MENTAL TASK CLASSIFICATION UTILIZING INTERCHANNEL RELATIONSHIP OF EEG SIGNAL BASED ON WAVELET AND EMPIRICAL MODE DECOMPOSTITION

by

Md. Mostafizur Rahman

MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

Department of Electrical and Electronic Engineering

BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

October 2017

The thesis entitled **"MENTAL TASK CLASSIFICATION UTILIZING INTERCHANNEL RELATIONSHIP OF EEG SIGNAL BASED ON WAVELET AND EMPIRICAL MODE DECOMPOSITION"** submitted by Md. Mostafizur Rahman, Student No.: 0413062209, Session: April, 2013 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING on October 07, 2017.

BOARD OF EXAMINERS

1. ____

(Dr. Shaikh Anowarul Fattah) **Chairman**
Professor (Supervisor)
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka - 1000, Bangladesh.

2. ____

(Dr. Quazi Deen Mohd Khosru) **Member**
Professor and Head (Ex-officio)
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka - 1000, Bangladesh.

3. ____

(Dr. Mohammed Imamul Hassan Bhuiyan) **Member**
Professor
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka - 1000, Bangladesh.

4. ____

(Dr. Mohammad Rakibul Islam) **Member**
Professor (External)
Department of Electrical and Electronic Engineering
Islamic University of Technology (IUT), Bangladesh
Board Bazar, Gazipur-1704, Bangladesh.

CANDIDATE'S DECLARATION

I, do, hereby declare that neither this thesis nor any part of it has been submitted elsewhere for the award of any degree or diploma.

Signature of the Candidate

Md. Mostafizur Rahman

Dedication

To my parents.

Acknowledgments

I would like to express my unbound respect and thankfulness to my supervisor Dr. Shaikh Anowarul Fattah for his proper guidance and support during the span of this research. It was him who motivated me to incite my own aptitude for knowing the unknown, encouraged me to the utmost and gave rise to the confidence in me that in an underdeveloped country like ours; it is possible to conduct research of superior quality. He not only acted as my supervisor but also helped me take important decisions of my academic career. Among many things I learned from Dr. Fattah, perseverance and constant hard work are the keys to open doors of solutions for several important issues in the area of mental task classification. I believe our work advanced the state of the art of the mental task classification problem. I also want to thank him for affording so much time for me in exploring new areas of my research and new ideas and improving the writing of this dissertation. Then, I would also like to thank the rest of the members of my thesis committee: Prof. Dr. Quazi Deen Mohd Khosru, Prof. Dr. Mohammed Imamul Hassan Bhuiyan, and Prof. Dr. Mohammad Rakibul Islam, for their encouragement and insightful comments. I would like to thank the Head of the department of Electrical and Electronic Engineering for allowing me to use the lab facilities, which contributed greatly in completing the work in time. I wish to give special thanks to Dr. Celia Shahnaz, for providing inspiration and guidance to walk the right path of research to success during the whole span of my research life in BUET. And most importantly, I wish to thank my parents, without whose prayers and constant support, I could never reach this stage of my life.

Abstract

In view of recent increase of brain computer interface (BCI) based applications, the importance of efficient classification of various mental tasks has increased prodigiously now-a-days. For effective classification purpose, efficient feature extraction scheme is necessary. Most of the reported algorithms are performed on the EEG signals or the processed EEG signals taken from various channels while the inter-channel relationship has not been utilized. Depending on the nature of the mental tasks, different spatial locations of brain become more actuated compare to other locations. It is expected that the correlation obtained from different combination of channels will be different for different mental tasks which can be exploited to extract distinctive feature. To corroborate this idea, in the proposed method, a feature extraction scheme based on cross-correlation among data (or decomposed data) obtained from various channels is proposed. Instead of directly utilizing EEG signal, various decomposition techniques, such as empirical mode decomposition (EMD), spectral band division, wavelet decomposition (WD) and wavelet packet decomposition (WPD) are employed on a test EEG signal obtained from a channel. Different well defined narrow frequency bands , corresponding to state of vigilance, are also investigated for feature extraction. Since EEG is a non-stationary signal, EMD, WD and WPD have the potential to perform better than the conventional time-frequency analysis method. Correlation coefficients are extracted from inter-channel pre-processed EEG signal. At the same time, different statistical features of decomposed EEG signals are also obtained. Finally, the feature matrix is formed utilizing inter-channel features and intra-channel features (statistical features) of the decomposed EEG signals. Different kernels of support vector machine (SVM) classifier are used to carry out classification result. For the purpose of demonstrating classification performance, ten different combinations of five different mental tasks, namely geometrical figure rotation, mathematical multiplication, mental letter composing, visual counting, base-line resting obtained from a publicly available dataset are utilized. It is found that the proposed scheme can classify mental tasks with a very high level of accuracy compared to some existing methods.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Classification of various mental tasks plays an essential role in various applications of brain-computer interface (BCI). Mental task is a specific type of human intentions or thoughts based task in which subjects are instructed to imagine themselves performing a specific arithmetic task (such as mathematical multiplication or counting task) or a particular motor imagery (MI) task (such as a hand or foot movement) without an overt motor output. There are various acquisition techniques for capturing brain activities while performing mental tasks in BCI system. Among these techniques, electroencephalogram (EEG) is one of the most promising tools, which is widely used now-a-days in the study of brain science, neural engineering and rehabilitation [1]. It is the most studied measure of potential for noninvasive BCI designs, mainly due to its excellent temporal resolution, non-invasiveness, usability, and low set-up costs [2–4]. In view of providing a direct communication interface between a brain and an external device, BCIs utilize various channels placed in the skull. Brain activities produce electrical signals detectable on the scalp, or within the brain. BCIs translate these signals into outputs that allow total lock-in patients suffering from brain or spinal cord injury to communicate without the participation of peripheral nerves and muscles via thoughts alone [5], [6]. One of the main challenges of BCI systems is to correctly and efficiently identify different EEG signals corresponding to different mental tasks using appropriate classification schemes. For the purpose of effectively classifying various mental tasks corresponding to a BCI system, in general, features are extracted from EEG signals obtained from various channels. Therefore, distinctive feature extraction utilizing EEG signal corresponding to various mental tasks can make the classification process more accurate and faster.

In this Chapter, a brief description of various feature extraction methods of mental task classification utilizing EEG signal is presented. It begins with a brief introduction to the acquisition technique of EEG signal. This Chapter presents the motivation of the thesis by providing the past and current research scenarios in the use of various types of feature extraction and classification techniques. The objectives and organization of the thesis are finally presented at the end.

1.1 EEG

An EEG is a process used to evaluate the electrical activity in the brain. Brain cells communicate with each other through electrical impulses. An EEG can be used to help detecting this activity. During EEG recordings, small sensors are attached to the scalp to pick up the electrical signals produced when brain cells send messages to each other. These signals are recorded by a machine and can be utilized to establish communication between man and machine. Since this recording process is non-invasive i.e. the electrode only picks up electric signal from the brain and does not affect the brain. Therefore, this process is totally painless and harmless. Despite limited spatial resolution, EEG continues to be a valuable tool for research and diagnosis, especially when millisecond-range temporal resolution is required.

1.1.1 Source of EEG Signal

EEG is a graphic representation of the difference in voltage between two different cerebral locations plotted over time. The scalp EEG signal generated by cerebral neurons is modified by electrical conductive properties of the tissues between the electrical source and the recording electrode on the scalp, conductive properties of the electrode itself, as well as the orientation of the cortical generator to the recording electrode. Because of the process of current flow through the tissues between the electrical generator and the recording electrode which is known as volume conduction, EEG provides a two-dimensional projection of our brain. It detects the summed ionic currents of thousands of pyramidal neurons beneath each of the 16 and 25 individual macro electrodes, and reports them as voltage differences across low resistance extracellular space. Specifically, the potentials recorded by the macro-electrodes on the skin of the skull are primarily generated by extracellular current

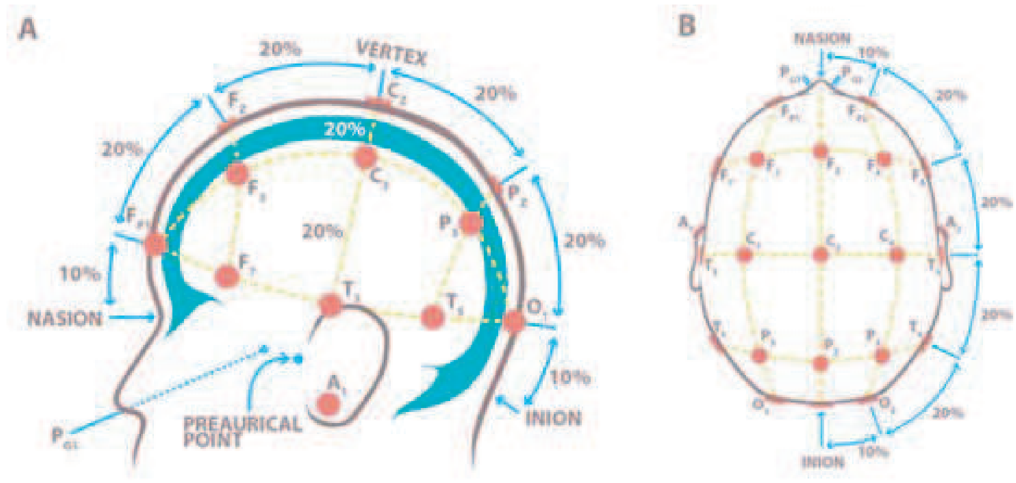

Figure 1.1: EEG electrodes position on the scalp in 10-20 EEG recording system

flow of synaptic potentials in pyramidal cells. Action potentials of the neurons are usually asynchronous and too fast-moving to generate detectable potentials on the skin's surface. As a result, brain cells other than pyramidal neurons such as interneurons and glial cells make relatively little contribution to skin potentials because, unlike pyramidal neurons, these cells are neither oriented in parallel to one another nor do their dendrites run perpendicular to the cortical surface. In contrast, pyramidal neurons run parallel to one another with large dendritic branches that run perpendicular to the cortical surface. Since voltage fields fall off with the square of distance, activity from deep sources is more difficult to detect than currents near the skull. The EEG waves obtained from the scalp electrodes show oscillations at different frequencies. Such oscillations at a variety of frequencies are associated with different states of brain functioning involving different parts of our brain. As a result, such oscillations depict synchronized activity over different networks of neurons which are known as neuronal networks. From such neuronal networks some of these oscillations are understood, while many others are not.

