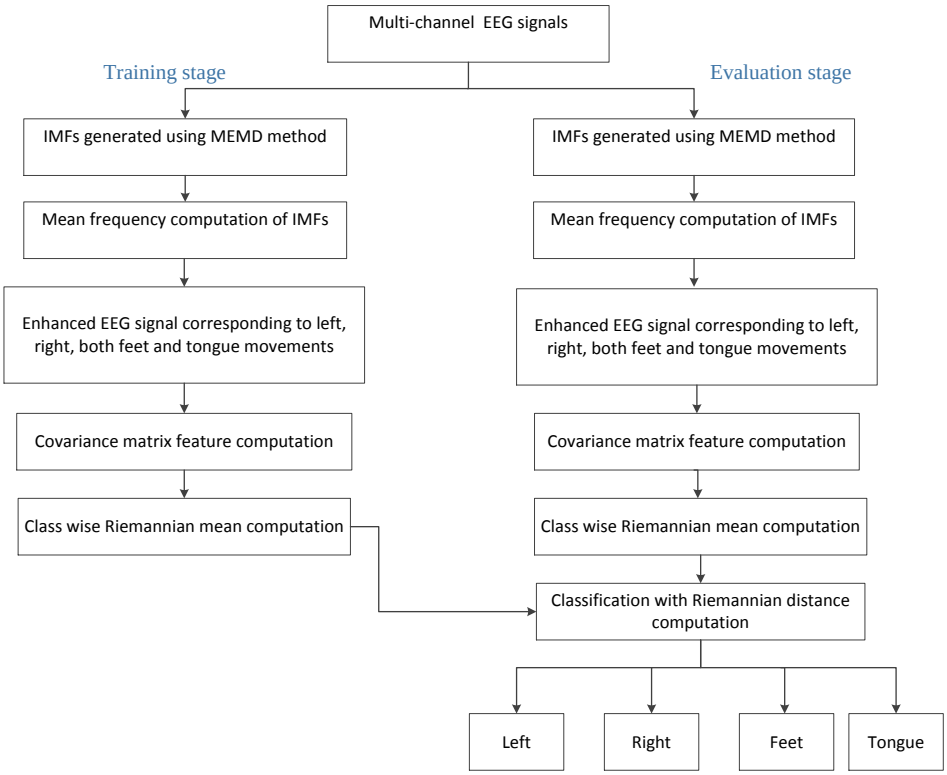


Figure 5.1: Block diagram for the proposed method.

5.3 Dataset

In this study, the publicly available BCI competition IV dataset 2A (Brunner et al., 2008) has been considered. All the twenty-two channels have been used in this study as shown in Figure 5.2. The reason for increasing the number of channels selected from three channels to fifteen channels and then to twenty-two channels was to test whether considering all the twenty-two channels helped to obtain more classification accuracy. There is no fixed method to select the number of channels in the literature with some considering three channels (Gandhi et al., 2014), 10 channels (Raza, 2016), and 22 channels (Lotte and Guan, 2011). The enhanced EEG signals from all twenty-two channels have been used to extract the sample covariance matrix (cf. 5.5) as a feature set.

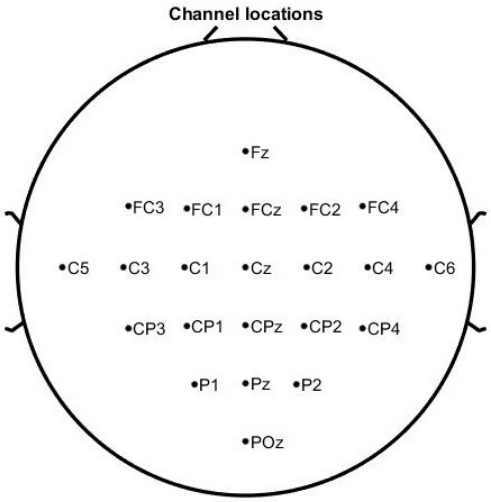


Figure 5.2: Head plot showing all the channels locations.

5.4 Multivariate Empirical Mode Decomposition

The left hand, right hand, feet and tongue MI EEG signals are decomposed using the MEMD method as in previous studies (cf. 4.2.3) and described in greater depth in (cf. 2.10.2). The decomposition dynamics of the MEMD technique may be explained using single trials of left hand and right hand MI EEG signals. Figs. 4.3 and 4.4 display the obtained MIMFs from the subject A08T. Similarly, the foot and tongue MI EEG signal and its obtained MIMFs are shown in Figs. 5.4(a) and 5.4(b) respectively.

5.5 Feature Extraction

In BCI literature, some research groups have studied EEG features such as Hjorth parameters, band power, power spectral density (PSD) and bispectrum (BSP) features (Brunner et al., 2008; Davies and James, 2013, 2014; Leeb et al., 2008; Lotte et al., 2007; Shahid and Prasad, 2011). The most commonly used features in MI based BCI applications for classification of right hand, left hand, foot and tongue MI tasks are Hjorth features and band power (Bajaj and Pachori, 2012; Gaur et al., 2015; Pachori, 2008; Park et al., 2011; Wolpaw et al., 2002). Also, some of the re-

search groups investigated spatial filters technique namely, common spatial pattern (CSP) as used in chapter 3 and its extension for multi-class MI tasks classification problems. This method enhances class separability by estimating the eigenvectors. Thereafter, the spatial patterns are derived from these eigenvectors (Ang et al., 2008).

In this chapter, the structure of sample covariance matrix has been exploited as a feature set, as the sample covariance matrix contains the spatial information present in EEG signal. The main objective is to devise a unique step by combining the spatial filtering and the classification. However, sample covariance matrices structure needs to be handled carefully in Riemannian manifold.

Sample Covariance Matrix

Let $x_j \in \Re^r$ represent the enhanced EEG signal vector at a particular time instant *j* where *r* represents the number of electrodes. The formal definition of spatial covariance matrix is given by Cov $= E\{(x_j - E\{x_j\})(x_j - E\{x_j\})^T\}$, wherein $E\{.\}$ depicts the expected value and superscript ^{T} represents matrix transposition. For designing a BCI, a short time segments are extracted from continuous EEG signals of a trial. They may be denoted in the form of a matrix $X_i = [x_{j+T_i} ... x_{j+T_i+T_n-1}] \in R^{r \times T_n}$ corresponding to i-th trial of an MI task beginning at time T_i. Furthermore, the spatial covariance matrix corresponding to the *i*-th trial is computed using the unbiased SCM $P_i \in R^{r \times r}$ such as,

$$Q_i = \frac{1}{T_n - 1} X_i X_i^T \tag{5.5.1}$$

where T_n represents the number of time instants taken from each trial.

Shrinkage covariance matrix

The shrinkage covariance matrix (SHCM) is computed from the enhanced EEG/MEG signals obtained from single channel or multi channel filtering methods. Let

$F_1, F_2F_3...F_f$ denote the f feature vectors. Let unbiased estimator of the mean is given as,

$$\hat{M} = \frac{1}{f}\sum_{i=1}^{f} F_i \tag{5.5.2}$$

Also, the unbiased estimator of the covariance matrix is denoted as,

$$\hat{C} = \frac{1}{f-1}\sum_{i=1}^{f}(F_i - \hat{M})(F_i - \hat{M})^\top \tag{5.5.3}$$

To account for the estimation error, $\hat{C}$ is substituted by

$$C(\gamma) = (1-\gamma)\hat{C} + \gamma\upsilon I \tag{5.5.4}$$

with a tuning parameter ($\gamma \in$) $[0, 1]$. υ denotes the average eigenvalue and I gives the identity matrix. More details may be obtained from Blankertz et al. (2011).

5.6 Riemannian Geometry Framework

Let, the space of all $r \times r$ symmetric matrices (SM) $\{P(r) \in Z(r), P^T = P\}$ is defined in the space of square real matrices $Z(r)$. Also, the set of all $r \times r$ symmetric positive definite (SPD) matrices is represented by $Q(r) = \{Q \in P(r), v^T Q v > 0, \forall v \in \Re^r\}$ and the set of all $r \times r$ invertible matrices is denoted by $Gl(r)$ in $Z(r)$. Lastly, the notation $Q^{1/2}$ denotes a symmetric matrix B such that it satisfies the relation $BB = Q$.

As shown in Figure 5.3, the space of SPD matrices $Q(r)$ is denoted by a differentiable Riemannian manifold Z (Moakher, 2005). The tangent space for the derivatives of a matrix Q lies in a vector space T_Q over the manifold. The dimensions of the tangent space and the manifold are $d_{ts} = r(r+1)/2$.

The inner product $<,>_Q$ of each tangent space changes smoothly from one point to another over the entire manifold. The definition of natural metric over the entire

SPD manifold is given by inner product represented as follows:

$$< P_1, P_2 >_Q = Tr(P_1 Q^{-1} P_2 Q^{-1}) \tag{5.6.5}$$

A norm for tangent vectors is induced by inner product on the tangent space given by, $||P||_Q^2 =< P, P >_Q =$ Tr$(PQ^{-1}PQ^{-1})$. It should be noted that such a norm reduces into the Frobenius norm in case of identity matrix i.e., $< P, P >_I = ||P||_F^2$.

In the following subsections, the Riemannian geometry framework details will be discussed and also, how the MI tasks classification is done using the Riemannian framework.

Geodesic Distance

Let $D(t) : [a, b] \rightarrow Q(r)$ denotes any form of differentiable path from $D(a) = Q_1$ to $D(b) = Q_2$. The length of $D(t)$ is expressed as,

$$L(D(t)) = \int_a^b ||D(t)||_{D(t)} dt \tag{5.6.6}$$

with the norm as discussed earlier. The geodesic distance on the manifold is defined as the minimum length curve joining these two points. The Riemannian distance is defined by the length of this curve. The geodesic distance is induced by the natural metric (5.6.5) and is given by,

$$\delta_R(Q_1, Q_2) = ||\text{log}(Q_1^{-1} Q_2)||_F = \left[\sum_{i=1}^{r} Log^2 \chi_i \right]^{1/2} \tag{5.6.7}$$

where $\chi_i, i = 1, ..., r$ denotes the real eigenvalues of $Q_1^{-1} Q_2$ and r represents the number of channels.

Following describes the main properties of the Riemannian geodesic distance, (Moakher, 2005):

1. $\delta_R(Q_2, Q_1) = \delta_R(Q_1, Q_2)$

2. $\delta_R(Q_1^{-1}, Q_2^{-1}) = \delta_R(Q_1, Q_2)$

3. $\delta_R(W^T Q_1 W, W^T Q_2 W) = \delta_R(Q_1, Q_2) \; \forall \, Gl(r)$.

Exponential Map

A tangent space for each point $Q \in Q(r)$ is composed by a set of tangent vectors defined at Q. Each tangent vector P_i is seen as the derivative at $t = 0$ of the geodesic $D_i(t)$ between exponential mapping $Q_i (= \text{Exp}_Q(P_i))$ and Q, defined as,

$$\text{Exp}_Q(P_i) = Q_i = Q^{1/2} \text{exp}(Q^{-1/2} P_i Q^{-1/2}) Q^{1/2} \tag{5.6.8}$$

The logarithmic mapping gives the inverse mapping which is defined as,

$$\text{Log}_Q(Q_i) = P_i = Q^{1/2} \text{log}(Q^{-1/2} Q_i Q^{-1/2}) Q^{1/2} \tag{5.6.9}$$

The geometrical procedure has been presented in Figure 5.3. Riemannian distance equivalent definitions are as follows,

$$\begin{aligned} \delta_R(Q, Q_i) = ||\text{Log}_Q(Q_i)||_Q = ||P_i||_Q \\ = ||\text{upper}(Q^{-1/2} \text{Log}_Q(Q_i) Q^{-1/2})||_2 = ||p_i||_2 \end{aligned} \tag{5.6.10}$$

where the upper(.) operator describes to keep the upper triangular part of a SM and vectorizing the diagonal elements with unity weight and non-diagonal elements with $\sqrt{2}$ weight (Moakher, 2005). Here p_i is the d_{ts}-dimensional vector upper$(Q^{-1/2} \text{Log}_Q(Q_i) Q^{-1/2})$ of the normalized tangent vector.

Some conditions over Q and the Q_i (Tuzel et al., 2008), (7) may leads to an approximation of computing distance between the manifold and the tangent space, such as

$$\forall i, j \qquad \delta_R(Q_i, Q_j) \approx ||p_i - p_j||_2 \tag{5.6.11}$$

These conditions can be verified if Q_i is locally distributed in the manifold. It means that Q must always represent the mean of the Q_i.