1.1.2 10-20 Standard EEG System

The international 10-20 system of electrode placement is the most widely used method to describe the location of scalp electrodes during an EEG recording or experiment. The 10-20 system is based on the relationship between the location of an electrode and the underlying area of cerebral cortex. Each site has a letter (to identify the lobe) and a number or another letter to identify the hemisphere location. The positions of the electrodes of the 10-20 system are shown in Fig. 1.1 [7]. This method was developed to ensure standardized reproducibility so that a subjects studies could be compared over time and subjects could be compared to each other. The letters F, T, C, P and O stand for frontal, temporal, central, parietal, and occipital lobes, respectively. Even numbers (2, 4, 6, and 8) refer to the right hemisphere and odd numbers (1, 3, 5 and 7) refer to the left hemisphere. “Z” refers to an electrode placed on the mid line. The smaller the number, the closer the position to the mid line. “Fp” stands for Front polar. Two anatomical land marks are used for the essential positioning of the EEG electrodes: first, the nasion which is the point between the forehead and the nose; second, the inion which is the lowest point of the skull from the back of the head and is normally indicated by a prominent bump. The “10” and “20” (10-20 system) refer to the 10% and 20% inter electrode distance. When recording a more detailed EEG with more electrodes, extra electrodes are added utilizing the spaces in-between the existing 10-20 system. This new electrode-naming-system is more complicated giving rise to the Modified Combinatorial Nomenclature (MCN). This MCN system uses 1, 3, 5, 7, 9 for the left hemisphere which represents 10%, 20%, 30%, 40%, 50% of the inion-to-nasion distance respectively. 2, 4, 6, 8, 10 are used to represent the right hemisphere. The introduction of extra letters allows the naming of extra electrode sites. These new letters do not necessarily refer to an area on the underlying cerebral cortex.

1.2 Literature Review

EEG signal is used extensively now-a-days by the researchers to handle different applications of BCI. EEG-based BCI systems employ electrical activity of brain to classify different EEG signals corresponding to various mental tasks precisely. Most popular way to classify the signals effectively is to acquire discriminative features from that signal. As a matter of fact, different schemes to extract distinctive features are available in literature. Several researchers concentrate in various MI task classification which is a special type of mental tasks. For example, in [8], a bayesian framework is proposed in order to find frequency band which can substantially segregate the feature vectors corresponding to two classes of MI tasks (right hand or right foot movement). However, the method offers moderate classification performance. With the ease of availability of high quality low cost EEG leads, one major concern now-a-days is the huge amount of data to be handled in case of multi-lead EEG signal analysis. For the purpose of EEG channel reduction for MI task, various types of spatial filters are widely employed, where regularized parameters need to be chosen manually. Choosing lesser pairs of channels reduce feature size and computation time effectively. In [9], the task of channel reduction is performed by using sparse spatial filter optimization method, where manual intervention is required for obtaining some parameters. In [10], the common spatial pattern (CSP) with generic learning is proposed for EEG channel reduction where optimal selection of regularized parameters needs further investigation.

Besides motor imagery type mental tasks, several researchers concentrate in other types of mental tasks, such as geometrical figure rotation, mathematical multiplication, mental letter composing, visual counting, and baseline-resting. For example, in [11], spectral power and asymmetry ratio based feature extraction scheme is proposed where an additional band ($24-37$ Hz) is used along with conventional lower spectral bands namely delta (<4 Hz), theta ($4-7$ Hz), alpha ($8-13$ Hz) and beta ($14-20$ Hz) for the classification of these five mental tasks. This method offers comparatively satisfactory classification performance but lacks consistency for all cases. In [12], similar feature extraction scheme used in [11] is proposed, however, the difference is that it utilizes an additional high frequency band ($40-100$ Hz) to obtain those features. In [13], a dictionary consisting of power spectral density and common spatial pattern (CSP) algorithm is introduced to classify various mental tasks.

Autoregressive (AR) model can be used to extract features for classification of mental tasks. For example, in [14], a feature extraction scheme based on AR modelling is proposed where sixth order AR system is considered to extract feature. Moreover, in [15], multivariate AR models are taken into consideration and four different representations of AR coefficients are tested to classify mental task. In [16], feature extraction scheme based on sparse AR model are investigated, which involves complex computation to exclude AR coefficients that are useless in the prediction stage. In the feature extraction for EEG signals, [17] extend the usual AR models for feature extraction. The extension model is an AR with exogenous input (ARX) model for combined filtering and feature extraction. In [18], a feature extraction method based on generalized Higuchi fractal dimension spectrum along with AR parameters is proposed.

Wavelet transform is also widely used in extracting EEG features [19], [20] in which EEG signals are decomposed by wavelet transform to calculate approximation and detail coefficients. Approximate entropy (ApEn) is a statistical parameter that measures the predictability of the current amplitude values of a physiological signal based on its previous amplitude values. A novel feature extraction method based on multi-wavelet transform and ApEn is proposed by [21]. The proposed method uses ApEn features derived from multiwavelet transform and combines with an artificial neural network to classify the EEG signals. Multiple kernel learning SVM based classification scheme is investigated in [22], where wavelet packet entropy and Granger causality is used for feature extraction.

Considering the non-stationary nature of EEG signal, empirical mode decomposition (EMD) is also applied for EEG analysis in different applications [23], [24], [25] . Recently in [23], the potential of EMD analysis in MI data handling is discussed with a note that relevant features must be carefully selected for getting better accuracy. In [24], empirical mode decomposition based classification method is proposed where feature selection method is utilized for better classification performance. Moreover, MI EEG signal classification scheme based on entropy of intrinsic mode function is reported in [25].

In [26], stockwell transform based algorithm is proposed and mean square root of standard deviation of signal after transformation is utilized as distinctive feature for mental task classification. A parametric feature extraction and classification strategy for brain-computer interfacing [27] use fast Fourier transform as feature extraction method. In [6], cross-correlation based feature extraction scheme is introduced where cross correlation is computed between two channels keeping one of them always a fixed reference channel. Since one channel is kept fixed, the effect of considering cross correlation between all channels on overall feature quality has not been investigated. Moreover, unique choice of a fixed channel depends on various reasoning.

1.3 Motivation

In EEG based BCI system, EEG recording is needed to capture the event of performing mental task, which can be used for further diagnosis and classification. EEG signals vary in time and frequency domain simultaneously, so instead of directly utilizing raw EEG data, time-frequency analysis can extract time-frequency information more precisely. Since, EEG is a non-stationary signal, EMD, wavelet decomposition (WD), wavelet packet decomposition (WPD) have the potential to perform better than the conventional time-frequency analysis method [28–31]. However, most of the reported algorithms are performed on the EEG signals or the processed EEG signals taken from various channels while the inter-channel relationship is ignored. It is considered that for different types of tasks, different channels corresponding to different parts of the brain are actuated. Measuring inter-channel relationship in some efficient spectro-temporal domains may play a significant role to cover the spatial and temporal relationship between different channels during various mental tasks. Thus, development of a proficient method capable of detecting and classifying different types of mental tasks utilizing the inter-channel relationship is still undiscovered and exploring this is the main motivation of this research.

1.4 Objectives and Scope

The objectives and scope of this thesis are:

1. To quantify inter-channel relationship in terms of statistical measure like correlation coefficient.
2. To obtain different statistical features in EMD, spectral band and wavelet domain.
3. To derive an efficient feature extraction scheme obtained by measuring inter-channel relationship in EMD, spectral band and wavelet domain.
4. To perform test validation of the proposed method on a publicly available mental task EEG dataset using different classifiers.

The outcome of this thesis is an efficient method of mental task classification from EEG signal exploiting inter-channel relationship in EMD, spectral band and wavelet domain. This can be utilized to develop EEG-based BCI systems in assisting, augmenting, and repairing human cognitive or sensory-motor functions, which play a significant role in the treatment of autism, depression, mental disorder etc.

1.5 Organization of the Thesis

In the first Chapter, a brief introduction to BCI technology and its acquisition technique utilizing EEG signal is presented. Moreover, it presents the motivation and objectives of the thesis by providing the past and current research scenarios in the use of various types of feature extraction and classification techniques. The rest of the thesis is organized as follows

In Chapter 2, a feature extraction scheme based on inter-channel relationship of intrinsic mode function of EEG signal corresponding to various mental tasks is proposed. The EMD technique is employed on a test EEG signal obtained from a channel, which provides a number of intrinsic mode functions (IMFs) and correlation coefficient extracted from inter-channel IMF data, referred to as inter-IMFCC in this paper, is computed for the selected IMF. At the same time, different statistical features are also obtained from each IMF. Finally, the feature matrix is formed utilizing inter-channel features (inter-IMFCC) and intra-channel features (statistical features) of the selected IMFs of EEG signal. Different kernels of support vector machine (SVM) classifier are used to carry out classification result. For the purpose of demonstrating classification performance, ten different combinations of five different mental tasks namely geometrical figure rotation, mathematical multiplication, mental letter composing, visual counting, and baseline-resting obtained from a publicly available dataset is utilized.

In Chapter 3, mental task classification scheme utilizing correlation coefficient extracted from inter-channel band limited signals is presented. Instead of directly utilizing EEG signal, different frequency bands , namely delta (<4 Hz), theta ($4-7$ Hz), alpha ($8-13$ Hz), beta ($14-20$ Hz), gamma ($24-37$ Hz) and $40-100Hz$ band are investigated to extract correlation coefficient from inter-channel data of a selected band, which is referred to as inter-SBCC. Moreover, different intra-channel statistical features are also obtained from each band-limited signal. Comparative performance analysis is also presented between proposed method and few recent papers from the literature using the same dataset and same classifier.

In Chapter 4, an efficient scheme of extracting features from EEG signal is proposed for mental task classification based on inter-channel relationship in wavelet domain. It is shown that use of wavelet domain inter-channel relationship can drastically improve the classification performance obtained by conventional wavelet statistics. Both multi-level wavelet decomposition and node reconstruction are utilized for proposed inter-channel correlation features and intra-channel statistical features extraction. Detail experimental results are presented considering the same dataset and same classifier.

Chapter 5 summarizes the outcome of this thesis with some concluding remarks and possible future works.

Chapter 2

Mental Task Classification Scheme Utilizing Correlation Coefficient Extracted from Inter-channel Intrinsic Mode Function

In this Chapter, a feature extraction scheme based on inter-channel relationship of intrinsic mode function (IMF) of EEG signal corresponding to various mental tasks is proposed. Instead of directly dealing with temporal EEG data, it is expected that features extracted from decomposed EEG data will provide more consistent characteristics . In particular, in this Chapter, widely used empirical mode decomposition (EMD) is utilized to obtain IMFs and first three IMFs are selected. Each IMF is utilized to compute the correlation coefficient from inter-channel IMF data, referred to as inter-IMFCC. Moreover, in the proposed method, unlike [6], only the channel information of the test trial is utilized to extract correlation coefficient and no previously defined reference signal is required for that purpose. It is also observed whether the classification accuracy improves if different statistical features obtained from respective IMF are used along with inter-IMFCC. SVM classifier is used to carry out classification process. The effect of the variation of number of channels and that of using different kernels are investigated. Simulation results are reported for a publicly available EEG dataset on various mental tasks. It is observed that the proposed algorithm can classify various mental tasks with a satisfactory level of performance. The detail results of this Chapter is reported in [32].

2.1 Data Acquisition

A widely used EEG data set collected by Keirn and Aunon is utilized [33] in this Chapter. EEG signals are acquired from the locations C3, C4, P3, P4, O1, and O2 which are denoted as the $10-20$ international system of electrode placement. Measurements are made considering A1 and A2 as reference. Data are band pass filtered using an analog filter with band limit of $0.1-100$ Hz and sampled at 250 Hz with 12 bit quantizer. The recording is carried out for ten seconds during each session. EEG signals from seven subjects performing five different mental tasks, namely geometrical figure rotation (R), mathematical multiplication (M), mental letter composing (L), visual counting (C), and baseline-resting (B) are investigated. For notational convenience, hereafter, each task is abbreviated with an alphabet as shown in the parentheses. However, data obtained from three subjects contain fewer than ten sessions or have some recording errors. Hence, like some other existing research works [11], in this Chapter, data from four subjects, each having ten or more sessions, are taken into consideration.

For the purpose of analysis of each ten second session, a number of frames with shorter time interval are investigated as EEG signal is assumed to be non-stationary. In this case, one second frame duration is considered with 0.5 second frame shift (i.e. 50% overlap between successive frames) [12], which provides a reasonable number of samples (250 samples) in each frame.

2.2 Proposed Method

The proposed mental task classification scheme can be divided into four major steps: empirical mode decomposition, inter-channel relation, feature extraction and classification. These steps are described in detail in the following subsections.

2.2.1 Empirical mode decomposition

Due to random nature of recordings of EEG data, it is very difficult to obtain discriminative characteristic from the time domain EEG data. Therefore, instead of directly utilizing EEG data, it may be easier to extract distinctive characteristic if decomposition is imposed on EEG data. Empirical mode decomposition is found very effective as it decomposes the signals in particular patterns preserving the originality of the signal. EMD is intuitive and adaptive, with intrinsic mode functions (IMF) directly derived from the signal under test without changing their domains. Moreover, each IMF contains information about how the frequency of the original signal changes in time. In Fig. 2.1, a sample EEG signal and it's four IMFs obtained from counting task is plotted. It is observed that the four IMFs, obtained after employing EMD on the test EEG signal, are lesser irregular and complex in nature than the original signal and have particular patterns. As a result, it is expected that classification performance will improve if IMFs are utilized to obtain distinctive features instead of main signal.

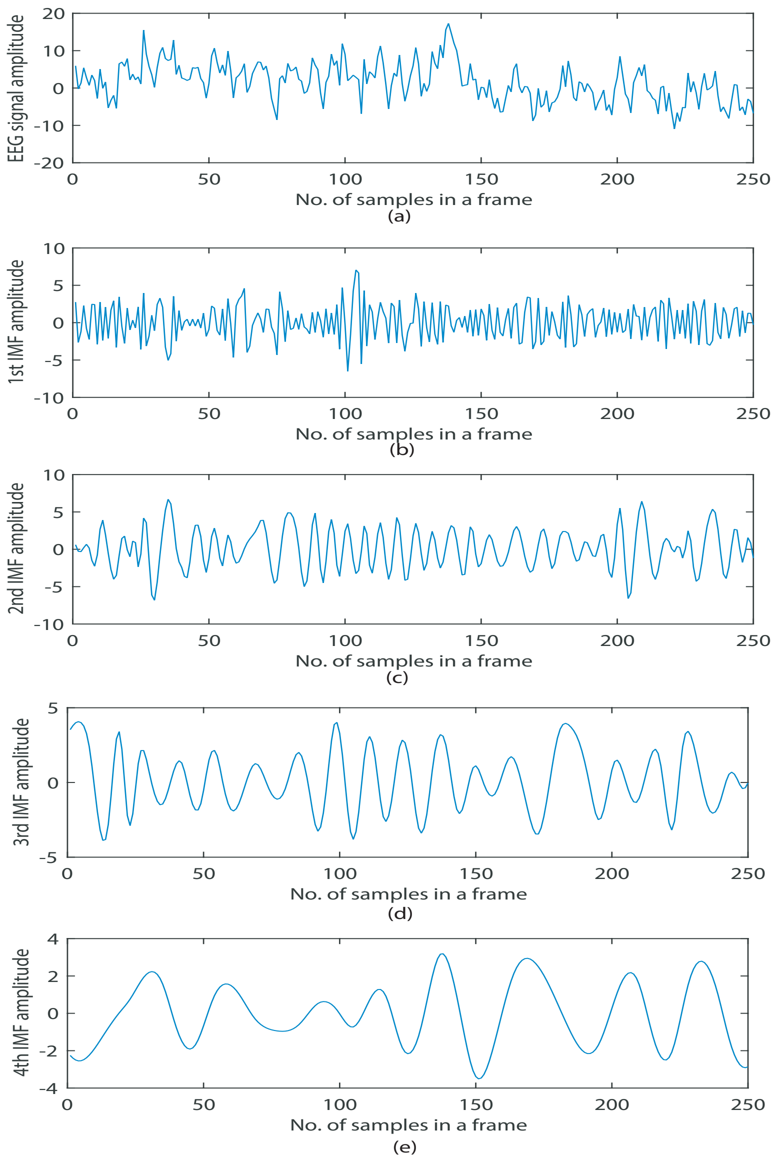

Figure 2.1: EEG signal and it's IMFs

An IMF can be defined as a function which has either equal number of maxima and minima or the difference between them is at most one. Moreover, the mean value of the envelope defined by the local maxima and the local minima is zero. In what follows, a brief description of obtaining IMFs by employing EMD on the EEG signal is described.

First, all the local maxima points of EEG data y_i, obtained from channel i, are connected to define the upper envelope and all the local minima points are connected to define the lower envelope. The new signal $h_1[n]$ is reconstructed as

$$h_1[n] = y_i[n] - \mu_1, \quad n = 1, 2, 3, ..., N, \tag{2.1}$$

where μ_1 is the mean value of the envelopes and N is the number of samples of EEG signal, y_i. The whole process is iterated w times until an IMF signal is generated according to the definition. The first IMF, u_1 is defined by

$$u_1[n] = h_w[n] \;\; and \;\; h_w[n] = h_{w-1}[n] - \mu_w, \tag{2.2}$$

where μ_w is the mean value of the envelopes at w-th iteration. The residue signal is found by subtracting the constructed IMF from the main signal,i.e.,

$$r_1[n] = y_i[n] - u_1[n]. \tag{2.3}$$

This residue signal is considered as the main signal to estimate the next IMF. The process continues until the residue signal is either a signal consisting of a single maxima or minima or a constant value. Finally, L IMFs and a residue signal are generated after performing the whole decomposition process. For L level of decomposition, $y_i[n]$ can be reconstructed as

$$y_i[n] = \sum_{m=1}^{L} u_m[n] + r_L[n], \quad n = 1, 2, 3, ..., N. \tag{2.4}$$

Here $u_1, u_2, ..., u_L$ represent the IMFs.

In the proposed method, it is observed that the number of IMFs that can be extracted considering any frame is four or more. It is expected that higher order IMFs which contain low frequency information may not be necessarily required while mental task is evaluated. Alternatively, it is expected to find more distinguishable characteristics in the IMFs that contains relatively high frequency information. As shown in Fig. 2.1(e), it is clearly observed that fourth IMF contains very low frequency information. Considering all these facts, for the sake of consistency, each channel data of a given frame is decomposed into only three IMFs. However, effect of varying the number of IMFs on the classification performance is delineated next in Sec. 2.3.1.

2.2.2 Inter-channel relation

In general, it is considered that for different types of tasks, different spatial locations of brain, such as central, parietal or occipital are stimulated. It is expected that data obtained from locations of the brain that are highly stimulated due to a specific type of task will be less correlated with data obtained from other less stimulated locations. For example, tasks involving visual effects are most likely to stimulate occipital regions predominantly. Therefore, EEG data obtained from the channels located in the occipital region will be significantly different from the data obtained from other less stimulated regions. Measuring inter-channel relationship may play a significant role to cover this spatial and temporal relationship between different channels for a particular type of task. In the proposed method, correlation coefficient is utilized to measure inter-channel relationship.

Correlation coefficient is a kind of statistical measure to quantify relationship between two or more signals. In this Chapter, it is utilized as a measuring tool to obtain inter-channel correlation of i-th and j-th channel. Instead of directly using EEG data, correlation coefficient is obtained from the m-th IMF, u_m decomposed from EEG signal. The correlation coefficient extracted from inter-channel IMF data is referred to as inter-IMFCC in this Chapter. The inter-IMFCC $R_e(i,j)$ obtained from i-th and j-th channel can be estimated as

$$R_e(i,j) = \frac{C_e(i,j)}{\sqrt{C_e(i,i)C_e(j,j)}}, \tag{2.5}$$

where $C_e(i,j)$ is the (i,j)-th component of the covariance matrix $\mathbf{C_e}$ of the i-th and j-th channel IMFs $u_m^{(i)}$ and $u_m^{(j)}$, each consists of N samples. It is expressed as

$$\mathbf{C_e} = \begin{bmatrix} cov\langle u_m^{(i)}, u_m^{(i)}\rangle & cov\langle u_m^{(i)}, u_m^{(j)}\rangle \\ cov\langle u_m^{(j)}, u_m^{(i)}\rangle & cov\langle u_m^{(j)}, u_m^{(j)}\rangle \end{bmatrix}. \tag{2.6}$$

The covariance of $u_m^{(i)}$ and $u_m^{(j)}$ denoted by $cov\langle u_m^{(i)}, u_m^{(j)}\rangle$ is calculated considering the following formula

$$cov\langle u_m^{(i)}, u_m^{(j)} \rangle = \frac{1}{N-1} \sum_{n=1}^{N} (u_m^{(i)}[n] - \mu_i)^\star (u_m^{(j)}[n] - \mu_j). \tag{2.7}$$

Here μ_i and μ_j indicate the mean of IMF data obtained from i-th and j-th channels, respectively and $\star$ denotes the complex conjugate. In the proposed method, all possible pairs of i-th and j-th channels are taken into consideration to obtain inter-IMFCC which is expected to provide maximum utilization of channel information. However, effect of choosing lesser pairs of channels are also investigated and presented in Sec. 2.3.4.

One of the major advantages of utilizing inter-IMFCC as feature is that it's values are bounded , which is $|R_e(i,j)| < 1$. If the IMF data obtained from the channels are same, inter-IMFCC is one, otherwise if there is no relationship, it is zero. To investigate the differentiating quality of inter-IMFCC as feature, a sample experiment considering multiplication and rotation task is performed. All fifteen different combination of six channels denoted as C3-C4, C3-P3, C3-P4 etc. are universally taken into consideration to measure inter-IMFCC. In Fig. 2.2, the box plot corresponding to inter-IMFCC obtained for fifteen different combinations of channels is presented. The boxplot indicates various statistical information, such as median, 25th and 75th percentile, and outliers of inter-IMFCC. There are thirty boxplots in each subfigure, each boxplot represents inter-IMFCC measured from a particular combination of channel for a particular type of task performed by Subject 1. In comparison to the boxplots presented in Fig. 2.2(a)-2.2(c), the presence of outliers in boxplot presented in Fig. 2.2(d) is much higher. As discussed before, higher order IMF contains very low frequency information which is less relevant to mental task considered here and hence poor distinctive features are expected to be extracted if 4-th IMF is used. This fact is also reflected in boxplot presented in Fig. 2.2(d). Therefore, in what follows, our discussions are restricted only for the first three IMFs.