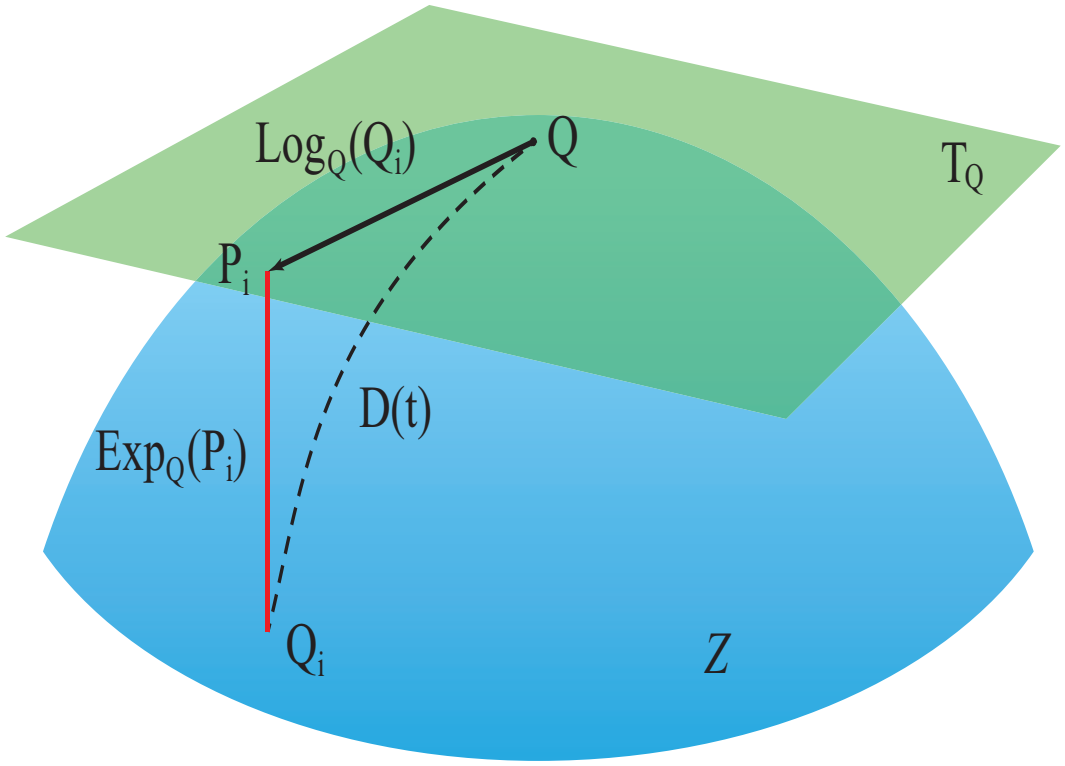


Figure 5.3: The tangent space at point Q. P_i is a tangent vector at Q. $D(t)$ is the geodesic between Q and Q_i.

Riemannian Mean

The Riemannian mean of $J \times 1$ SPD matrices using Riemannian geodesic distance (5.6.7) is denoted by,

$$G(Q_1, ..., Q_J) = \underset{Q \in Q(r)}{\text{argmin}} \sum_{i=1}^{J} \delta_R^2(Q, Q_i) \tag{5.6.12}$$

This is called geometric mean. Considering 1×1 SPD matrices $y_i > 0_{1\times J}$, this definition gives $G(y_1, ..., y_J) = \underset{y>0}{\text{argmin}} \sum_{i=1}^{J} \log^2(y_i/y) = \sqrt[J]{y_1, ..., y_J}$

To compute the Riemannian mean of J SPD matrices an efficient algorithm is discussed in (Barachant et al., 2012; Moakher, 2005).

Motor Imagery Classification Using Riemannian Geometry Framework

For each MI task in the training stage, a set of trials, denoted by $\{X_i^{tr}, i \in J^{(K)}\}$ corresponding to K-th condition, is obtained. The $J^{(K)}$ denotes the set of trial indices corresponding to the K-th condition. For each single trial, the SCM is computed serving as a feature vector using (5.5.1). These SCMs belong to the manifold Z since they are SPD matrices. Let Q be the SCM of the trial X^{ts} from test data. It is possible to derive an efficient classification algorithm using the equations presented in Section 5.6 to compute the unknown test label of a trial X^{ts}.

Classification Algorithm using Minimum Distance to Riemannian Mean

For each of the MI tasks, the class-wise covariance matrices $Q_G^{(c)}$ have been calculated with the equations discussed in Section 5.6, where $c = [1 : K]$ depicts the class indices. A simple algorithm has been used to compute the distance between class-wise covariance matrix $Q_G^{(c)}$ and an unknown test trial SCM typically known as Riemannian distance. The class giving the minimum distance is assigned to the unknown test trial X^{ts}. This procedure has been done to assign the class to each of the new test trials. The details of the algorithm are as follow:

Algorithm:

Input : a set of trials X_i^{ts} of different unknown classes.
Input : $J^{(c)}$ denotes the set of indices for the trials corresponding to the c-th class.
Output : $\hat{K}$ gives the estimated class of unknown test trial X^{ts}

1. Compute SCM of X_i^{ts} to obtain Q_i^{ts}, from eq. (5.5.1).
2. Compute SCM of X_i^{tr} to obtain Q_i^{tr}, from eq. (5.5.1).
3. $for\ c\ =\ 1\ to\ K\ do$
4. $Q_G^{(c)} = G(Q_i^{tr}, i \in J^{(c)})$, from eq. (5.6.12).

5. $end\ for$

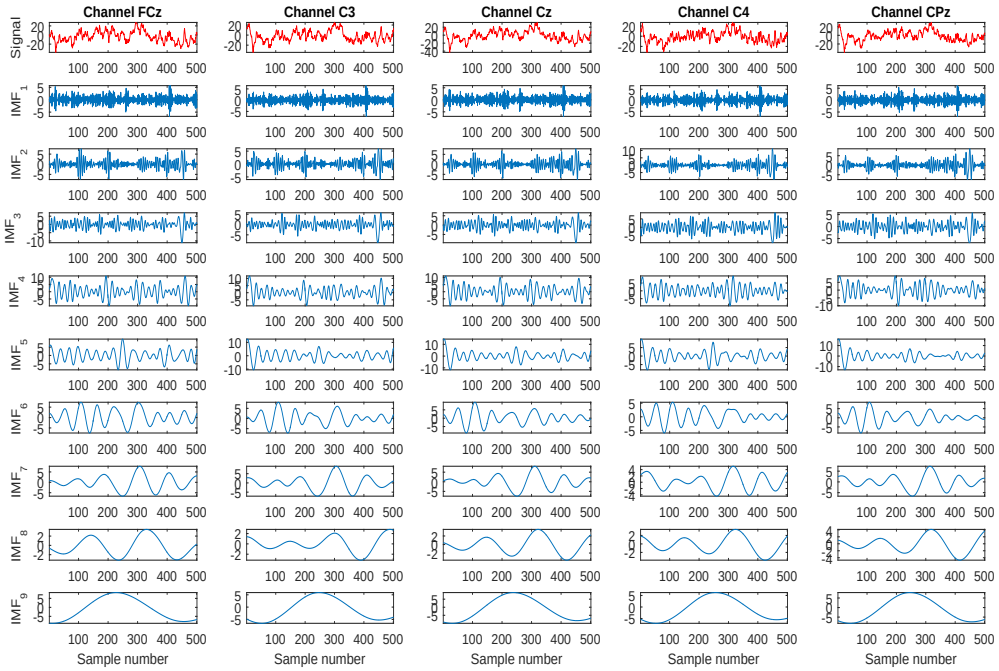
6. $\hat{K} = \underset{x}{\text{argmin}}\ \delta_R(Q_i^{ts}, Q_G^{(c)})$, from eq. (5.6.7).

5.7 Results and Discussion

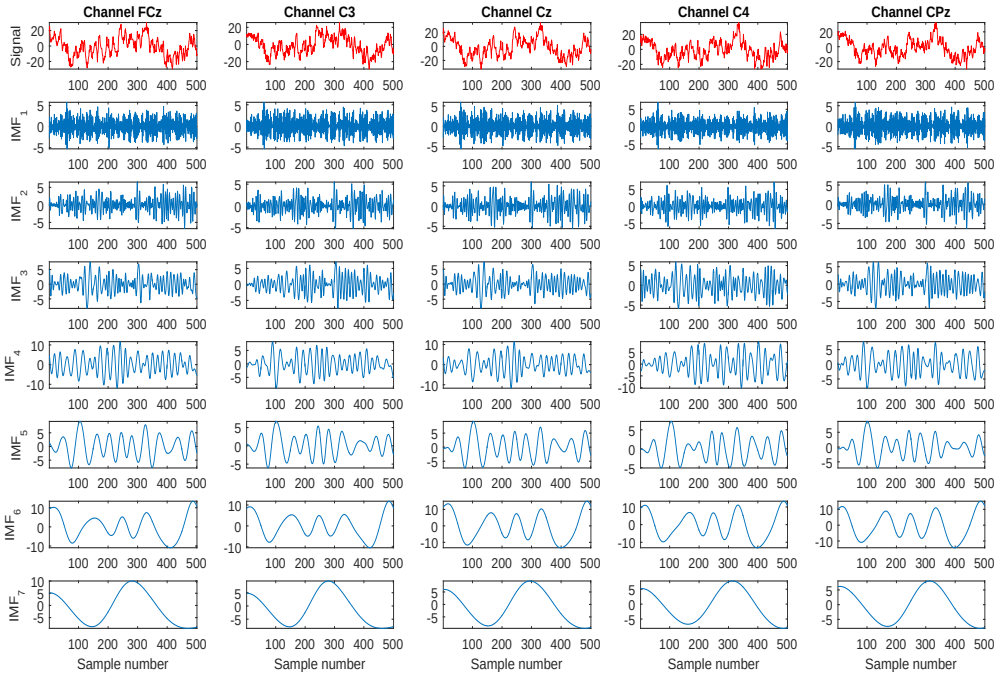
In the training session, a 5-fold cross-validation has been done to avoid the overfitting issue during the classification of multi-class MI EEG signals. For the evaluation session, the Riemannian mean covariance matrix for each of the classes has been obtained with all training trials from the A0ST and evaluated on all unknown test trials of the corresponding evaluation session A0SE on a trial-by-trial basis, where S represents the subject number. As the MI task began at 2 seconds, the covariance matrix feature is computed for the EEG signals from the 2.5 second to 4.5 second time-interval of the MI paradigm. The Kappa value (cf. eq. 2.5.2) (between 0 and 1) and classification accuracy for all the nine subjects have been computed.

The mean frequency measure (cf. eq. 3.2.1) has been computed for all the MIMFs corresponding to right hand, left hand, both feet and tongue MI tasks. This mean frequency was used to identify the MIMFs which contribute to μ and β rhythms and remaining of the MIMFs were discarded. To enhance EEG signals corresponding to left hand, right hand, foot and tongue MI tasks, a subject specific filtering range is identified based on the feature separability in the training session. The obtained MIMFs have been summed to attain the enhanced EEG signals. The identified subject specific frequency range has been reported in Table 5.5.

The results in Figure 5.5(a) and Figure 5.5(b) show the feature distribution of first five best features. Figure 5.5(a) depicts the boxplot obtained using Kruskal-Wallis test for the first five best features with the SS-MEMDBF. After applying the SS-MEMDBF, the first five best features extracted have shown statistically significant improvement in separability with p-value < 0.05 in the training session for the left hand and right hand MI tasks. The feature selection has been done by ranking the



(a)



(b)

Figure 5.4: The EEG signals and its MIMFs corresponding to channels FCz, C3, Cz, C4 and CPz of A08T for (a) Foot MI task (b) Tongue MI task.

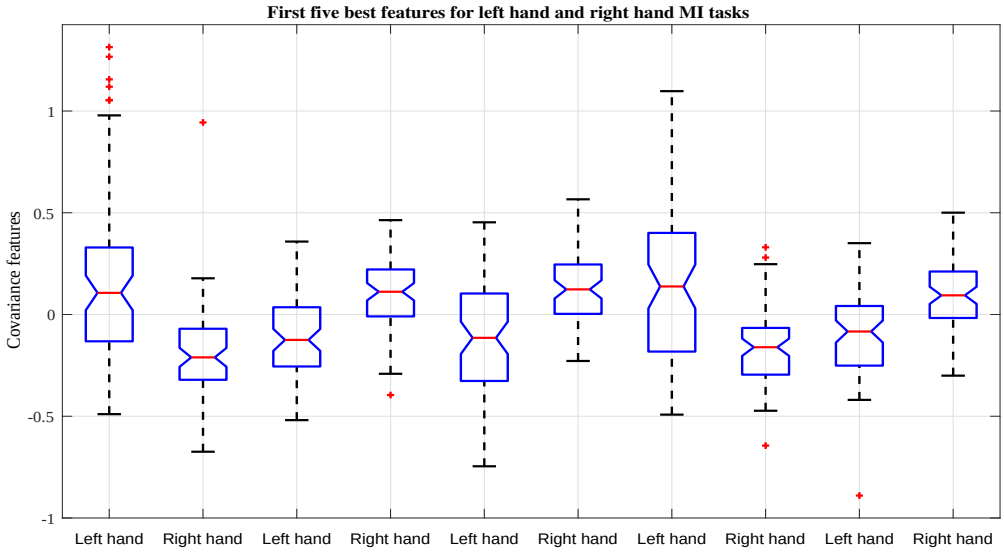
available features with the Wilcoxon test method for left hand and right hand MI tasks. The feature set has been arranged based on the rank in decreasing order of class separability. Then the best five features were selected giving the highest class discrimination for left hand and right hand MI task classification as shown in Figure 5.5(a) with the proposed method. For illustration, the first five features are shown although there are more features in the training session which are statistically significant in terms of features separability. Figure 5.5(b) shows the same five features computed from the raw EEG signals giving p-values of 0.1065, 0.2861, 0.0643, 0.2111 and 0.6983. The p-values indicate that the two classes have no significant difference in their feature distribution for left hand and right hand MI tasks. However, with the proposed method, the p-values show statistically significant difference in feature distribution.

Table 5.1 gives the details of the trials rejected from each subject in the evaluation session as indicated with the event 1023 (Brunner et al., 2008). Subject A06 has maximum number of the trials rejected. The rejected trials corresponding to each MI tasks are as follow: left hand 19 trials, right hand 17 trials, foot 18 trials and tongue 19 trials respectively. This gives the total of 73 trials rejected for subject A06 in the evaluation session.

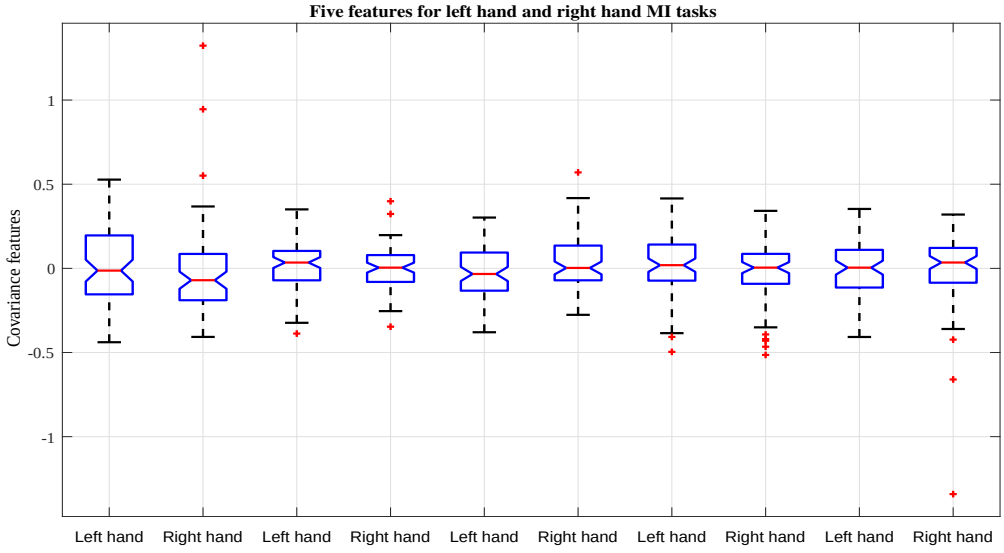
Table 5.1: Trials rejected from all subjects.

Subject	Total trials	Correct trials	Rejected trials	Left hand	Right hand	Foot	Tongue
A01	288	281	7	1	2	3	1
A02	288	283	5	1	1	3	0
A03	288	273	15	5	2	4	4
A04	288	228	60	13	15	13	19
A05	288	276	12	2	7	0	3
A06	288	215	73	19	17	18	19
A07	288	277	11	1	3	1	6
A08	288	271	17	6	4	3	4
A09	288	264	24	7	7	3	7

Table 5.2 presents the results obtained using the pairwise binary classification for



(a) Kruskal-Wallis Test on enhanced EEG signals with SS-MEMDBF for first five best features.



(b) Kruskal-Wallis Test on raw EEG signals for same five features.

Figure 5.5: The box plot depicts the first five best features obtained in the training session for left hand and right hand MI tasks (a) SS-MEMDBF (b) raw EEG signals.

multiple MI tasks. Since there are four classes, we can have six possible pairs of imagery tasks: left vs right (LvR), left vs foot (LvF), left vs tongue (LvT), right vs foot (RvF), right vs tongue (RvT), foot vs tongue (FvT). The features obtained with the SS-MEMDBF have helped to achieve superior classification accuracy.

Table 5.2: Classification accuracy (in %) for the proposed method with and without SS-MEMDBF with one vs one scheme applied on BCI competition IV dataset 2A.

Subject	Accuracy with SS-MEMDBF (in %)					
	LvR	LvF	LvT	RvF	RvT	FvT
A01	91.49	97.14	98.59	98.56	100	75
A02	60.56	78.57	67.13	80.71	68.53	74.47
A03	94.16	91.11	88.89	97.1	95.65	69.85
A04	76.72	91.53	87.5	92.24	86.36	66.07
A05	58.52	68.31	71.94	67.88	70.9	70.92
A06	68.52	69.16	73.58	66.97	75	67.29
A07	78.57	92.25	89.05	93.57	90.37	78.83
A08	97.01	88.89	83.58	91.97	83.09	81.02
A09	93.85	92.54	98.46	79.85	83.85	89.55
Average	79.93	85.5	84.3	85.43	83.75	74.78

Table 5.3 displays the comparison of classification accuracy using SS-MEMDBF and bandpass filtering along with Riemannian geometry framework. In addition, it shows the comparison of SS-MEMDBF with CSP features as well. Method-1 shows the results obtained by performing bandpass filtering in the range of 4 and 30 Hz with the same time window from which features are extracted. These features have been classified using Riemannian geometry framework. **With the proposed method, the group accuracy has improved by $>$ 4 % (p = 0.0078) and eight of the nine subjects have shown improvement in the accuracy**. Method-2 displays the classification accuracy in the evaluation session using the proposed SS-MEMDBF with Method-2 displays the classification accuracy in the evaluation session using the proposed SS-MEMDBF with CSP features. In this case, although an improvement of $>$ 1.6 % (p = 0.7422) was obtained across the nine subjects but the results obtained are not significant in terms of p-value. Method-3, method-

4 and method-5 display the comparison of classification accuracy obtained in the evaluation session with other published results. An average improvement in classification accuracy was achieved across nine subjects in comparison to those obtained by three most recent advanced techniques, namely the method-3 ($p = 0.20$) (Lotte and Guan, 2011), the method-4 ($p = 0.027$) (Raza, Cecotti, Li, and Prasad, 2015), and the method-5 ($p = 0.0039$) (Raza, Cecotti, Li, and Prasad, 2015). These methods studied the binary classification to classify the left hand and right hand MI tasks. Method-3 implemented CSP on bandpass filtered EEG between 8-30 Hz. Then computed the log variance from three pairs of filters in the feature extraction, further classified by linear discriminant analysis (LDA)(Lotte and Guan, 2011). Method-4 and method-5 studied the CSP feature and identified the covariate shift followed by adaptive learning and transductive learning respectively (Raza et al., 2015). Using the proposed technique, there is a substantial improvement in four of the nine subjects in terms of classification accuracy as reported in Table 5.3 marked in boldface. The p-value has been computed using the Wilcoxon signed rank test using signrank command in Matlab. Subject A05 has shown very low classification accuracy except in the case when the number of channels is reduced. Tables 4.2 and 5.3 show the classification accuracy obtained for the subject A05 and when comparing classification accuracies with other methods. In Table 4.2, method-1 used only three channels and reported the accuracy of 71.53%. In Table 5.3, method-4 and method-5 reported a classification accuracy of 57.64% and 61.11% with ten channels. Method-3 reported a classification accuracy of 54.86% with twenty-two channels. With our proposed method, we obtained a classification accuracy of 60.74% with fifteen channels (reported in Table 4.2) and 58.52% with all twenty-two channels (reported in Table 5.3). To conclude, for subject A05, finding the optimum number of channels may help to obtain better classification accuracy.

Subject-specific filtering range was identified for all nine subjects for each pair of MI tasks presented in Table 5.4.

The Kappa value was also computed using Riemannian geometry with one vs rest (OVR) scheme for the multi-class classification of MI tasks on the same dataset.

Table 5.3: Classification accuracy comparison (in %) for LvR task of the proposed method with other published results using one vs one scheme applied on BCI competition IV dataset 2A.

Subject	Evaluation session comparison with other groups (in %)					
	Proposed method	Method-1	Method-2	Method-3 (Lotte and Guan, 2011)	Method-4 (Raza et al., 2015)	Method-5 (Raza et al., 2015)
A01	91.49	85.11	92.91	88.89	90.28	90.28
A02	**60.56**	56.34	59.86	51.39	54.17	57.64
A03	94.16	91.97	95.62	96.53	93.75	95.14
A04	**76.72**	70.69	66.38	70.14	64.58	65.97
A05	58.52	56.3	62.96	54.86	57.64	61.11
A06	68.52	68.52	68.52	71.53	65.28	65.28
A07	78.57	65	68.57	81.25	62.5	61.11
A08	**97.01**	94.78	96.27	93.75	90.97	91.67
A09	**93.85**	92.31	93.08	93.75	85.42	86.11
Average	79.93	75.67	78.24	78.01	73.84	74.92
p-value		0.0078	0.7422	0.2031	0.0039	0.0273

Table 5.4: Subject specific filtering range for all six possible MI tasks with the proposed method applied on BCI competition IV dataset 2A.

Subject	LvR		LvF		LvT		RvF		RvT		FvT	
	Low	high	Low	high	Low	high	Low	high	Low	high	Low	high
A01	9	30	9	25	9	25	4	26	8	25	9	26
A02	6	25	9	29	9	29	7	26	8	22	9	30
A03	7	27	7	23	4	25	8	24	5	26	10	26
A04	7	30	5	25	10	26	9	27	9	25	8	22
A05	10	26	10	30	4	30	9	28	4	26	4	25
A06	4	25	6	28	5	22	7	28	8	27	5	22
A07	10	27	10	29	6	29	10	30	6	27	7	27
A08	6	28	5	23	8	22	8	24	10	28	10	26
A09	5	24	6	22	5	23	6	25	5	25	4	26

Table 5.5: Kappa values of the proposed method with and without SS-MEMDBF applied on BCI competition IV dataset 2A.

Subject	Subject specific filtering range (%)		Kappa value with SS-MEMDBF		Kappa value with raw EEG	
	Lower	higher	Training	Evaluation	Training	Evaluation
A01	8	26	0.78	0.86	0.58	0.53
A02	7	26	0.16	0.24	0.25	0.21
A03	7	26	0.7	0.7	0.58	0.54
A04	6	28	0.4	0.68	0.22	0.38
A05	4	29	0.24	0.36	0.18	0.15
A06	6	28	0.22	0.34	0.33	0.28
A07	6	29	0.79	0.66	0.49	0.4
A08	8	22	0.77	0.75	0.77	0.61
A09	5	25	0.81	0.82	0.67	0.63
Average			0.54	0.60	0.45	0.41
p-value					0.0781	0.0039

The results obtained using the proposed methodology are presented in Table 5.5. The Kappa value measures the agreement across the two outcomes (Schlogl et al., 2007). A subject specific filtering range was obtained providing the best Kappa value shown in Table 5.5 for the multi-class problem. With subject specific MEMD based filtering, the average Kappa value across all the nine subjects has improved by 0.08 ($p = 0.0508$) in the evaluation sessions as compared to 0.52 reported using the Riemannian geometry in (Barachant et al., 2012). The average Kappa value of 0.54 has been obtained with 5-fold cross-validation in the training stage. In addition, a maximum improvement of 0.17 was observed for the subject A08 when compared to that reported using Riemannian geometry in (Barachant et al., 2012). Subjects A04 and A09 have improved by 0.15 and 0.14 in Kappa value. There is an improvement in Kappa value with the SS-MEMDBF filtering in eight of the nine subjects considering only the evaluation session compared to that reported using Riemannian geometry in (Barachant et al., 2012).

Table 5.6 shows the classification accuracy when evaluated on BCI competition IV dataset 3 using EMD based filtering and MEMD based filtering. It also shows the comparison with the other BCI competition IV winners (Sardouie and Shamsollahi, 2012; Tangermann et al., 2012). The classification accuracy has been computed for

Table 5.6: Classification accuracies with the proposed method when evaluated on BCI competition IV dataset 3.

Subject	MEMDBF	EMDBF	MEMDBF-SCM	Winner1	Winner2	Winner3	Winner 4
S01	52.7	40.54	54.05	**59.5**	31.1	16.2	23.0
S02	49.31	43.83	46.57	34.3	19.2	31.5	17.8
Average	51.00	42.18	50.31	46.90	25.15	23.85	20.4

subjects S01 and S02. These results are computed using two different features. The SHCM and SCM features have been computed from the enhanced MEG signals. The Columns first and second present the classification accuracy results obtained using the SHCM feature with MEMDBF and EMDBF filtering techniques. The third column gives the results computed using SCM feature as MEMDBF-CSP. The main highlights observed based on the classification accuracy are as follows: (1) The average classification accuracy computed with the EMD based filtering and MEMD based filtering for both subjects gives the minimum standard deviation of 2.33 and 2.4 as compared to other research groups. (2) Subject S02 gives the maximum classification accuracy of 49.31 % which is higher than ($> 5\%$) with EMDBF method and > 15 % with the competition winner. (3) Since the higher classification is achieved in multiclass classification problem using the MEMDBF technique, thus the features are more separable as compared to EMDBF method. These filtering techniques have served as a preprocessing step. It should be noted that no complexity has been introduced at feature extraction and classification steps.