It is observed that the values of inter-IMFCCs obtained for three combination of channels, namely C4-O1, P4-O1 and O1-O2 are found significantly higher in case of multiplication task than that in case of rotation task. It is to be noted that for these three combinations, O1 is considered as the reference channel obtained from left hemisphere and other non-reference channels, namely C4, P4 and O2,

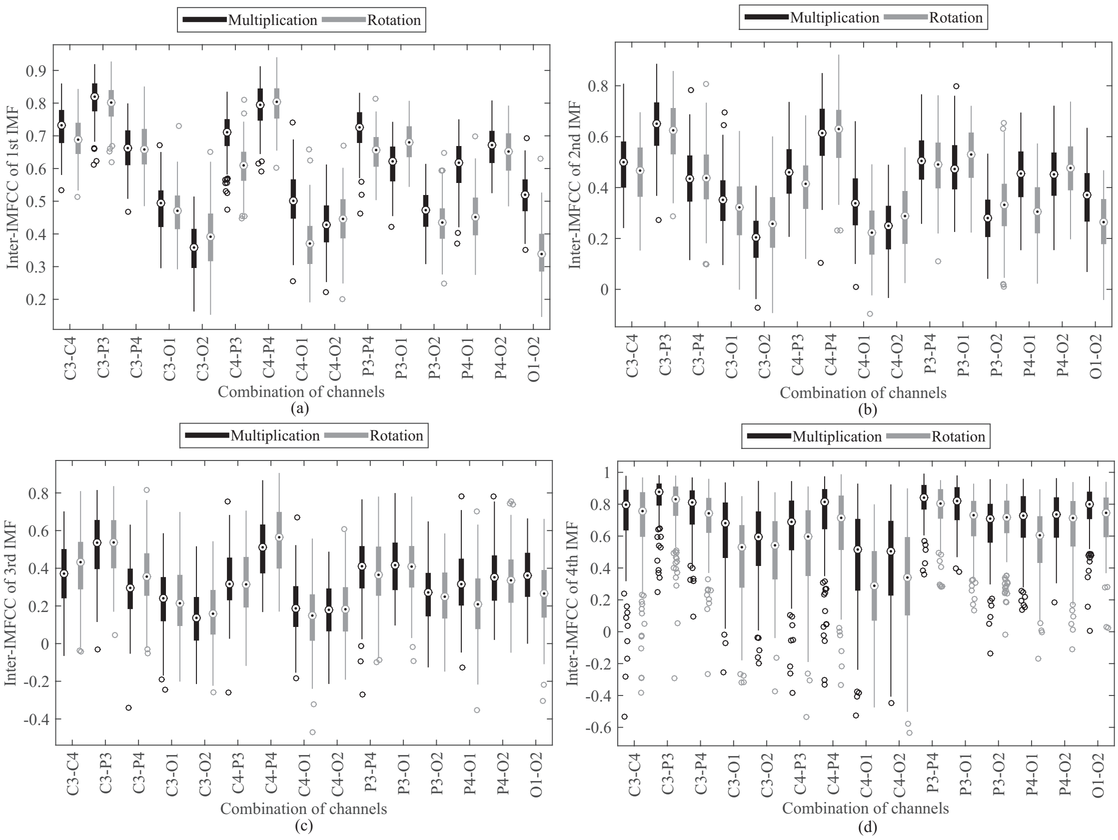

Figure 2.2: Inter-IMFCC obtained from different IMFs of Subject 1.

are from right hemisphere. One possible reason behind this observation is that in case of rotation task, O1 channel may get more stimulated than other channels due to the fact that in rotation task, visually observed objects are required to rotate around their axis mentally. As a result, data obtained from O1 channel is less correlated with data obtained from other channels, specially the channels located in right hemisphere in case of rotation task. This observation corroborates the hypothesis that during performing a particular type of task if any location of the brain becomes more excited than other locations, data obtained from stimulated location will be significantly different from data obtained from comparatively less stimulated locations resulting lower inter-IMFCC. Moreover, it is also found that inter-IMFCC values are comparatively higher in case of multiplication task than that obtained in case of rotation task. It is expected that being an arithmetic task, multiplication involves more complexity in comparison to rotation task. As a result, all locations of brain get more excited while performing multiplication task than rotation task, which in turn leads to more correlation between channels in case of multiplication task. However, location of stimulation may vary from person to person depending on the nature of task.

2.2.3 Feature extraction

In the proposed method, for the purpose of feature extraction, inter-IMFCCs are utilized to exploit the relationship among various channels. Moreover, statistical parameters, such as root mean square(RMS), standard deviation and entropy are also included in the feature vector to represent statistical measure of IMF data obtained from various channels. RMS depicts statistical measure of numerical values of varying quantity of the data obtained from channel i of corresponding IMF. For IMF data u_m consisting of N samples, RMS can be expressed as

$$rms(u_m^{(i)}) = \sqrt{\frac{1}{N}\sum_{n=1}^{N} u_m^{(i)}[n]^2}, \quad n = 1, 2, ..., N. \tag{2.8}$$

To measure the dispersion of the IMF data around it's mean value μ_i, standard deviation is proposed as a distinctive feature. Standard deviation of IMF data $u_m^{(i)}$ obtained from channel i is given by

$$std(u_m^{(i)}) = \sqrt{\frac{1}{N}\sum_{n=1}^{N} (u_m^{(i)}[n] - \mu_i)^2}, \quad n = 1, 2, ..., N. \tag{2.9}$$

For the purpose of measuring uncertainty of the IMF data $u_m^{(i)}$, entropy is introduced in the feature vector. Entropy is a statistical measure of randomness that is defined as

$$ent(u_m^{(i)}) = -\sum_{r=0}^{N} (p\langle u_m^{(i)}[r]\rangle \times log_2(p\langle u_m^{(i)}[r]\rangle)), \tag{2.10}$$

where $p\langle u_m^{(i)}[r]\rangle$ indicates the probability of occurrence of a particular value $u_m^{(i)}[r]$ of IMF data $u_m^{(i)}$ of i-th channel and is denoted by

$$p\langle u_m^{(i)}[r]\rangle = n_r/N. \tag{2.11}$$

n_r indicates the number of occurrence of $u_m^{(i)}[r]$ among the N number of samples of $u_m^{(i)}$, i.e. $\sum n_r = N$.

In brief, for the purpose of feature extraction, at first, the raw EEG signal is preprocessed with a 60 Hz notch filter. After that, the eeg data corresponding to a channel is decomposed utilizing empirical mode decomposition where from each channel data three IMFs are extracted. Finally, the feature vector is formed utilizing inter-IMFCC and statistical parameters of IMFs, such as rms, standard deviation and entropy obtained from each channel. For L number of selected IMFs and N_c number of channels for each IMF, number of inter-IMFCC obtained is $L \times^{N_c} \mathbf{C}_2$. The number of features obtained from statistical parameters (std, rms and entropy) of IMFs for a test frame is $L \times (N_c + N_c + N_c)$. Finally the total feature dimension of the proposed method is $L \times (^{N_c}\mathbf{C}_2 + 3 \times N_c)$.

2.2.4 Classification

Classifier selection is essential to obtain satisfactory result while performing test validation of the proposed method. In the proposed method, kernel based SVM classifier is chosen to effectively classify mental tasks due to its effectiveness and acceptability in supervised classification. To generate an N dimensional decision vector $\mathbf{w} = [w_1 \ w_2 \ \cdots \ w_N]^T$, features extracted from the IMF data are provided into the classifier instead of raw EEG data. The extracted features from the the training data set consisting of P frames are converted from the original space to a new representative vector space to discriminate different classes more efficiently. A class label is provided for each N_i dimensional frame $\mathbf{x_i} = x_i(n), n = 1, \cdots, N_i$. For two class problem with two class label $+1$ and -1, each frame $\mathbf{x_i}$ fulfill the following inequalities:

$$\begin{aligned} \mathbf{w^T x_i} + b &\geq +1, \text{for all positive } x_i \\ \mathbf{w^T x_i} + b &\leq -1, \text{for all negative } x_i. \end{aligned} \tag{2.12}$$

In kernel based SVM classifier, to match with class label of the training data set, the following discriminant function $f(x)$ is utilized to form the decision vector, which can be expressed as

$$f(x) = \sum_{i=1}^{P} c_i k(x_i, x) + b. \tag{2.13}$$

Here c_i is an empirical vector and kernel matrix $\mathbf{K}$ is given by

$$\mathbf{K} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_P) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_P) \\ \vdots & \vdots & \cdots & \vdots \\ k(x_P, x_1) & k(x_P, x_2) & \cdots & k(x_P, x_P) \end{bmatrix}. \tag{2.14}$$

For the purpose of classification, the performance of different kernel functions in SVM classifier is observed considering various feature extraction methods. It is found that polynomial kernel based classification approach outperforms other kernels in terms of classification accuracy. In all calculations of the reported classification accuracies, leave-one-out cross validation scheme is employed to generate classification result. In this scheme, each frame is tested one by one, i.e. when a frame is left out for testing, remaining frames are used for training. Let us consider total $N_A + N_B$ number of frames where N_A number of frames belong to class A and N_B number of frames belong to class B. In the leave-one-out cross validation scheme, when one of those $N_A + N_B$ frames is left out for testing, remaining $N_A + N_B - 1$ frames are used for training. This process is repeated $N_A + N_B$ times. Finally, classification accuracy is defined as the percentage of correctly identifying the class of each frame. Among total $N_A + N_B$ number of frames if N_t number of frames are correctly classified, the classification accuracy can be expressed as

$$Accuracy = \frac{N_t}{(N_A + N_B)} \times 100\%. \tag{2.15}$$

2.3 Simulation and Results

In this Section, performance of various feature extraction methods is investigated considering classification accuracy obtained under different conditions, such as varying the feature dimension, utilizing different statistical parameters as feature and use of various EEG channel locations. Moreover, effect of utilizing different kernel functions of SVM classifier on classification accuracy is also analyzed. A comparative analysis on classification performance between the proposed method and some other methods is also presented.

In the proposed method, instead of directly using channel data, corresponding IMFs are used to extract inter-IMFCC and statistical parameters (std, rms and entropy) using (2.4)-(2.10). Unless otherwise specified, polynomial kernel of SVM classifier is employed in leave one out cross-validation manner to obtain classification accuracy. The classification task is carried out considering two types of mental tasks at a time, as conventionally done by other researchers [11], [12]. In this way, ten different combinations of the five types of tasks, as mentioned in Sec. 2.1, are possible. Here, for notational convenience, each combination of tasks is denoted with two alphabets from two different tasks. For example, MC refers to a two class (multiplication and counting) classification problem, BL corresponds to another two class (baseline-resting and mental letter composing tasks) classification problem. In what follows, detail results and analyses are presented.

2.3.1 Effect of varying the number of IMFs

The number of IMFs to be used in the feature matrix directly dictates the feature dimension. It is already mentioned that higher order IMFs which contain very low frequency information are not necessary to be considered. The distinctive quality of the proposed inter-IMFCC feature deteriorates for 4-th IMF as shown in Fig. 2(d). Hence, in the proposed method, only first three IMFs are considered. In this subsection, effect of variation of number of IMFs are demonstrated on overall classification accuracy for four subjects. Here the number of IMFs is varied from 1 to 4 and different cases like extracting only one IMF (1IMF), two IMFs (2IMFs) etc. are considered.

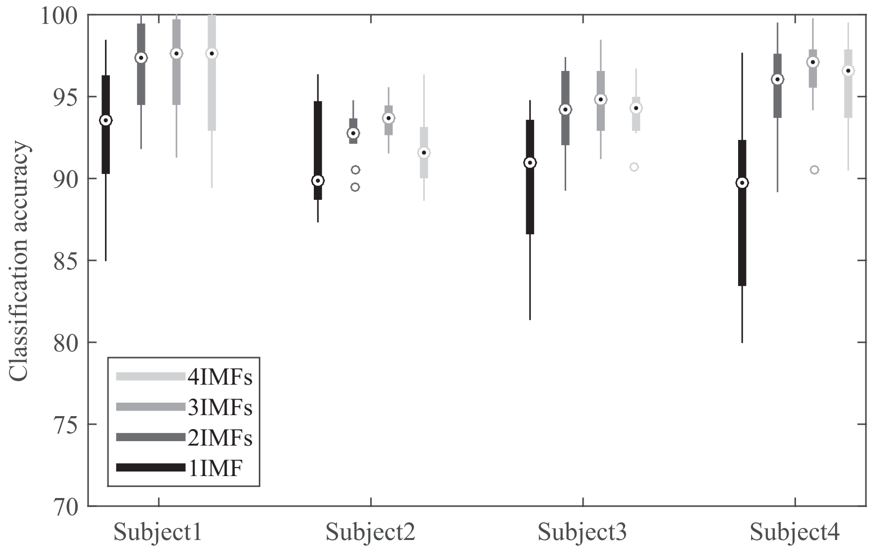

Figure 2.3: Effect of number of IMFs variation on classification accuracy

In Fig. 2.3, the box plot corresponding to classification performance obtained by varying number of IMFs is presented. The sixteen boxplots indicate various statistical information, such as median, 25th and 75th percentile, and outliers of classification accuracy. Each boxplot represents classification accuracy of ten different combination of tasks for a subject considering particular number of IMFs to be used for feature extraction. It is found that with the increase in number of IMFs, classification accuracy becomes more consistent for each subject until number of IMF is three. Moreover, it is observed that for all subjects, features extracted considering the first three IMFs offer the best classification accuracy with respect to all other combinations of IMFs. That is why, although all channel data of a frame can be decomposed into four or more IMFs, only three IMFs are considered to extract feature. Meanwhile, considering three IMFs rather than four or higher number of IMFs offer a reduced feature dimension.

2.3.2 Effect of different statistical feature

In the proposed method, as mentioned in Sec. 2.2.3, some statistical parameters are used as features, which are extracted from the channel IMF data. Effect of using conventional statistical features on classification accuracy is investigated considering ten widely used higher and lower order statistical parameters namely average (avg), median (med), mode (mod), maxima (max), minima (min), standard deviation (std), root mean sqare (rms), entropy (ent), skewness (skew) and kurtosis (kurt). For notational convenience, hereafter, each statistical feature is abbreviated as shown in the parentheses. It is to be noted that the main objective of this Chapter is to demonstrate the efficacy of proposed correlation feature (inter-IMFCC) obtained from inter-channel IMFs. It is expected that the use of proposed inter-IMFCC feature along with the conventional statistical features of IMFs will offer better classification performance. In this regard, two different cases are considered:

1. Use of only statistical features: Each statistical feature is extracted from each of three IMFs of a channel i.e., for N_C number of channels with L number of IMFs extracted from each channel, feature dimension is $N_C \times L$.

2. Use of proposed inter-IMFCC feature along with statistical feature: In this case, number of inter-channel correlation coefficients (inter-IMFCC) to be obtained from N_c channels for each IMF is ${}^{N_c}\mathbf{C}_2$. Hence, for N_C number of channels with L number of IMFs extracted from each channel, total feature dimension is $({}^{N_c}\mathbf{C}_2 + N_C) \times L$.

In Fig. 2.4, classification accuracies considering the previously discussed two cases for the ten statistical features obtained for all subjects are shown. It is observed that classification accuracy increases if inter-IMFCC is combined with channel statistical information of each IMF. Statistical parameters, such as std, rms and entropy of

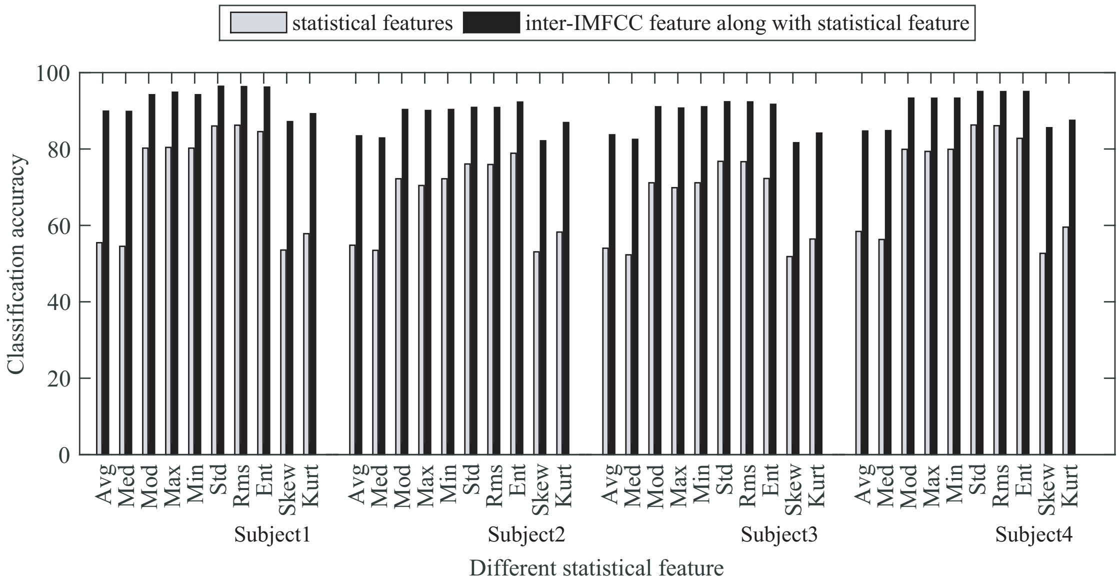

Figure 2.4: Classification accuracy obtained utilizing different statistical feature of IMFs

IMFs offer better classification performance compared to some higher order statistical feature, namely skewness and kurtosis. Moreover, features like max and min which are likely to be more biased because of the presence of noise are avoided. First order statistical parameters, such as avg, med, mod are also excluded as EEG signals are very random in nature. Due to distinctive nature of std, rms and entropy, these three statistical parameters are finally chosen for the feature vector along with proposed inter-IMFCC feature to classify mental tasks.

2.3.3 Effect of utilizing kernel of SVM classifier

The effect of using different kernels in SVM classifier on overall classification performance of the proposed method is thoroughly investigated. In order to demonstrate the performance variation due to change in kernels, three widely used kernels are considered, namely linear, quadratic and polynomial kernel. To observe the variation of classification accuracies for different kernels, all 10 different combinations of tasks, namely MC, MB, ML, MR, CB, CL, CR, BL, BR and LR from each subject are considered and average classification accuracy of those combination of tasks are measured from four subjects. In Fig. 2.5, average classification accuracies for 10 different combination of tasks by using three different kernels are plotted.

It is found that between linear and quadratic kernel, the later offers better classification performance. However, it is observed that the classification performances

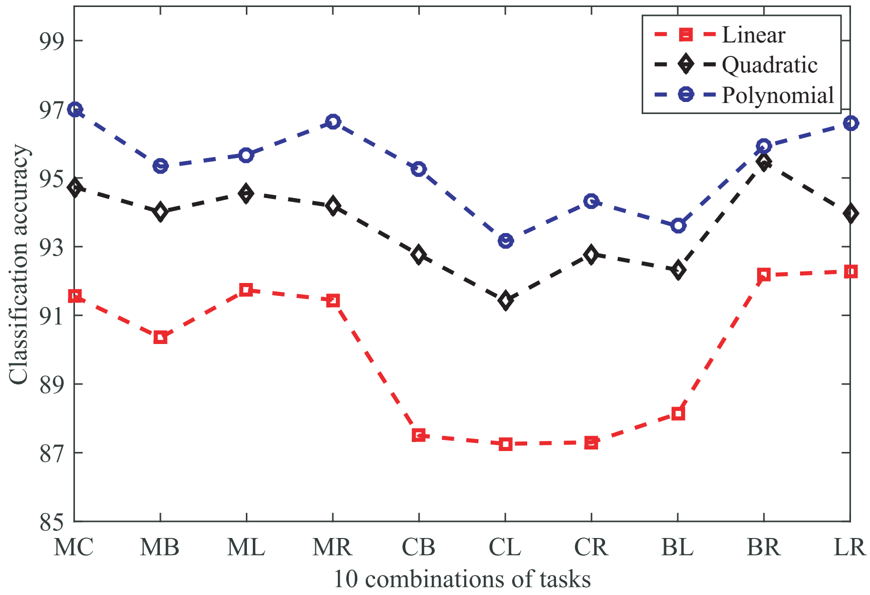

Figure 2.5: Effect of different SVM kernel on proposed IMF based method

of polynomial kernel are consistently better in comparison to those obtained by linear and quadratic kernels in all cases. For that purpose, polynomial kernel of SVM classifier is chosen to classify the tasks in the proposed method.

2.3.4 Effect of variation of number of channel pairs

In the proposed method, all possible pairs of channels are taken into consideration to obtain inter-IMFCC so that maximum channel information can be utilized. However, choosing lesser pairs of channels reduce feature size effectively. Reduction in feature size definitely helps in reducing computation time. Hence, effect of variation of the number of channel pairs is presented in this subsection. It is to be noted that in [11], [12], asymmetry ratio of a pair of channel is computed considering one channel from left hemisphere and the other channel from right hemisphere. Similarly, in this Chapter, the effect of measuring inter-IMFCC considering one channel from left hemisphere and the other from right hemisphere is investigated. This investigation is performed considering counting and baseline resting task and denoted as Exp1 in Fig. 2.6. Moreover, the effect of measuring inter-IMFCC with respect to a specific region, denoted as Exp2, is also observed.

Exp1: For three channels located in left hemisphere C3, P3, O1 and three channels located in right hemisphere C4, P4, O2, possible nine combinations of computing inter-IMFCCs are (C3, C4), (C3, P4), (C3,O2), (P3, C4), (P3, P4), (P3,O2),(O1, C4), (O1, P4), (O1,O2).

Exp2: Depending on the choice of region, such as parietal, central or occipital, to obtain reference signals, three different investigations can be performed:

(a) Considering signals of parietal region as reference, eight combination of channels for computing inter-IMFCCs are possible, such as (P3, C3), (P3, C4), (P3, O1), (P3, O2), (P4, C3), (P4, C4), (P4, O1), (P4, O2).