5.8 Comparison with Other Published Results

As this chapter represents the culmination of all the previous studies reported in this thesis it would be pertinent to assess SS-MEMDBF against others in the field who are presenting similar works to assess the effectiveness of the proposed technique. Table 5.7 displays the Kappa values computed with the proposed

Table 5.7: Kappa value comparison with other published results. The best result for each subject is displayed in bold characters.

Subjects	SS-MEMDBF	MDRM (Barachant et al., 2012)	Winner 1 (Ang et al., 2008)	Winner 2 (Guangquan et al., 2008)	Winner 3 (Song, 2008)	Winner 4 (Coyle, 2008)	Winner 5 (Jin, 2008)
A01	**0.86**	0.75	0.68	0.69	0.38	0.46	0.41
A02	0.24	0.37	**0.42**	0.34	0.18	0.25	0.17
A03	0.70	0.66	**0.75**	0.71	0.48	0.65	0.39
A04	**0.68**	0.53	0.48	0.44	0.33	0.31	0.25
A05	0.36	0.29	**0.40**	0.16	0.07	0.12	0.06
A06	**0.34**	0.27	0.27	0.21	0.14	0.07	0.16
A07	0.66	0.56	**0.77**	0.66	0.29	0.00	0.34
A08	0.75	0.58	**0.75**	0.73	0.49	0.46	0.45
A09	**0.82**	0.68	0.61	0.69	0.44	0.42	0.37
Average	**0.60**	0.52	0.57	0.52	0.31	0.30	0.29
p-value		0.0508	0.4844	0.0469	0.0039	0.0078	0.0039

methodology and comparison with other similar works based on p-value computation. The SS-MEMDBF has shown substantial performance improvement in multi-class MI based BCI classification when compared to that reported using Riemannian geometry in (Barachant et al., 2012). The SS-MEMDBF done at preprocessing stage has thus helped to achieve a mean Kappa of 0.60 which is superior to all the results reported so far. The results obtained have been computed using a one versus rest mechanism to classify the MI EEG signals into multiple classes. Moreover, we have exploited the frequency domain information as done by the competition winner.

The top competition winner implemented filter bank CSP (FBCSP) and multiple one-against the rest classifiers (Ang et al., 2008) and reported average Kappa value 0.57 across nine subjects; while the runner-up implemented CSP on bandpass filtered data (between 8-30 Hz) and computed log variance of best eight components as the features and then classified by LDA and Bayesian classifiers and computed average Kappa 0.52 (Guangquan et al., 2008). The third group (Song, 2008) have applied bandpass filtering between 8-25Hz and used a recursive channel elimination for the channels selection. They, thereafter extracted the CSP feature and used ensemble multi-class classifier using three SVM classifiers and computed mean Kappa value 0.31. The fourth group (Coyle, 2008) applied CSP on spectrally filtered neural time-series prediction pre-processing (NTSPP) signals at

pre-processing stage and used the log variance of each filtered channel with a one second sliding window as features and then the best classifier among two variants of support vector machine (SVM) and three variants of LDA was chosen for each subject individually for classification purposes and calculated mean Kappa value 0.30.

There is a significant difference in the methodology followed between competitors and us: we have obtained the enhanced EEG signal using SS-MEMDBF and computed feature set as SCM for each trial across all the nine subjects, then did the classification using Riemannian mean distance. A research group however did study the usage of the Riemannian geometry but their pre-processing approach was very different and has reported mean kappa value of 0.52 (Barachant et al., 2012). Comparing these results, we have achieved highest Kappa value in four subjects and same Kappa value in one subject of the provided nine subjects in the evaluation session. These results clearly show that without the prior knowledge about the non-stationary characteristics of the EEG signals, the SS-MEMDBF method has shown promising performance and thus, has potential to enhance the feature separability when incorporated as a pre-processing method and significantly enhance BCI applications.

5.9 Conclusion

The SS-MEMDBF method was explored with sample covariance as a feature set to enhance the performance of multiple class MI based BCI. The main idea in the proposed method is to provide subject specific MEMD based filtering range in the preprocessing stage reducing the effect of the intra- and inter-subject non-stationarity present in the EEG signals. This preprocessing step provides the enhanced EEG signals from which the extracted feature's distributions have statistically significant differences. Also, mean frequency ranges have been identified when EEG signals are recorded from the same cortical areas across the subjects. The results were obtained in terms of Kappa value and classification accuracy when single trials are

classified. The filtering method is demonstrated to reduce the effect of intrinsic non-stationarity in the EEG signals to some extent. Also, it reduces the noise, and artifacts. Moreover, highly significant performance improvement was obtained in binary class and multi-class classification problems of MI based BCI by enhancing the EEG signals at the pre-processing stage. In the SS-MEMDBF, a selection of one or multiple MIMFs is done when the MIMFs have mean frequencies falling in the frequency range of μ and β rhythms specific to a subject. Further, the enhanced EEG signal obtained has helped to achieve improvement in the Kappa value while classifying EEG signals into left hand, right hand, foot, and tongue MI tasks when compared to raw EEG signals. The Kappa value obtained with the SS-MEMDBF has shown significant improvement in both the training session and the evaluation session across the multiple subjects. Overall the results compare favourably with a research group that used Riemannian geometry and the BCI competition IV dataset 2A winners for the multi-class and binary class classification problems. Although SS-MEMDBF has helped to obtain enhanced feature separability and reduce the error rates due to intrinsic non-stationarities present in EEG signals to a large extent, adaptive classification methods may also be explored to handle the non-stationarities more efficiently which is further discussed in future work section.

The proposed method in this chapter is studied on publicly available BCI competition IV dataset 2A for discrimination of two-class and four-class MI based EEG signals in offline mode. Additionally, the proposed has been evaluated on the BCI competition IV dataset 3 MEG data also. In future, it may be interesting to evaluate the proposed method in an online classification problem in real-time environment using EEG or MEG recording techniques. Additionally, it may be possible to further extend the proposed method through its application on more than four-class. This chapter has addressed contribution (C4) of this thesis and the next chapter seeks to address the fifth (C5) contribution by proposing a novel method built on top of transfer learning when applied to multichannel EEG data in a motor imagery based BCI paradigm.

Chapter 6

Tangent Space based Transfer Learning

6.1 Introduction

The previous chapter investigated the effectiveness of subject specific multivariate empirical mode decomposition based filtering (SS-MEMD) in Riemannian geometry framework when evaluated on multichannel electroencephalogram (EEG) data in a motor imagery (MI) based brain-computer interface (BCI) paradigm. With the method presented in the previous chapter, significant improvements were achieved in terms of classification accuracy of two-class and four class classification problems of MI EEG signals. This chapter will address the last contribution for this thesis (C5) by introducing a novel tangent space based transfer learning classification method which seeks to further improve the classification accuracy.

This chapter concludes by comparing the results obtained by using a combination of the SS-MEMD technique and a new tangent space based transfer learning classification method with the method previously proposed in the last chapter based on a distance formula in Riemannian geometry framework. It also compares the proposed method to other similar work demonstrating a substantial performance

improvement in multi-class MI based BCI classification.

The aims of this chapter are as follow:

1. To study the session-to-session non-stationarity present in the EEG signals;
2. To reduce the long calibration time in BCI systems by proposing a model which can be directly used for evaluating unseen single trials for a subject without any training session data.
3. To automatically identify the subject specific MIMFs based on mean frequency measure;
4. To present the results in terms of classification accuracy when single trials are classified.

6.2 Methods

As discussed previously (cf. 2.8.1), a challenging task is to classify motion intentions into two classes since the recorded EEG signals are highly subject-specific and sensitive to noise and have inherent non-stationarities which are due to changes in the signal properties over time within the session as well as across sessions. They also require long calibration time, which is limiting the use of BCI in patients and healthy people. Every time there is a need to collect numerous training data trials for machine learning methods particularly used in BCI paradigm. Thus it becomes difficult to achieve a stable operation of BCI. A unique approach was already discussed in the study 2 of chapter 4 by combining the training data available from all of the nine subjects. This approach yields improvement but in this work, a novel tangent space based transfer learning method is proposed to further improve the classification performance. This classification technique exploits the tangent space features shared structure between the training data of multiple sessions and subjects instead of combining the training data. A block diagram of the proposed

method is shown in Figure 6.1.

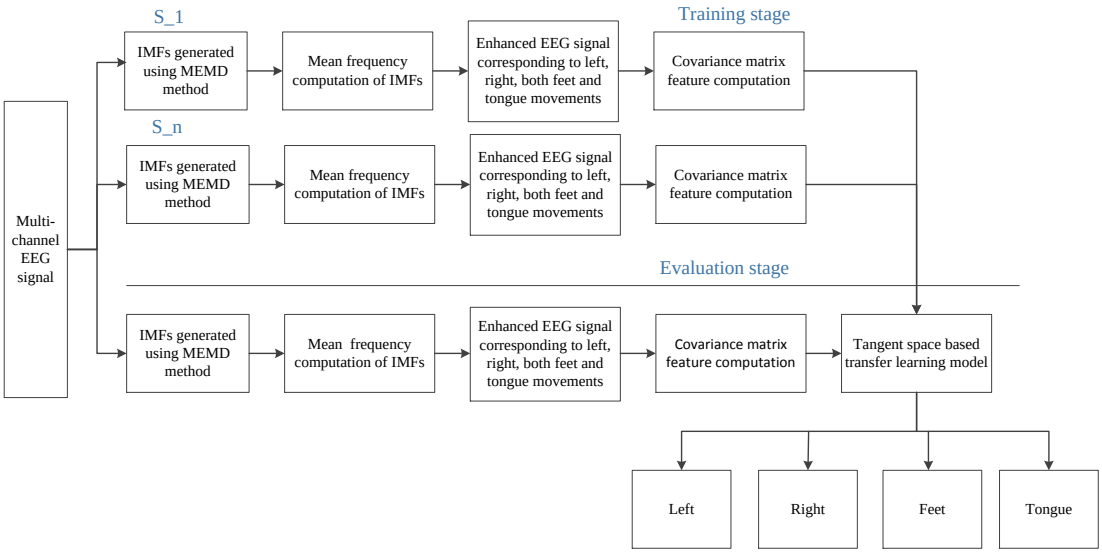


Figure 6.1: A pipeline for the proposed methodology.

6.3 Dataset

In this study, the publicly available BCI competition IV dataset 2A (Brunner et al., 2008) has been used. All of the twenty-two channels have been considered in this study as shown in Figure 6.2. The enhanced EEG signals from all twenty-two channels have been used to extract the sample covariance matrix (cf. 5.5) as a feature set. The feature set contains $n(n+1)/2$ features where n denotes the number of channels. Here, in this study the number of channels are twenty-two, so the feature obtained from the enhanced EEG signals are 253.

6.4 Multivariate Empirical Mode Decomposition

The feet, tongue, left hand, and right hand MI EEG signals were decomposed using the MEMD method as in previous studies (cf. 4.2.3) and was described in more depth in (cf. 2.10.2). The decomposition mechanism of the MEMD method was also explained using single trials of left hand and right hand MI EEG signals. It

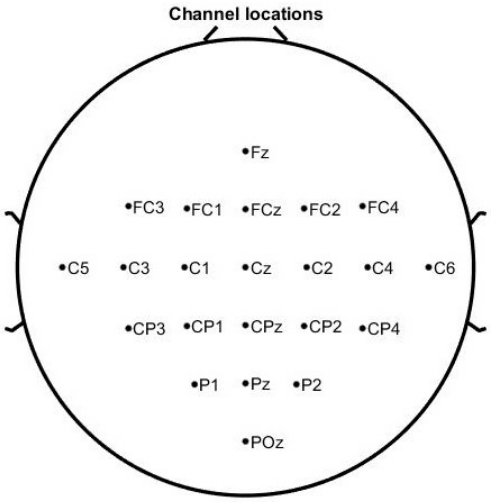


Figure 6.2: Head plot showing all the channels locations

was also shown in Figures 4.3 and 4.4 using single trials from Subject A08T how the trials are decomposed with the MEMD decomposition technique (cf. 5.4).