(b) Considering signals of central region as reference, eight combination of channels for computing inter-IMFCCs are possible, such as (C3, P3), (C3, P4), (C3, O1), (C3, O2), (C4, P3), (C4, P4), (C4, O1), (C4, O2).

(c) Considering signals of occipital region as reference, eight combination of channels for computing inter-IMFCCs are possible, such as (O1, C3), (O1, C4), (O1, P3), (O1, P4), (O2, C3), (O2, C4), (O2, P3), (O2, P4).

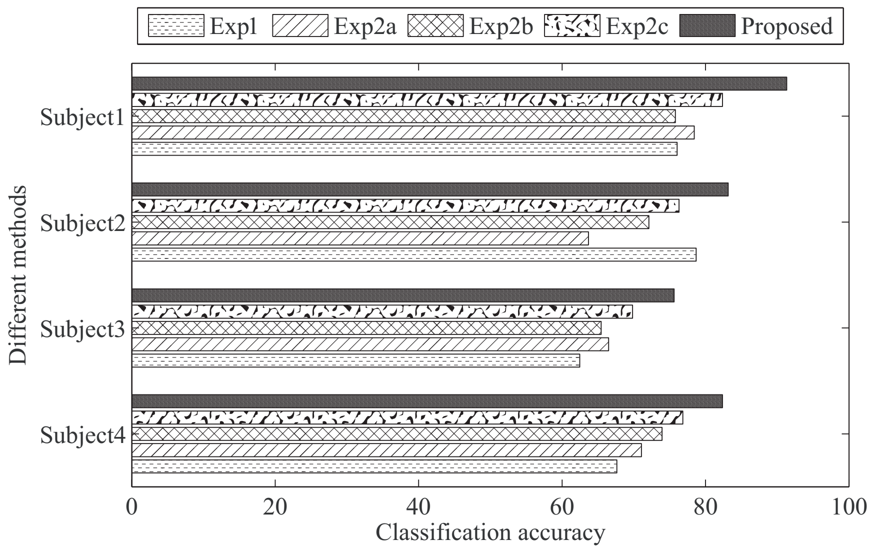

Figure 2.6: Effect of variation of number of channel pairs on proposed inter-IMFCC feature in case of MB tasks

In Fig. 2.6, a comparative analysis among these experiments is presented in terms of classification accuracy. In the above two experiments, reduced number of channel pairs are utilized and lower classification accuracy compared to the proposed method is achieved. As a result, it is not possible to select any one particular choice of reduced number of channels to obtain acceptable classification performance in all subjects.

2.3.5 Performance Comparison among Various Methods

With a view to comparing the classification performance, five methods referred to as PAR4, PAR5, PAR6, EF8 and EF3 have been considered. Among these five methods, three methods are based on power asymmetry ratio (namely PARq) computed from q number of spectral bands [11], [12]. Remaining two methods are based on EMD feature (namely EFn) where n corresponds to number of features to be extracted from each IMF obtained by EMD decomposition [24].

In PARq methods, depending on the number of frequency bands utilized, the methods are referred to as PAR4, PAR5 and PAR6. For example, in PAR4 method, features are extracted from the four traditionally used bands, namely delta (<4 Hz), theta ($4-7$ Hz), alpha ($8-13$ Hz), beta ($14-20$ Hz) while PAR5 utilizes an additional gamma band ($23-37$ Hz). In PAR6, one more additional band ($40-100$ Hz), along with these five bands, is proposed to compute power of spectral bands and asymmetry ratios. For one pair of channels, the asymmetry ratio for each spectral band is computed as [11]

$$A(i,j) = \frac{(P(i) - P(j))}{(P(i) + P(j))} \tag{2.16}$$

where two indices i and j are used to correspond channel pairs placed in the left and right hemispheres, respectively. For example, $P(i)$ corresponds to the spectral band power of the i-th channel placed in the left hemisphere and $P(j)$ corresponds to that obtained from the j-th channel placed in the right hemisphere. Depending on the number of channels (N_i and N_j) in each hemisphere, total $N_i \times N_j$ number of asymmetry ratios, denoted by $A(i,j)$, can be computed for each spectral band. As a result, the feature dimension for PAR4, PAR5 and PAR6 method is $(q \times N_i \times N_j + q \times (N_i + N_j))$ where q denotes number of spectral band considered for these methods.

On the otherhand, in EF8 method, eight features are extracted from each IMF, namely RMS, variance, Shannon entropy, Lempel-Ziv complexity measure, central frequency, maximum frequency, skewness and kurtosis. However, in the proposed method, the first three of these eight statistical features are employed along with the proposed inter-IMFCC feature. In order to better demonstrate the effect of incorporating the inter-IMFCC feature, another method EF3 is considered where only the first three features are used without the proposed inter-IMFCC feature and classification performance of the EF3 method is also compared with that of the proposed mehtod.

For the purpose of performance evaluation, leave one out cross validation technique is carried out in all methods. In Tables 2.1-2.4, the classification accuracies obtained by using four different subjects are separately reported for six methods. It is found that the classification accuracies obtained from different subjects are 90.5% or more in the proposed method. In all cases, it is observed that the proposed feature extraction method outperforms other existing methods reported in this Chapter in terms of classification accuracy. However, in some combinations of mental tasks, existing methods offer competitive classification performance with respect to proposed method. For example, in case of BR combination of Subject 1 reported in Table 2.1, both EF8 and proposed method achieve 99.74% classification accuracy. In Table 2.4, it is observed that the average classification accuracies obtained by PAR6 and EF8 are very comparable with those obtained by the proposed method. For all subjects, it is found that the average classification accuracy obtained for EF8 is very close to EF3 despite having a larger feature dimension. However, after adding inter-IMFCC along with the three parameters used in EF3, the average classification accuracy increases drastically and for Subject 2 and Subject 3, it increases around 7.5% from EF3. In each reported existing method, it is observed that for various combination of mental tasks, classification accuracy varies a lot. For example, in PAR4 method, for Subject 1 and Subject 4, the standard deviation of classification accuracies for various subjects are found 8.91% and 7.75% compared to 3.21% and 2.52% of the proposed method. It is found that the classification performance obtained by the proposed method varies from subject to subject, but not at a very large scale. For Subject 2, the standard deviation obtained from different combination of mental tasks is found 1.16% which is the least among all four subjects. It is clearly observed that the proposed method offers consistently satisfactory classification accuracy in all cases irrespective of subjects and combination of mental tasks.

2.3.6 Computation time

Average computational time is measured to extract features from one test signal for six methods namely PAR4, PAR5, PAR6, EF8, EF3 and proposed method.

Table 2.1: Classification performance comparison of proposed IMF based feature extraction method with existing methods for Subject 1

Task	PAR4 [11]	PAR5 [11]	PAR6 [12]	EF8 [24]	EF3	Proposed
MC	88.42	88.95	91.58	96.32	96.32	97.89
MB	82.11	83.95	87.37	95.79	90.53	97.37
ML	86.05	87.37	90.26	98.68	96.84	99.47
MR	89.47	91.05	93.95	98.95	97.63	100.00
CB	72.37	78.95	82.63	92.11	92.11	98.95
CL	65.26	69.74	77.11	84.47	83.95	91.32
CR	71.58	74.74	80.26	86.32	81.58	91.84
BL	69.74	71.05	82.89	90.53	86.58	94.74
BR	83.42	86.58	92.37	99.74	97.89	99.74
LR	70.26	77.37	81.84	94.47	89.21	95.53
Avg	77.87	80.97	86.03	93.74	91.26	96.68
Stddev	8.91	7.65	5.82	5.30	5.92	3.21

Table 2.2: Classification performance comparison of proposed IMF based feature extraction method with existing methods for Subject 2

Task	PAR4 [11]	PAR5 [11]	PAR6 [12]	EF8 [24]	EF3	Proposed
MC	69.47	76.58	83.42	82.37	80.79	93.95
MB	78.95	86.32	91.05	83.16	84.74	93.16
ML	70.53	83.42	87.63	90.79	88.42	95.00
MR	71.32	79.21	90.00	85.79	82.63	92.89
CB	74.74	80.53	92.11	82.89	84.74	92.63
CL	64.74	75.53	89.47	91.32	92.37	93.95
CR	68.68	73.42	87.11	79.74	84.47	93.42
BL	71.84	79.47	86.84	87.89	89.47	94.21
BR	76.05	79.74	86.05	85.53	84.47	91.58
LR	71.58	80.79	84.21	88.16	85.79	95.53
Avg	71.79	79.50	87.79	85.76	85.79	93.63
Stddev	4.01	3.74	2.85	3.79	3.41	1.16

Table 2.3: Classification performance comparison of proposed IMF based feature extraction method with existing methods for Subject 3

Task	PAR4 [11]	PAR5 [11]	PAR6 [12]	EF8 [24]	EF3	Proposed
MC	68.42	74.91	79.82	86.32	88.25	96.32
MB	73.16	74.56	81.40	85.79	86.49	93.68
ML	71.58	74.39	80.70	85.96	86.32	92.46
MR	74.74	79.47	87.89	92.63	87.89	96.49
CB	70.53	72.98	81.05	87.89	85.26	91.23
CL	72.81	77.19	80.35	88.60	83.68	93.16
CR	68.60	75.44	81.40	92.63	90.00	94.91
BL	74.39	74.56	84.21	87.54	83.68	94.91
BR	73.86	75.96	84.21	92.28	85.61	94.74
LR	77.89	83.33	87.54	92.98	93.33	98.42
Avg	72.60	76.28	82.86	89.26	87.05	94.63
Stddev	2.92	3.05	2.96	3.03	2.96	2.11

Table 2.4: Classification performance comparison of proposed IMF based feature extraction method with existing methods for Subject 4

Task	PAR4 [11]	PAR5 [11]	PAR6 [12]	EF8 [24]	EF3	Proposed
MC	83.95	90.53	97.63	99.47	98.42	99.74
MB	86.84	90.53	94.74	96.84	95.00	97.11
ML	86.05	88.42	92.89	95.00	92.89	95.79
MR	84.21	88.95	93.95	92.63	91.32	97.11
CB	81.58	82.37	86.58	94.21	95.26	98.16
CL	78.16	81.58	87.63	85.79	87.89	94.21
CR	88.68	92.89	96.32	95.26	92.63	97.11
BL	68.16	77.63	83.95	85.26	86.58	90.53
BR	94.47	95.26	97.11	97.89	93.95	97.63
LR	94.47	96.05	97.37	96.05	93.16	96.84
Avg	84.66	88.42	92.82	93.84	92.71	96.42
Stddev	7.75	6.09	4.99	4.78	3.48	2.52

Table 2.5: Feature dimension and feature extraction time comparison of proposed IMF based method with existing methods

Different methods	PAR4 [11]	PAR5 [11]	PAR6 [12]	EF8 [24]	EF3	Proposed
Feature dimension	60	75	90	192	72	99
Average time (ms)	52.63	60.02	76.36	2749.30	128.80	108.11

The whole process of computation is performed using Intel(R) Core(TM) i5-4200M processor with 2.50 GHz clock speed and 4 GB ram. The feature dimension and the feature extraction time for six methods are listed in Table 2.5.

It is found that the proposed method uses a very small computation time for feature extraction compared to recently reported EF8 method. One of the reasons for such a small computation time for the proposed method is it's feature dimension compared to EF8. For three selected IMFs and six channels for each IMF, the feature dimension of the proposed method is $3\times({}^{6}\mathbf{C}_2+3\times6) = 99$. On the contrary, for four selected IMFs and similar number of channels for each IMF, the feature dimension of the EF8 and EF3 method is $4 \times (8 \times 6) = 192$ and $4 \times (3 \times 6) = 72$ respectively. In case of PAR4, PAR5 and PAR6, feature dimension is $4 \times 3 \times 3 + 4 \times 6 = 60$, $5 \times 3 \times 3 + 5 \times 6 = 75$ and $6 \times 3 \times 3 + 6 \times 6 = 90$ respectively. The PAR4, PAR5 and PAR6 method utilizes lesser time and features but the classification accuracies are lesser in these methods than the proposed method.

2.4 Conclusion

In the proposed mental task classification scheme, inter-channel correlation coefficient of each IMF is utilized to explore the relationship between channels, which is referred to as inter-IMFCC method. Moreover, intra-channel features, such as standard deviation, rms and entropy of each IMF are also measured. Finally, both inter-channel features and intra channel features of each IMF are utilized to form feature vector and a quite satisfactory classification performance is achieved. It is observed that increase in feature dimension by considering more IMFs not necessarily provides better classification performance and thus only three IMFs from each channel are found sufficient. Effect of selecting different combinations of channels is also investigated and it is observed that considering all combination of channels provide the best classification performance irrespective of the tasks or the subjects. Classification performance for various feature extraction methods are listed considering polynomial kernel and it is observed that the proposed method outperforms other methods in terms of classification accuracy. Results obtained from various types of investigation verify that the proposed mental task classification scheme is capable of classifying EEG signals with high classification accuracy.

Chapter 3

Mental Task Classification Scheme Utilizing Correlation Coefficient Extracted from Inter-channel Spectral Band Limited Signal

In this Chapter, mental task classification scheme utilizing correlation coefficient extracted from inter-channel spectral band limited signals is presented. It is expected that features extracted from band-pass filtered EEG data will provide more consistent characteristics in comparison to that obtained from full band raw EEG data. For this purpose, different well defined narrow frequency bands , namely delta (< 4 Hz), theta ($4-7$ Hz), alpha ($8-13$ Hz), beta ($14-20$ Hz), and wide frequency bands, namely gamma ($24-37$ Hz) and $40-100$ Hz bands are used to preprocess the EEG signal. Each band-limited signal is utilized to compute the correlation coefficient from inter-channel spectral band limited signals, referred to as inter-SBCC. In view of obtaining the proposed inter-SBCC feature, only the channel information of the test frame is utilized, therefore no previously defined reference channel data is required for that purpose. It is shown that use of proposed inter-SBCC feature can drastically improve the classification performance obtained by conventional statistics of spectral band-limited signals. Classification process is carried out by using the SVM classifier. The effect of extracting features from each band limited signal on classification accuracy is also investigated. Extensive experimentation is carried out on the same dataset used in the previous Chapter.

3.1 Proposed Method

The proposed mental task classification scheme can be divided into four major steps: preprocessing, inter-channel relation, feature extraction and classification. These steps are described in detail in the following subsections.

3.1.1 Preprocessing

Due to random nature of recordings of EEG data, it is very difficult to obtain discriminative characteristic from the time domain EEG data. Therefore, instead of directly utilizing EEG data, it may be easier to extract distinctive characteristic from spectral band limited EEG data. In EEG signal analysis, depending on the nature of practical applications, different well defined narrow frequency bands , namely delta (< 4 Hz), theta ($4-7$ Hz), alpha ($8-13$ Hz), beta ($14-20$ Hz), and gamma ($24-37$ Hz) are widely investigated for feature extraction [11]. However, in the current application of mental task classification, it may not be useful to restrict the EEG signal analysis only to these low frequency bands. The reason behind is explained as follows. It is well known that while performing mental tasks, relatively high frequency bands (e.g. beta bands or even higher) remain active. Considering this fact in [11], for the purpose of mental task classification, frequency band up to 37 Hz and in [12] frequency band up to 100 Hz is used. In view of investigating the presence of high frequency components in EEG signal while performing mental tasks, spectral analysis on a large number of EEG frames taken from different channels is carried out. In Fig. 3.1, average values of magnitude spectra along with standard deviations, computed from 19 consecutive overlapping frames of EEG signal obtained from Subject 1 considering mathematical multiplication task, are plotted. As mentioned before, these frames correspond to one complete session within which the mental task is performed. It is clearly observed from this figure that substantial amount of spectral information exists in high frequency region (> 40 Hz) of the averaged magnitude spectra. It is found that the patterns of the averaged spectrum obtained in different other frames exhibit quite similar nature. As a result, in the proposed method, like [12] six bands are utilized in order to extract spectral information residing in higher frequency region. However, effect of selecting a specific band-limited signal on the classification performance is delineated next in the result section. It is to be mentioned that in order to remove 60 Hz artifact, at the beginning a digital notch filter is used.

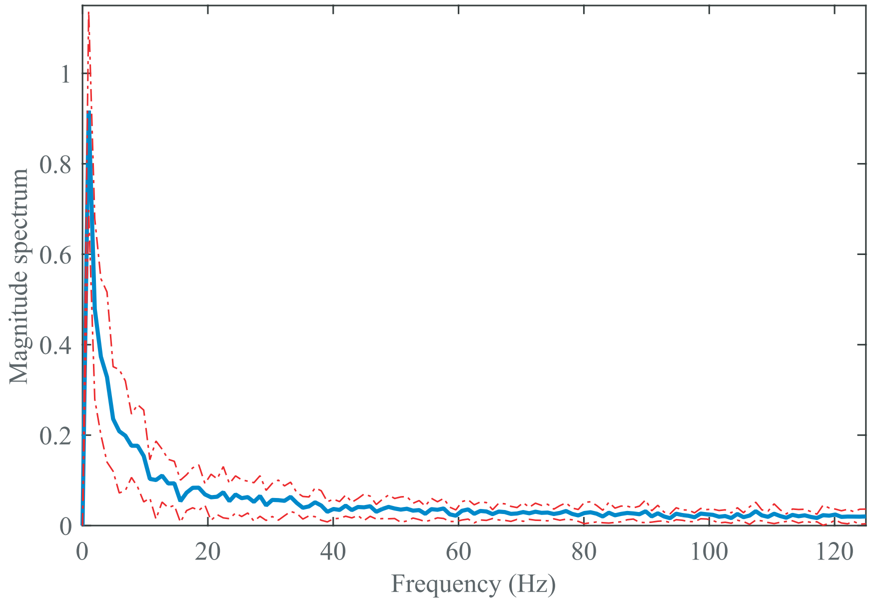

Figure 3.1: Average magnitude spectrum corresponding to a session of mathematical multiplication task obtained from C3 channel of Subject 1. The red dotted line on both sides of the average spectrum indicates the standard deviation.

3.1.2 Inter-channel relation

As discussed in Chapter 2, measuring inter-channel relationship can play a significant role to to cover the spatial and temporal relationship between different channels for a particular type of task. It is expected that data obtained from locations of the brain that are highly stimulated due to a specific type of task will be less correlated with data obtained from other less stimulated locations. This hypothesis may be utilized to obtain distinctive feature by measuring inter-channel relationship. In this regard, in the proposed method, correlation coefficient is utilized to measure inter-channel relationship.

Correlation coefficient is a kind of statistical measure to quantify relationship between two or more signals. In this Chapter, it is utilized as a measuring tool to obtain inter-channel correlation of i-th and j-th channel. Instead of directly using EEG data, correlation coefficient is obtained from a particular band limited EEG signal, denoted as f_m. f_1, f_2, f_3, f_4, f_5 and f_6 correspond to EEG signal extracted from delta, theta, alpha, beta, gamma and $40 - 100Hz$ band respectively. The correlation coefficient extracted from inter-channel spectral band data is referred to as inter-SBCC in this Chapter. The inter-SBCC $R_f(i,j)$ obtained from i and j-th electrode can be estimated as

$$R_f(i,j) = \frac{C_f(i,j)}{\sqrt{C_f(i,i)C_f(j,j)}} \tag{3.1}$$

where $C_f(i,j)$ is the (i,j)-th component of the covariance matrix $\mathbf{C_f}$ obtained from the channel band-limited signal $f_m^{(i)}$ and $f_m^{(j)}$, each consists of N samples. It is expressed as

$$\mathbf{C_f} \quad = \quad \begin{bmatrix} cov\left\langle f_m^{(i)}, f_m^{(i)} \right\rangle & cov\left\langle f_m^{(i)}, f_m^{(j)} \right\rangle \\ cov\left\langle f_m^{(j)}, f_m^{(i)} \right\rangle & cov\left\langle f_m^{(j)}, f_m^{(j)} \right\rangle \end{bmatrix}. \tag{3.2}$$

The covariance of $f_m^{(i)}$ and $f_m^{(j)}$ denoted by $cov\left\langle f_m^{(i)}, f_m^{(j)} \right\rangle$ is calculated considering the following formula

$$cov\left\langle f_m^{(i)}, f_m^{(j)} \right\rangle = \frac{1}{N-1}\sum_{n=1}^{N}(f_m^{(i)}[n] - \mu_i)^{\star}(f_m^{(i)}[n] - \mu_j). \tag{3.3}$$

Here μ_i and μ_j indicate the mean of band-limited signal obtained from i-th and j-th channels respectively and $\star$ denotes the complex conjugate. Similar to inter-IMFCC feature extraction method described in Chapter 2, to obtain inter-SBCC, all possible pairs of i-th and j-th channels are taken into consideration, The reason behind that is to utilize the information obtained from all possible channel pairs. To investigate the distinctive quality of inter-SBCC as feature, a sample experiment, similar to the one investigated in Sec. 2.2.2 considering multiplication and rotation task, is performed. Fifteen different combination of six channels denoted as namely 'C3-C4', 'C3-P3', 'C3-P4', 'C3-O1', 'C3-O2', 'C4-P3', 'C4-P4', 'C4-O1', 'C4-O2', 'P3-P4', 'P3-O1', 'P3-O2', 'P4-O1', 'P4-O2' and 'O1-O2' are taken into consideration to measure inter-SBCC. In Fig. 3.2, the box plot corresponding to inter-SBCC obtained for these fifteen different combinations of channels is presented. As described in the previous chapter, the boxplot indicates various statistical information, such as median, 25th and 75th percentile, and outliers of inter-SBCC. There are thirty boxplots, each boxplot represents inter-SBCC measured from a particular combination of channel for a particular type of task.