6.5 Tangent space based transfer learning

This section introduces the classification model used in this chapter. The training session across multiple subjects are indexed $s = \{1, ..., S\}$ and and has n_s^{tr} trials, $Z_s = \{(x_s^i, y_s^i)\}_{i=1}^{n_s^{tr}}$. As discussed in Section 5.6, the tangent space concept in the Riemannian geometry framework and the logarithmic mapping gives the inverse mapping which is defined as,

$$\text{Log}_Q(Q_i) = P_i = ||lower\{Q^{1/2}\text{log}(Q^{-1/2}Q_iQ^{-1/2})Q^{1/2}\}|| \tag{6.5.1}$$

These are features derived from the tangent space and named as tangent space features. Only $n(n+1)/2$ are considered by taking the lower triangular matrix of the provided $n \times n$ features.

$$F = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nn} \end{bmatrix} \tag{6.5.2}$$

Here, $x_s^i \in \mathbb{R}^d$ denotes the features derived from the EEG signals of training session in the subject s during trial i, with d denoting the number of features selected from the feature matrix F. Since the sample covariance matrices (SCM) is symmetric, the lower triangular matrix is being considered for this study giving a total of $d = n(n+1)/2$ features. The x_s^i consists of tangent space features computed at different scalp locations. Variable y_s^i gives the subject's stimulus such as motor imagery task of either the left or right hand imagination in trial i for session s. This approach is applicable for solving two class classification problems. Additionally, it is a linear regression problem with $y_s^i \in \{+1, -1\}$ for all i and s. Based on this assumption, the model is linear with a noise term ν, the linear function model is denoted by,

$$y_s^i = w_s^T x_s^i + \nu \tag{6.5.3}$$

related to each subject/training session s. The parameters w_s shows the weights assigned to the individual features which are further used to evaluate the stimulus for trials in evaluation session of a new subjects s. Given a new EEG signal x for subject/ evaluation session s, the stimulus is evaluated by

$$\hat{y}_s^{i+1} = w_s^T x_s^i \tag{6.5.4}$$

Each subject/training session has a shared structure (Σ, μ) that represents the invariable properties for stimulus prediction. To be specific, (Σ, μ) represents the covariance and mean vectors of features. The divergence of training session model from shared structure of each subject $||w_s - \mu||$ gives the session specific properties of the stimulus prediction (Jayaram et al., 2016). This shared structure is unknown, thus the main goal is to find the shared structure across all the subjects. This is achieved by combining the optimization problem as,

$$\underset{K,\Sigma,\mu}{min} LF(K,\Sigma,\mu;Z,\theta) = \underset{K,\sum,\mu}{min} \frac{1}{\theta}\sum_{1}^{s}||X_s w_s - y_s||^2 + \sum_{1}^{s}\Psi(w_s;\Sigma,\mu) \tag{6.5.5}$$

where d denotes the dimension of each feature vector and $K = [w_1, ..., w_s]^T$ and $Z = \{Z_s\}_{s=1}^{S}$. The w_s may be computed by solving the above optimization problem.

Algorithm 1 Shared structure computation across multiple training subjects using multi-task BCI

1: Input: Z,θ
2: Set : $\{(\Sigma,\mu)\} = \{\mathrm{I}, 0\}$
3: **repeat**
4: Update $w_s = \frac{\left(\frac{1}{\theta}X_s^T y_s \Sigma + \mu\right)}{\left(\frac{1}{\theta}X_s X_s^T \Sigma\right)}$
5: Update Σ using $\hat{\Sigma} = \frac{\sum_{s=1}^{S}\left(w_s-\mu\right)\left(w_s-\mu\right)^T}{Tr\left(\left(w_s-\mu\right)\left(w_s-\mu\right)^T\right)} + \varepsilon\mathrm{I}$
6: Update μ using $\hat{\mu} = \frac{1}{S}\sum_{s=1}^{S} w_s$
7: **until** converge
8: **Output :** $\{(\Sigma,\mu)\}$

Algorithm 1 is used for computing shared structure and w_s across training sessions for all the subjects. More details about the optimization problem to compute w_s can be obtained from Alamgir et al. (2010); Jayaram et al. (2016).

6.6 Results and discussion

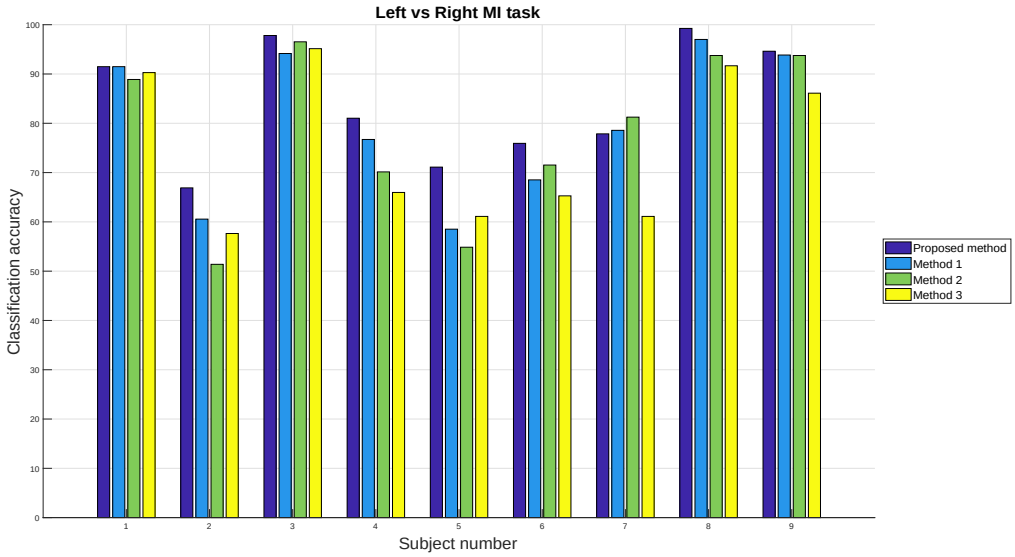
In the training session, a novel tangent space based transfer learning model has been created by exploiting the features obtained from the enhanced EEG signals using all nine subjects. The features share some common information because the feature set is generated when a subject is asked to perform the same motor imagery task. The proposed method exploits the tangent space features shared structure between the training data of multiple subjects. For the computation of classification accuracy (in %) for each subject in the evaluation session, 100% of

A0ST data has been constructing the classification model from all the nine subjects. Furthermore, the tangent space feature has been computed on all unknown test trials of the corresponding evaluation session A0SE on a trial-by-trial basis, where S represents the subject number. These features have been classified using the proposed classification model and assigned a particular class. The proposed method is suitable for two-class classification problems. Now, there are four motor imagery tasks so the total number of combinations obtained are six as discussed later in this thesis. The possible combination are as follow : left vs right (LvR), left vs foot (LvF), left vs tongue (LvT), right vs foot (RvF), right vs tongue (RvT), foot vs tongue (FvT).

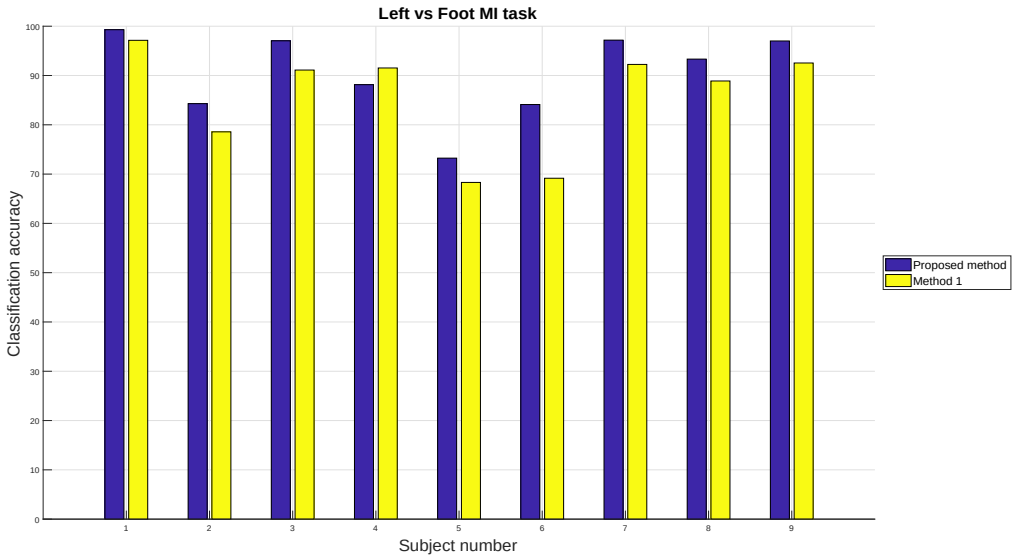
As the MI task began at 2 seconds, the covariance matrix feature has been computed for the EEG signals from the 2.5 second to 4.5 second time-interval of the MI paradigm. The classification accuracy for all the evaluation session in nine subjects has been computed. In this study, the mean frequency measure (cf. eq. 3.2.1) has been calculated to identify the MIMFs from all the obtained MIMFs corresponding to the right hand, left hand, both feet and tongue MI tasks. These MIMFs are identified based on mean frequency measure which provide a major contribution to μ and β rhythms and remaining MIMFs are discarded as noise. These identified MIMFs were summed together to obtain the enhanced EEG signals.

Figures 6.3, 6.4 and 6.5 display the difference between the classification accuracy computed using SS-MEMDBF with Riemannian geometry framework and SS-MEMDBF with tangents space based transfer learning obtained in the evaluation session for all of the six possible binary MI tasks. The performance improvement is demonstrated with bar graphs for all of the nine subjects.

Figure 6.3(a) shows the classification accuracy comparison of the proposed pipeline (SS-MEMDBF with tangents space-based transfer learning) with SS-MEMDBF with Riemannian geometry framework, and other state-of-the-art methods obtained in the evaluation session for the left hand and right hand MI tasks. With the proposed tangent space based transfer learning method, we have achieved the best results

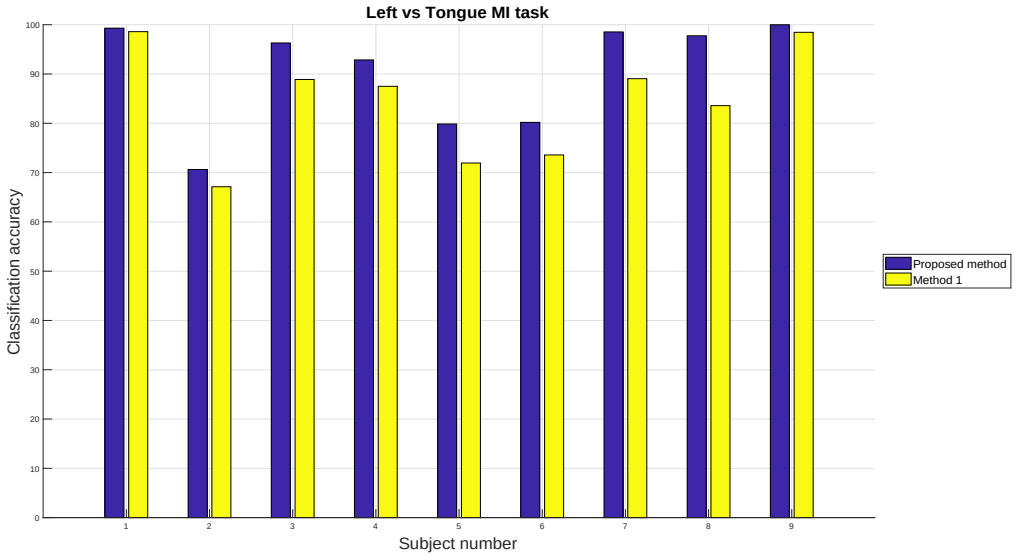


(a) Comparison of classification accuracy for left hand and right hand MI task.

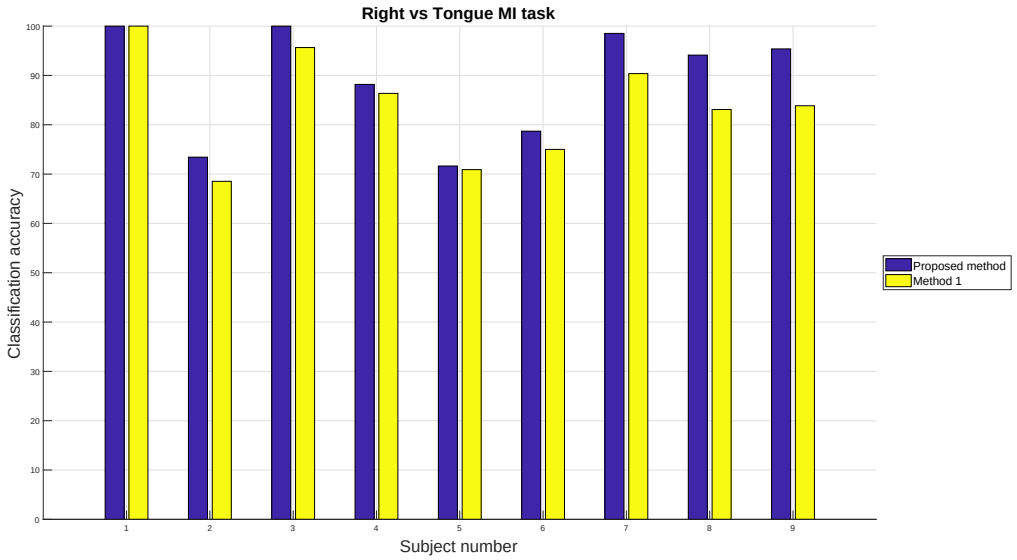


(b) Comparison of classification accuracy for left hand and foot MI task.

Figure 6.3: The bar graph displays the classification accuracy comparison using proposed pipeline with other published results (a) left hand and right hand MI tasks (b) left hand and foot MI tasks obtained in the evaluation session.

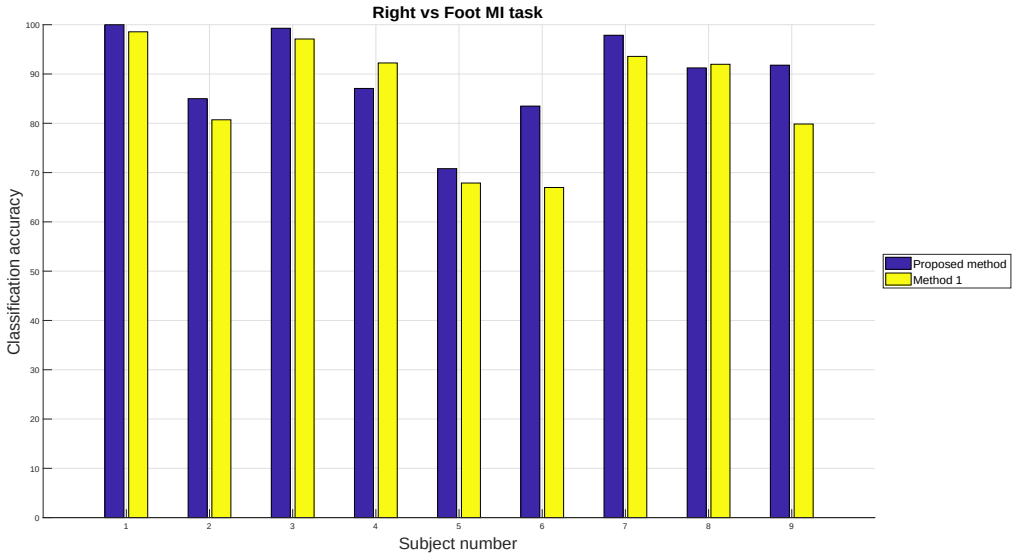


(a) Comparison of classification accuracy for left hand and tongue MI task.

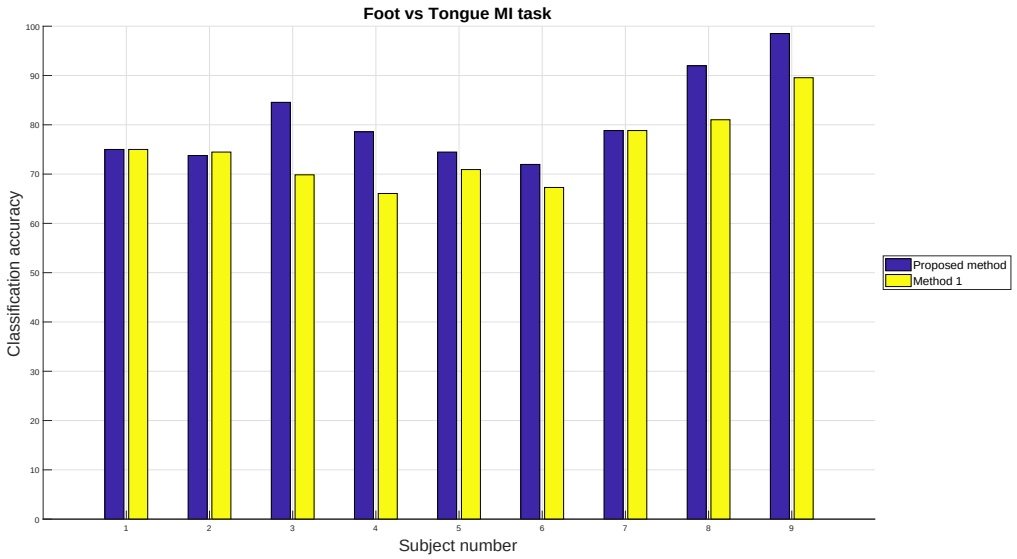


(b) Comparison of classification accuracy for right hand and tongue MI task.

Figure 6.4: The bar graph displays the classification accuracy comparison (a) left hand and tongue MI task (b) right hand and tongue MI task, using proposed pipeline with other state-of-the-art methods in the evaluation session.



(a) Comparison of classification accuracy for right hand and foot MI task.



(b) Comparison of classification accuracy for foot and tongue MI task.

Figure 6.5: The bar graph displays the classification accuracy comparison using proposed pipeline with other published results (a) right hand and foot MI task and (b) foot and tongue MI task in the evaluation session.

Table 6.1: Classification accuracy (in %) for the proposed classification method with one vs one scheme applied on BCI competition IV dataset 2A.

Subject	SS-MEMDBF and proposed method						Comparison with other groups		
	LvR	LvF	LvT	RvF	RvT	FvT	Method-1	Method-2	Method-3
A01	91.49	99.29	99.3	100	100	75	88.89	90.28	90.28
A02	66.9	84.29	70.63	85	73.43	73.76	51.39	54.17	57.64
A03	97.81	97.04	96.3	99.28	100	84.56	96.53	93.75	95.14
A04	81.03	88.14	92.86	87.07	88.18	78.57	70.14	64.58	65.97
A05	71.11	73.24	79.86	70.8	71.64	74.47	54.86	57.64	61.11
A06	75.93	84.11	80.19	83.49	78.7	71.96	71.53	65.28	65.28
A07	77.86	97.18	98.54	97.86	98.52	78.83	81.25	62.5	61.11
A08	99.25	93.33	97.76	91.24	94.12	91.97	93.75	90.97	91.67
A09	94.62	97.01	100	91.79	95.38	98.51	93.75	85.42	86.11
Average	84	90.4	90.6	89.61	88.89	80.85	78.01	73.84	74.92
Std	12.05	8.64	10.83	9.33	11.46	9.1	17.01	15.93	15.43
p-value							0.0273	0.0039	0.0039

when compared with the other state-of-art methods. As discussed earlier, the results obtained with SS-MEMDBF with Riemannian geometry framework are still impressive but few of the subjects have performed badly due to the effect of the non-stationarity present. There was a need to perform transfer learning to handle these issues. The proposed method has helped to overcome non-stationarity to a larger extent by combining two approaches together to form a novel pipeline. The proposed pipeline handles the non-stationarity in the pre-processing stage which is evident by the results reported using SS-MEMD filtering technique in chapter 5 and then applying transfer learning on the tangent space features. It is evident that the results are more impressive when tangent space based transfer learning is compared with Riemannian geometry framework. Seven of the nine subjects have shown improvement but subject A05 has shown exceptional improvement of > 12 % using the proposed methods. The average classification accuracy for the nine subjects has increased by $> 4\%$ in the evaluation session. The bar graph shown in Fig 6.3(b) displays the classification accuracy comparison of SS-MEMDBF with

Riemannian geometry framework and SS-MEMDBF with tangents space based transfer learning obtained in the evaluation session for left hand and foot MI tasks. There is an improvement in classification accuracy for eight of the nine subjects using tangent space based transfer learning method. The average accuracy across the nine subjects has improved by $> 4.9\%$. Subject A06 has shown a maximum improvement of >14% in the evaluation session. Six of the nine subjects have shown an improvement between > 4 % and $< 6\%$.

In the Figure 6.4(a), the classification accuracy comparison using SS-MEMDBF with Riemannian geometry framework and SS-MEMDBF with tangents space based transfer learning obtained in the evaluation session for the left hand and tongue MI tasks have been presented. All of the nine subjects have shown improvement in the classification accuracy using the proposed method. The average classification accuracy has been improved by $> 6\%$. Subject A09 has shown maximum improvement of $> 14\%$ in the classification accuracy.

Figure 6.4(b) shows the classification accuracy comparison using SS-MEMDBF with Riemannian geometry framework and SS-MEMDBF with tangents space based transfer learning obtained in the evaluation session for the right hand and tongue MI tasks. Eight of the nine subjects have shown improvement in the classification accuracy using the proposed method. The average classification accuracy has been improved by $> 5\%$. Subject A09 and A08 have shown significant improvement of $> 11\%$ in the classification accuracy. Subject A07 has shown an improvement of $>$ 8% in terms of classification accuracy.

Figure 6.5(a) similarly shows the classification accuracy comparison using SS-MEMDBF with Riemannian geometry framework and SS-MEMDBF with tangents space based transfer learning obtained in the evaluation session but for the right hand and foot MI tasks. Seven of the nine subjects have shown improvement in classification accuracy using the proposed method. The average classification accuracy has been improved by >4%. The greatest improvement was seen in subject A06 at >16% whilst subject A09 has shown an improvement of almost 12%

in terms of classification accuracy.

Figure 6.5(b) again compares SS-MEMDBF with Riemannian geometry framework and SS-MEMDBF with tangents space based transfer learning obtained in the evaluation session but this time for the foot and tongue MI tasks. All but one subject showed improvement in classification accuracy using the proposed method with an average classification accuracy improvement of $> 6\%$. Subject A03 showed significant improvement of $> 14\%$ with subject A04 improving by $> 12\%$. Subjects A08 and A09 also improved by $> 10\%$ and $> 8\%$ respectively.

6.7 Conclusion

The tangent space based transfer learning method has been explored with sample covariance as a feature set to enhance the performance of two class MI based BCI. The main idea is to provide subject specific MEMD based filtering range in the preprocessing stage reducing the effect of the inter-subject non-stationarity present in the EEG signals and then performing the classification using tangent space based transfer learning method. This preprocessing step enables to achieve the enhanced EEG signals from which the extracted feature's distributions have statistically significant differences. The results were obtained in terms of classification accuracy when single trials are classified in the evaluation session. The filtering method is demonstrated to reduce the effect of intrinsic non-stationarity in the EEG signals to some extent. The proposed method in this chapter along with the previously proposed filtering method in chapter 4 has provided significant performance improvement in two class classification problems of MI based BCI. The classification accuracy obtained with the SS-MEMDBF and proposed method has shown significant improvement in the evaluation session across the multiple subjects. The SS-MEMDBF has thus helped to obtain enhanced feature separability and reduce the error rates due to intrinsic non-stationarities present in EEG signals to a large extent. Additionally, a tangent space based transfer learning method was able to handle the non-stationarities more efficiently. The proposed classifica-

tion method in this chapter was studied on publicly available BCI competition IV dataset 2A for discrimination of two-class MI based EEG signals. In future, it may be interesting to evaluate the proposed method in for multi-class classification problems using EEG or MEG recording techniques.

This chapter has addressed the fifth contribution (C5) by proposing a novel method built on top of transfer learning when applied to multichannel EEG data in a motor imagery based BCI paradigm.

Chapter 7

Conclusions and Recommendations

7.1 Concluding Summary

Non-stationarity is a major issue and is often perceived across sessions and subjects in MEG/EEG-based brain-computer interface (BCI) systems. In this phenomenon, the statistical properties of the recorded brain signals change with time. Due to this issue, the performance of the BCI system is often degraded while using traditional machine learning algorithms which are built on the assumption that the statistical property should remain stationary across the trials, which is often violated. This thesis has addressed this shortcoming through the development of novel and robust single and multichannel filtering techniques for the analysis of both EEG and MEG brain signals in the pre-processing stage leading to improved BCI classification accuracy. These filtering techniques are able to handle the adverse effect of inherent non-stationarity in the brain signals. Another issue is the long calibration time needed to record training data, which limits the usefulness of BCI both for patients and healthy users. There is also a need to collect numerous training data trials for machine learning methods frequently used in BCI paradigms. Thus it becomes difficult to achieve a stable operation of BCI.

In this thesis, single channel and multichannel filtering techniques are studied for the analysis of brain signals. The main work was done based on EEG, but to check the effectiveness of the proposed techniques they were also evaluated on MEG signals. The aim of these filtering techniques was to enhance MEG/EEG in the pre-processing stage which led to improved classification accuracy for use in BCI systems. Also, a tangent space transfer learning approach was proposed which is an important step towards zero training for BCI systems. This has been achieved through:

1. Development of an empirical mode decomposition based filtering method.
2. Development of a multivariate empirical mode decomposition filtering method.
3. Development of a novel tangent space based transfer learning classification model.

To benchmark the performances of the two proposed filtering techniques, they were evaluated on two publically available BCI competition IV EEG datasets and one publically available BCI competition IV MEG dataset. They were also compared against other state-of-the-art research methods.

7.2 Contributions of the Thesis

The research work within the thesis has been peer-reviewed in three international conference papers (Gaur et al., 2015, 2016a,b), and contributed one journal paper (Gaur et al., 2018) with two journal papers due to be submitted along with one book chapter . There were three posters also presented and the results have been reported in this thesis (Gaur et al., 2017, 2016c; Kaushik et al., 2017). As discussed earlier in Section 1.2 of chapter 1, there were five research objectives set, each of which has been addressed in each of the preceding contribution chapters and which will now be discussed in the context of this contributions:

Contribution 1 (C1) - Chapter 3: The EEG signals were enhanced using the single channel filtering method. The features namely, Hjorth and band power features were computed from these enhanced EEG signals. These features were classified into the left hand and right hand motor imagery (MI) using a linear discriminant analysis (LDA) based classification method. This filtering method was also used to classify multi-direction wrist movement MEG signals into right, forward, left and backward classes. The MEG signals were similarly enhanced using the single channel filtering method. The sample entropy feature was extracted from these enhanced MEG signals and the feature set was classified using an ensemble classifier.

Contribution 2 (C2) - Chapter 4: A novel signal processing pipeline was introduced to classify left hand and right hand motor imagery based EEG signals. These signals were first enhanced using multichannel filtering method and then CSP features were extracted from these enhanced EEG signals and further classified using LDA. Additionally, a subject independent classification model was proposed which helped to reduce the training time by using a general model to classify the MI based EEG signals into the left hand and right hand.

Contributions 3 (C3) and 4 (C4) - Chapter 5: An automated classification system was introduced to classify motor imagery based EEG signals into binary and multiple classes of associated MI tasks. A subject specific filtering range has been identified for the motor imagery tasks, namely, left hand, right hand, foot and tongue. The covariance matrix feature was computed as a feature set and classified in the Riemannian geometry framework.

Contribution 5 (C5) - Chapter 6: A classification pipeline was introduced to classify motor imagery based EEG signals. This pipeline was used to solve six combinations of the two-class classification problem for motor imagery tasks, namely, left hand, right hand, foot and tongue. These signals were enhanced in a specific range corresponding to a particular motor imagery task classification problem with the multichannel filtering method. A tangent space feature was computed

from covariance matrix as a feature set and classified using the linear regression method.

7.3 Future Work

Although the techniques presented in this thesis are both novel and robust there are several obvious shortcomings which could be made to improve the work, a short discussion of which will now follow:

1. As discussed in chapter 4, EEG channels have been selected due to their proximity to the motor cortex region in BCI competition IV dataset 2a. In the future, it would be interesting to explore a method to implement a channel selection mechanism based on correlation in the time domain and/or coherence in the frequency domain to achieve better classification accuracy with a minimum number of channels.

2. As discussed in chapter 4, spatial filters have been heuristically selected as also reported in (Lotte and Guan, 2011), they have considered $m = 3$, where m represents the first and the last column vectors of the CSP matrix. Also, Gaur et al. (2016a) have selected $m = 4$ and $m = 5$ spatial filters, while Raza et al. (2015) considered $m = 2$ spatial filters of the CSP matrix. In future, it may be interesting to automatically select the spatial filters for a particular subject. This subject specific selection of spatial filters should help achieve higher classification accuracy because EEG data is highly subject specific.

3. The single channel and multichannel filtering methods may be evaluated in an online MEG-based BCI paradigm for single trials classification problem to validate the performance of a real-time BCI system.

4. Impact analysis on the functional connectivity can be done to see how the connectivity pattern is changing when the single channel, multichannel, and bandpass filtering is done on the EEG/MEG data.

5. As discussed in Chapter 5, n(n+1)/2 features were considered in the Riemannian geometry framework for classification purpose where n denotes the number of channels. Since all twenty two channels were considered, the filtering method may be time-consuming in terms of execution time. There is a need to find a method which can help to find optimum number of channels without compromising with the classification accuracy. One possible solution may be to first apply cross-validation to find when additional features may lead to overfitting. Then, it would be interesting to find their optimum features using a one-way ANOVA by ranking them according to the p-value and then using the identified feature for classification purposes.

6. The results reported in the literature and in this thesis (cf. Chapter 6) demonstrate that transfer learning methods provide a general improvement when compared to other classification methods, but they don't solve the BCI training time problem. The approach followed in this thesis (cf. Chapter 6), all the training session features have been considered for a specific motor imagery task. Future work may involve the selection of subjects based on the significant difference between training data's feature distribution for a specific motor imagery task. In the literature, transfer learning methods are studied where they have exploited the feature structure across all the subjects (Jayaram et al., 2016). There is a need for an automatic method to decide whether the new subject training session feature structure should be considered to update the existing shared structure based on the algorithm discussed in Chapter 6. The final shared structure should be able to classify the new subject evaluation session data. This automatic method will be an important step and may address the zero training time issue to a large extent.

References

Alamgir, M., Grosse-Wentrup, M., Altun, Y., 2010. Multitask learning for brain-computer interfaces. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. pp. 17–24.

Ang, A. Q.-X., Yeong, Y. Q., Ser, W., 2017. Emotion Classification from EEG Signals Using Time-Frequency-DWT Features and ANN. Journal of Computer and Communications 5 (03), 75.

Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., Zhang, H., 2012. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. Frontiers in Neuroscience 6.

Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., Zhang, H., Phua, K. S., Hamadicharef, B., Tee, K. P., 2008. BCI Competition IV Results. `http://bbci.de/competition/iv/results/ds2a/KaiKengAng_desc.pdf`, last accessed 16 December 2017.

Arvaneh, M., Guan, C., Ang, K. K., Quek, C., 2013. EEG data space adaptation to reduce intersession nonstationarity in brain-computer interface. Neural computation 25 (8), 2146–2171.

Bajaj, V., Pachori, R. B., 2012. Classification of seizure and nonseizure EEG signals using empirical mode decomposition. IEEE Transactions on Information Technology in Biomedicine 16 (6), 1135–1142.

Bajaj, V., Pachori, R. B., 2013. Epileptic seizure detection based on the instantaneous area of analytic intrinsic mode functions of EEG signals. Biomedical Engineering Letters 3 (1), 17–21.

Bamdadian, A., Guan, C., Ang, K. K., Xu, J., 2013. Improving session-to-session transfer performance of motor imagery-based BCI using adaptive extreme learning machine. In: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2013. IEEE, pp. 2188–2191.

Barachant, A., Bonnet, S., Congedo, M., Jutten, C., 2012. Multiclass brain–computer interface classification by Riemannian geometry. IEEE Transactions on Biomedical Engineering 59 (4), 920–928.

Bashashati, A., Fatourechi, M., Ward, R. K., Birch, G. E., 2007. A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. Journal of Neural Engineering 4 (2), R32.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., Müller, K.-R., 2011. Single-trial analysis and classification of ERP components - a tutorial. NeuroImage 56 (2), 814–825.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Muller, K.-R., 2008. Optimizing spatial filters for robust EEG single-trial analysis. Signal Processing Magazine, IEEE 25 (1), 41–56.

Brodu, N., Lotte, F., Lécuyer, A., 2011. Comparative study of band-power extraction techniques for motor imagery classification. In: Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2011 IEEE Symposium on. IEEE, pp. 1–6.

Brunner, C., Leeb, R., Müller-Putz, G., Schlögl, A., Pfurtscheller, G., 2008. BCI Competition 2008–Graz data set A. Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology, 136–142.

Carlson, N. R., 2005. Foundations of Physiological Psychology, 6th Edition. Allyn & Bacon, Boston, MA.

Chen, C., Song, W., Zhang, J., Hu, Z., Xu, H., 2010. An adaptive feature extraction method for motor-imagery BCI systems. In: Computational Intelligence and Security (CIS), 2010 International Conference on. IEEE, pp. 275–279.

Coyle, D., 2008. BCI Competition IV Results. `http://bbci.de/competition/iv/results/ds2a/DamienCoyle_desc.pdf`, last accessed 16 December 2017.

Coyle, D., 2009. Neural network based auto association and time-series prediction for biosignal processing in brain-computer interfaces. IEEE Computational Intelligence Magazine 4 (4), 47–59.

Coyle, D., Prasad, G., McGinnity, T. M., 2005. A time-series prediction approach for feature extraction in a brain-computer interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering 13 (4), 461–467.

Coyle, D., Prasad, G., McGinnity, T. M., 2009. Faster self-organizing fuzzy neural network training and a hyperparameter analysis for a brain–computer interface. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 39 (6), 1458–1471.

Davies, S. R., James, C. J., 2013. Novel use of Empirical Mode Decomposition in single-trial classification of motor imagery for use in brain-computer interfaces. In: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 5610–5613.

Davies, S. R., James, C. J., 2014. Using Empirical Mode Decomposition with Spatio-Temporal dynamics to classify single-trial Motor Imagery in BCI. In: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 4631–4634.

Flandrin, P., Rilling, G., Goncalves, P., 2004. Empirical mode decomposition as a filter bank. IEEE Signal Processing Letters 11 (2), 112–114.

Gandhi, V., Arora, V., Behera, L., Prasad, G., Coyle, D., McGinnity, T. M., 2011. Eeg denoising with a recurrent quantum neural network for a brain-computer interface. In: Neural Networks (IJCNN), The 2011 International Joint Conference on. IEEE, pp. 1583–1590.

Gandhi, V., Prasad, G., Coyle, D., Behera, L., McGinnity, T. M., 2014. Quantum

Neural Network-Based EEG Filtering for a Brain–Computer Interface. IEEE Transactions on Neural Networks and Learning Systems 25 (2), 278–288.

Gandhi, V., Prasad, G., Coyle, D., Behera, L., McGinnity, T. M., 2015. Evaluating Quantum Neural Network filtered motor imagery brain-computer interface using multiple classification techniques. Neurocomputing 170, 161–167.

Gao, L., Wang, J., Chen, L., 2013. Event-related desynchronization and synchronization quantification in motor-related EEG by Kolmogorov entropy. Journal of Neural Engineering 10 (3), 036023.

Gaur, P., Bornot, J., Prasad, G., Wang, H., Pachori, R., 2017. Decoding of Multi-direction Wrist Movements Using Multivariate Empirical Mode Decomposition. In: MEG UK 2017, University of Oxford, UK.

Gaur, P., Pachori, R. B., Wang, H., Prasad, G., 2015. An empirical mode decomposition based filtering method for classification of motor-imagery EEG signals for enhancing brain-computer interface. In: International Joint Conference on Neural Networks. pp. 1–7.

Gaur, P., Pachori, R. B., Wang, H., Prasad, G., 2016a. A multivariate empirical mode decomposition based filtering for subject independent BCI. In: 27th Irish Signals and Systems Conference (ISSC). IEEE, pp. 1–7.

Gaur, P., Pachori, R. B., Wang, H., Prasad, G., 2016b. Enhanced Motor Imagery Classification in EEG-BCI using Multivariate EMD based Filtering and CSP Features. In: Proceedings of the Sixth International Brain-Computer Interface Meeting 2016: BCI Past, Present, and Future.

Gaur, P., Pachori, R. B., Wang, H., Prasad, G., 2018. A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry. Expert Systems with Applications 95, 201–211.

Gaur, P., Prasad, G., Wang, H., Pachori, R., 2016c. An MEG based BCI for classification of multi-direction wrist movements using empirical mode decomposition. In: MEG UK 2016 , York, UK.

Gilles, J., 2013. Empirical wavelet transform. IEEE Transactions on Signal Processing 61 (16), 3999–4010.

Grosse, P., Cassidy, M., Brown, P., 2002. EEG–EMG, MEG–EMG and EMG–EMG frequency analysis: physiological principles and clinical applications. Clinical Neurophysiology 113 (10), 1523–1531.

Guangquan, L., Gan, H., Xiangyang, Z., 2008. BCI Competition IV Results. `http://bbci.de/competition/iv/results/ds2a/LiuGuangquan_desc.txt`, last accessed 16 December 2017.

Hajipour Sardouie, S., Shamsollahi, M. B., 2012. Selection of efficient features for discrimination of hand movements from MEG using a BCI competition IV data set. Frontiers in Neuroscience 6, 42.

Herman, P., Prasad, G., McGinnity, T. M., Coyle, D., 2008. Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering 16 (4), 317–326.

Hjorth, B., 1970. EEG analysis based on time domain properties. Electroencephalography and Clinical Neurophysiology 29 (3), 306–310.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., Liu, H. H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. Vol. 454. The Royal Society, pp. 903–995.

Huang, N. E., Wu, M.-L. C., Long, S. R., Shen, S. S., Qu, W., Gloersen, P., Fan, K. L., 2003. A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. In: Proceedings of the royal society of london a: Mathematical, physical and engineering sciences. Vol. 459. The Royal Society, pp. 2317–2345.

Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., Grosse-Wentrup, M., 2016.

Transfer learning in brain-computer interfaces. IEEE Computational Intelligence Magazine 11 (1), 20–31.

Jia, C., Gao, X., Hong, B., Gao, S., 2011. Frequency and phase mixed coding in SSVEP-based brain–computer interface. IEEE Transactions on Biomedical Engineering 58 (1), 200–206.

Jin, W., 2008. BCI Competition IV Results. `http://bbci.de/competition/iv/results/ds2a/JinWu_desc.txt`, last accessed 16 December 2017.

Kanamori, T., Hido, S., Sugiyama, M., 2009. A least-squares approach to direct importance estimation. Journal of Machine Learning Research 10 (Jul), 1391–1445.

Kaushik, G., Gaur, P., Prasad, G., Wang, H., Pachori, R., 2017. An MEG based multi direction wrist movements analysis using empirical mode decomposition and multivariate empirical mode decomposition. In: MEG UK 2017, University of Oxford, UK.

Krauledat, M., Dornhege, G., Blankertz, B., Müller, K.-R., 2007. Robustifying EEG data analysis by removing outliers. Chaos and Complexity Letters 2 (3), 259–274.

Leeb, R., Brunner, C., Müller-Putz, G., Schlögl, A., Pfurtscheller, G., 2008. BCI Competition 2008–Graz data set B. Graz University of Technology, Austria.

Lemm, S., Blankertz, B., Curio, G., Müller, K.-R., 2005. Spatio-spectral filters for improving the classification of single trial EEG. IEEE Transactions on Biomedical Engineering 52 (9), 1541–1548.

Li, Y., Kambara, H., Koike, Y., Sugiyama, M., 2010. Application of covariate shift adaptation techniques in brain–computer interfaces. IEEE Transactions on Biomedical Engineering 57 (6), 1318–1324.

Li, Y., Koike, Y., Sugiyama, M., 2009. A framework of adaptive brain computer interfaces. In: 2009 2nd International Conference on Biomedical Engineering and Informatics. IEEE, pp. 1–5.

Liyanage, S. R., Guan, C., Zhang, H., Ang, K. K., Xu, J., Lee, T. H., 2013. Dynamically weighted ensemble classification for non-stationary EEG processing. Journal of Neural Engineering 10 (3), 036007.

Looney, D., Li, L., Rutkowski, T. M., Mandic, D. P., Cichocki, A., 2008. Ocular artifacts removal from EEG using EMD. In: Advances in Cognitive Neurodynamics ICCN 2007. Springer, pp. 831–835.

Lotte, F., 2014. A tutorial on EEG signal-processing techniques for mental-state recognition in brain–computer interfaces. In: Guide to Brain-Computer Music Interfacing. Springer, pp. 133–161.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., 2007. A review of classification algorithms for EEG-based brain–computer interfaces. Journal of Neural Engineering 4.

Lotte, F., Guan, C., 2011. Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. IEEE Transactions on Biomedical Engineering 58 (2), 355–362.

Lu, H., Plataniotis, K. N., Venetsanopoulos, A. N., 2009. Regularized common spatial patterns with generic learning for EEG signal classification. In: Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE. IEEE, pp. 6599–6602.

Moakher, M., 2005. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. SIAM Journal on Matrix Analysis and Applications 26 (3), 735–747.

Mohammadi, R., Mahloojifar, A., Coyle, D., 2013. Unsupervised short-term covariate shift minimization for self-paced BCI. In: Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2013 IEEE Symposium on. IEEE, pp. 101–106.

Moreno-Torres, J. G., Raeder, T., Alaiz-RodríGuez, R., Chawla, N. V., Herrera, F.,

2012. A unifying view on dataset shift in classification. Pattern Recognition 45 (1), 521–530.

Nicolas-Alonso, L. F., Gomez-Gil, J., 2012. Brain computer interfaces, a review. Sensors 12 (2), 1211–1279.

Norani, N. M., Mansor, W., Khuan, L., 2010. A review of signal processing in brain computer interface system. In: Biomedical Engineering and Sciences (IECBES), 2010 IEEE EMBS Conference on. IEEE, pp. 443–449.

Pachori, R. B., 2008. Discrimination between ictal and seizure-free EEG signals using empirical mode decomposition. Research Letters in Signal Processing 2008, 14.

Park, C., Looney, D., Ahrabian, A., Mandic, D. P., et al., 2013. Classification of motor imagery BCI using multivariate empirical mode decomposition. IEEE Transactions on Neural Systems and Rehabilitation Engineering 21 (1), 10–22.

Park, C., Looney, D., Kidmose, P., Ungstrup, M., Mandic, D. P., 2011. Time-frequency analysis of EEG asymmetry using bivariate empirical mode decomposition. IEEE Transactions on Neural Systems and Rehabilitation Engineering 19 (4), 366–373.

Park, C., Plank, M., Snider, J., Kim, S., Huang, H. C., Gepshtein, S., Coleman, T. P., Poizner, H., 2014. EEG Gamma Band Oscillations Differentiate the Planning of Spatially Directed Movements of the Arm Versus Eye: Multivariate Empirical Mode Decomposition Analysis. IEEE Transactions on Neural Systems and Rehabilitation Engineering 22 (5), 1083–1096.

Pfurtscheller, G., Allison, B. Z., Bauernfeind, G., Brunner, C., Solis Escalante, T., Scherer, R., Zander, T. O., Mueller-Putz, G., Neuper, C., Birbaumer, N., 2010. The hybrid BCI. Frontiers in Neuroscience 4, 3.

Pfurtscheller, G., Brunner, C., Schlögl, A., Da Silva, F. L., 2006. Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. NeuroImage 31 (1), 153–159.

Pfurtscheller, G., Neuper, C., 2001. Motor ımagery and direct brain-computer ınterface. In: Proc. IEEE. v89. pp. 1123–1134.

Pfurtscheller, G., Neuper, C., Flotzinger, D., Pregenzer, M., 1997. EEG-based discrimination between imagination of right and left hand movement. Electroencephalography and Clinical Neurophysiology 103 (6), 642–651.

Phinyomark, A., Phukpattaranont, P., Limsakul, C., 2012. Feature reduction and selection for EMG signal classification. Expert Systems with Applications 39 (8), 7420–7431.

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., Lamantia, A.-S., Mcnamara, J. O., Williams, S. M. (Eds.), 2004. Neuroscience, 3rd Edition. Sinauer Associates, Inc., Sunderland, Massachusetts U.S.A.

Quionero-Candela, J., Sugiyama, M., Schwaighofer, A., Lawrence, N. D., 2009. Dataset shift in machine learning. The MIT Press.

Raza, H., 2016. Adaptive learning for modelling non-stationarity in EEG-based brain-computer interfacing. Ph.D. thesis, Ulster University.

Raza, H., Cecotti, H., Li, Y., Prasad, G., 2015. Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface. Soft Computing, 1–12.

Raza, H., Prasad, G., Li, Y., Cecotti, H., 2014. Covariate shift-adaptation using a transductive learning model for handling non-stationarity in EEG based brain-computer interfaces. In: Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on. IEEE, pp. 230–236.

Rehman, N., Mandic, D. P., 2009. Multivariate empirical mode decomposition. In: Proceedings of The Royal Society of London A: Mathematical, Physical and Engineering Sciences. The Royal Society, p. rspa20090502.

Richman, J. S., Moorman, J. R., 2000. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology 278 (6), H2039–H2049.

Rilling, G., Flandrin, P., Goncalves, P., et al., 2003. On empirical mode decomposition and its algorithms. In: IEEE-EURASIP workshop on nonlinear signal and image processing. Vol. 3. IEEER, Grado, Italy, pp. 8–11.

Samek, W., Vidaurre, C., Müller, K.-R., Kawanabe, M., 2012. Stationary common spatial patterns for brain–computer interfacing. Journal of Neural Engineering 9 (2), 026013.

Sardouie, S. H., Shamsollahi, M. B., 2012. Selection of efficient features for discrimination of hand movements from MEG using a BCI competition IV data set. Frontiers in Neuroscience 6.

Satti, A., Guan, C., Coyle, D., Prasad, G., 2010. A covariate shift minimisation method to alleviate non-stationarity effects for an adaptive brain-computer interface. In: Pattern Recognition (ICPR), 2010 20th International Conference on. IEEE, pp. 105–108.

Schlogl, A., Kronegg, J., Huggins, J. E., Mason, S. G., 2007. Evaluation Criteria for BCI Research. Toward brain-computer interfacing.

Schlögl, A., Vidaurre, C., Müller, K.-R., 2009. Adaptive methods in bci research-an introductory tutorial. In: Brain-Computer Interfaces. Springer, pp. 331–355.

Shahid, S., Prasad, G., 2011. Bispectrum-based feature extraction technique for devising a practical brain–computer interface. Journal of Neural Engineering 8 (2), 025014.

Shahid, S., Sinha, R. K., Prasad, G., 2010. Mu and beta rhythm modulations in motor imagery related post-stroke EEG: a study under BCI framework for post-stroke rehabilitation. BMC Neuroscience 11 (Suppl 1), P127.

Sharma, R., Pachori, R. B., 2015. Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. Expert Systems with Applications 42 (3), 1106–1117.

Sharma, R., Pachori, R. B., Acharya, U. R., 2015a. An integrated index for the identification of focal electroencephalogram signals using discrete wavelet transform and entropy measures. Entropy 17 (8), 5218–5240.

Sharma, R., Pachori, R. B., Acharya, U. R., 2015b. Application of entropy measures on intrinsic mode functions for the automated identification of focal electroencephalogram signals. Entropy 17 (2), 669–691.

Sharma, R., Pachori, R. B., Upadhyay, A., 2017. Automatic sleep stages classification based on iterative filtering of electroencephalogram signals. Neural Computing and Applications, 1–20.

Shenoy, P., Krauledat, M., Blankertz, B., Rao, R. P., Müller, K.-R., 2006. Towards adaptive classification for BCIPart of the 3rd Neuro-IT and Neuroengineering Summer School Tutorial Series. Journal of Neural Engineering 3 (1), R13.

Song, W., 2008. BCI Competition IV Results. `http://bbci.de/competition/iv/results/ds2a/WeiSong_desc.txt`, last accessed 16 December 2017.

Song, Y., Liò, P., et al., 2010. A new approach for epileptic seizure detection: sample entropy based feature extraction and extreme learning machine. Journal of Biomedical Science and Engineering 3 (06), 556.

Sugiyama, M., 2012. Learning under non-stationarity: Covariate shift adaptation by importance weighting. In: Handbook of Computational Statistics. Springer, pp. 927–952.

Sugiyama, M., Krauledat, M., MÃžller, K.-R., 2007. Covariate shift adaptation by importance weighted cross validation. Journal of Machine Learning Research 8 (May), 985–1005.

Sugiyama, M., Nakajima, S., Kashima, H., Buenau, P. V., Kawanabe, M., 2008. Direct importance estimation with model selection and its application to covariate shift adaptation. In: Advances in neural information processing systems. pp. 1433–1440.

Sun, S., Zhang, C., 2006. Adaptive feature extraction for EEG signal classification. Medical and Biological Engineering and Computing 44 (10), 931–935.

Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., Leeb, R., Mehring, C., Miller, K. J., Müller-Putz, G. R., et al., 2012. Review of the BCI competition IV. Frontiers in Neuroscience 6.

Tomioka, R., Müller, K.-R., 2010. A regularized discriminative framework for EEG analysis with application to brain–computer interface. NeuroImage 49 (1), 415–432.

Tuzel, O., Porikli, F., Meer, P., 2008. Pedestrian detection via classification on riemannian manifolds. IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (10), 1713–1727.

ur Rehman, N., Park, C., Huang, N. E., Mandic, D. P., 2013. EMD via MEMD: multivariate noise-aided computation of standard EMD. Advances in Adaptive Data Analysis 5 (02), 1350007.

Vidaurre, C., Blankertz, B., 2010. Towards a cure for BCI illiteracy. Brain topography 23 (2), 194–198.

Vidaurre, C., Cabeza, R., Scherer, R., Pfurtscheller, G., et al., 2007. Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces. IEEE Transactions on Biomedical Engineering 54 (3), 550–556.

Vidaurre, C., Kawanabe, M., Von Bunau, P., Blankertz, B., Muller, K., 2011. Toward unsupervised adaptation of LDA for brain–computer interfaces. IEEE Transactions on Biomedical Engineering 58 (3), 587–597.

Vidaurre, C., Schlögl, A., Blankertz, B., Kawanabe, M., Müller, K.-R., 2008. Unsupervised adaptation of the LDA classifier for brain–computer interfaces. In: Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course. Vol. 2008. Citeseer, pp. 122–127.

Von Bünau, P., Meinecke, F. C., Király, F. C., Müller, K.-R., 2009. Finding stationary subspaces in multivariate time series. Physical Review Letters 103 (21), 214101.

Von Bünau, P., Meinecke, F. C., Scholler, S., Müller, K.-R., 2010. Finding stationary brain sources in EEG data. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. IEEE, pp. 2810–2813.

Waldert, S., Mehring, C., Preissi, H., Braun, C., 2008a. BCI Competition 2008–dataset 3. Brain Machine Interfacing Initiative, Albert-Ludwigs-University Freiburg, the Bernstein Center for Computational Neuroscience Freiburg and the Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen.

Waldert, S., Preissl, H., Demandt, E., Braun, C., Birbaumer, N., Aertsen, A., Mehring, C., 2008b. Hand movement direction decoded from MEG and EEG. Journal of Neuroscience 28 (4), 1000–1008.

Wang, L., Xu, G., Wang, J., Yang, S., Guo, M., Yan, W., 2012. Motor imagery BCI research based on sample entropy and SVM. In: Electromagnetic Field Problems and Applications (ICEF), 2012 Sixth International Conference on. IEEE, pp. 1–4.

Wojcikiewicz, W., Vidaurre, C., Kawanabe, M., 2011. Stationary common spatial patterns: towards robust classification of non-stationary eeg signals. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 577–580.

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., Donchin, E., Quatrano, L. A., Robinson, C. J., Vaughan, T. M., et al., 2000. Brain-computer interface technology: a review of the first international meeting. IEEE Transactions on Rehabilitation Engineering 8 (2), 164–173.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., Vaughan, T. M., 2002. Brain–computer interfaces for communication and control. Clinical Neurophysiology 113 (6), 767–791.

Wu, Z., Huang, N. E., 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in Adaptive Data Analysis 1 (01), 1–41.

Zhang, H., Yang, H., Guan, C., 2013. Bayesian learning for spatial filtering in an EEG-based brain–computer interface. IEEE Transactions on Neural Networks and Learning Systems 24 (7), 1049–1060.