It is observed that the presence of outliers decrease gradually from boxplot presented in Fig. 3.2(a) to boxplot presented in Fig. 3.2(f). The boxplot in Fig. 3.2(a) represents the inter-SBCC feature quality obtained from delta band signal which

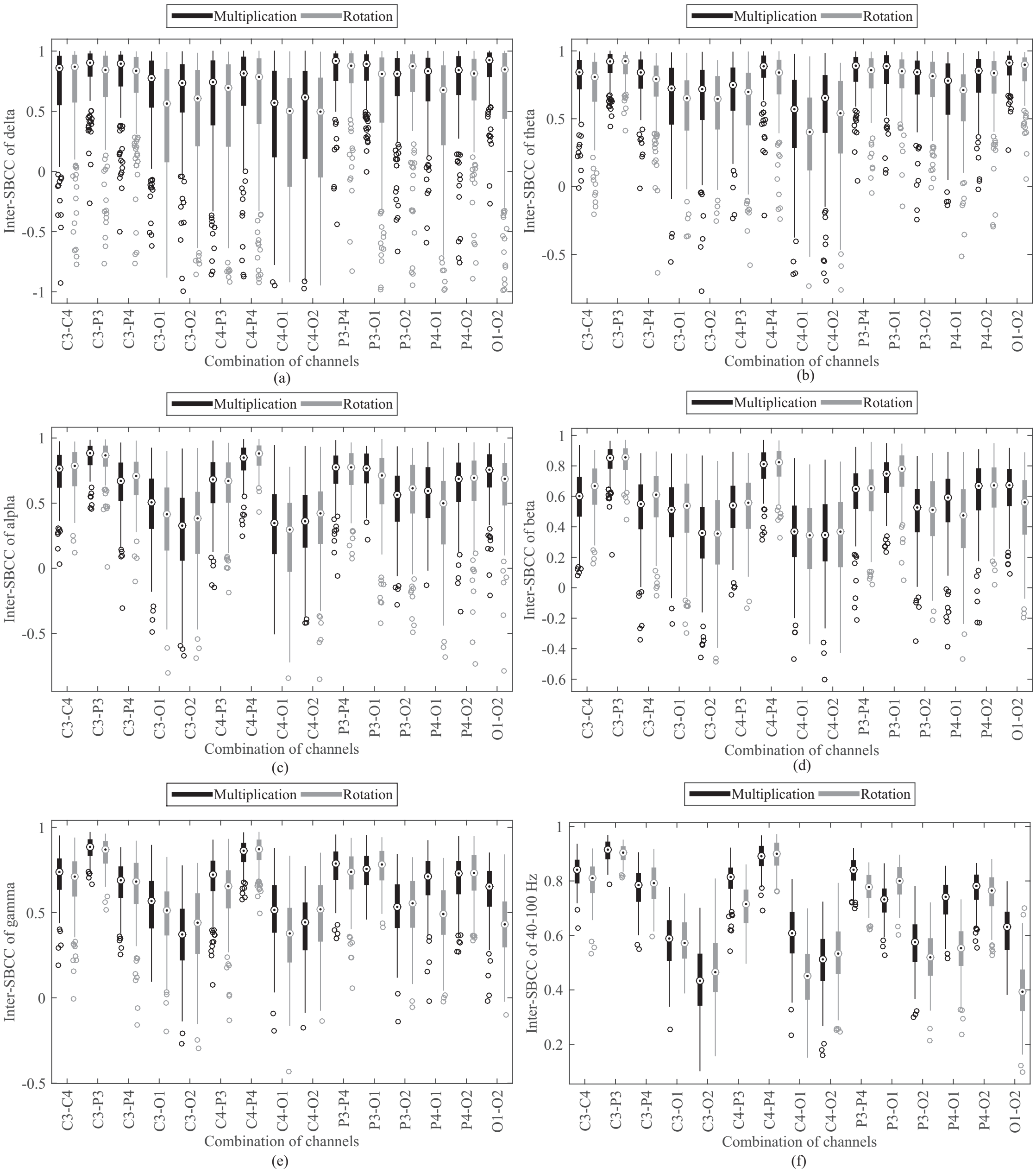

Figure 3.2: Inter-SBCC obtained from different band-limited signals of Subject 1

contains very low frequency information. Similarly, boxplots in Fig. 3.2(b)-Fig. 3.2(f) represent the statistical information of the proposed inter-SBCC feature obtained from theta, alpha, beta, gamma and $40-100$ Hz band respectively. As discussed before, in mental task application, relatively high frequencies remain active. Therefore, comparatively better feature quality is obtained in high frequency band-limited signals than that obtained by low frequency narrow band-limited signals as shown in Fig. 3.2.

3.1.3 Feature extraction

In the proposed method, for the purpose of feature extraction, inter-SBCCs are utilized to exploit the relationship among various channels. Moreover, root mean square(RMS), standard deviation and entropy are also included in the feature vector to represent statistical measure of data obtained from various channels. Like inter-SBCC method, statistical parameters are also obtained from band-limited signals obtained from EEG data instead of EEG data itself. RMS depicts statistical measure of numerical values of varying quantity of the data obtained from channel i of corresponding band, f_m, which can be expressed as

$$rms(f_m^{(i)}) = \sqrt{\frac{1}{N}\sum_{n=1}^{N} f_m^{(i)}[n]^2} \tag{3.4}$$

Standard deviation of band-limited signal obtained from i-th channel is given by

$$std(f_m^{(i)}) = \sqrt{\frac{1}{N}\sum_{n=1}^{N} (f_m^{(i)}[n] - \mu_i)^2} \tag{3.5}$$

Here, μ_i indicates mean value of the data.

For the purpose of measuring uncertainty of the band-limited signal $f_m^{(i)}$, entropy is introduced,which is defined as

$$ent(f_m^{(i)}) = -\sum(p\langle f_m^{(i)}[r]\rangle \star log_2(p\langle f_m^{(i)}[r]\rangle)) \tag{3.6}$$

where $p\langle f_m^{(i)}\rangle$ indicates the probability of occurrence of a particular value $f_m^{(i)}[r]$ of $f_m^{(i)}$ and is denoted by

$$p\langle f_m^{(i)}[r]\rangle = n_r/N. \tag{3.7}$$

n_r indicates the number of occupance of $f_m^{(i)}[r]$ among the N number of values of $f_m^{(i)}$, i.e. $\sum n_r = N$.

In brief, for the purpose of feature extraction, at first, the raw EEG signal is preprocessed with a 60 Hz notch filter. After that, six band limited EEG signals are extracted from the raw EEG data of a channel. The feature vector is formed utilizing inter-SBCC and statistical parameters of spectral bands, such as rms, standard deviation and entropy obtained from each channel. A method utilizing statistical features extracted from the spectral band (SB) is referred to as SBS and when inter-SBCC is extracted from SB, it is termed as SBC. The proposed method, denoted as

SBS+SBC utilizes both the SBS and the SBC methods simultaneously for extracting distinctive features from spectral bands. For N_c number of channels and L number of spectral bands extracted from each channel, number of inter-SBCC obtained is $L \times {}^{N_c}\mathbf{C}_2$. The number of features obtained from statistical parameters of band-limited signals for a test frame is $L \times (N_c + N_c + N_c)$. Finally the total feature dimension of the proposed method is $L \times ({}^{N_c}\mathbf{C}_2 + N_c + N_c + N_c)$.

3.1.4 Classification

The classification process is carried out via different kernels of SVM classifier similar to the previous Chapter. It is found that polynomial kernel based classification approach outperforms other kernels in terms of classification accuracy for the method proposed in this Chapter. In all cases, leave-one-out cross validation scheme is employed to generate classification result, where each frame is tested one by one. During the testing of a frame, all the remaining frames are used for training. The overall accuracy is calculated based on the classification results obtained in all the frames using (2.15) described in the previous Chapter.

3.2 Simulation Results and Discussion

In this Section, performance of various feature extraction methods is investigated considering classification accuracy obtained under different conditions, such as choosing a specific band limited signal, utilizing different statistical parameters as feature and use of various EEG channel locations. Moreover, effect of different kernel functions in SVM classifier on classification accuracy is also analyzed. A comparative analysis on classification performance between the proposed method and some other methods is also performed.

In the proposed method, instead of directly using channel data, corresponding band-limited EEG signals are used to extract inter-SBCC and statistical parameters (std, rms and entropy) using (3.1)-(3.7). Unless otherwise specified, polynomial kernel of SVM classifier is employed in leave one out cross-validation manner to obtain classification accuracy. In what follows, detail results and analyses are presented.

3.2.1 Effect of Frequency Band Selection

In different EEG signal analysis, most commonly band limited signals are used considering conventional frequency bands, namely delta, theta, alpha, beta, and gamma [11]. Estimating inter-SBCC and statistical parameters from a specific band-limited EEG signal may not be capable of providing representative characteristics. However, for the purpose of investigation, each band of EEG signal is separately generated by using narrow-band filters and fifteen inter-SBCC are estimated from the band-limited EEG signals. Classification performance for each band is separately computed. Moreover, various wide-band signals, such as $24-37$ Hz or $40-100$ Hz signals, are also taken in consideration and here also classification performance is computed considering the fifteen inter-SBCCs and statistical features.

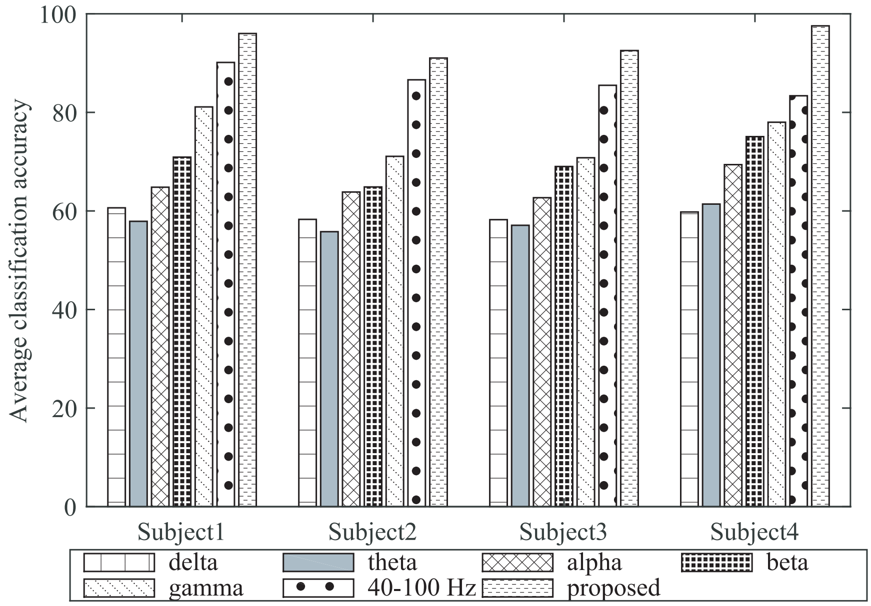

Figure 3.3: Effect of frequency band selection on classification accuracy for all four subjects in case of MC tasks.

The variation of classification performance for different band limited EEG signals is demonstrated in Fig. 3.3. It is observed that extracting features from different narrow-band EEG signals cannot provide satisfactory performance. However, considering wide-band EEG signals offer comparatively better performance than narrow-band EEG signals. In particular, the best classification performance is achieved when features from all bands are merged together. That is why, specific band limitation of the given EEG data is not adopted in this Chapter.

3.2.2 Effect of Different Statistical Feature

In the proposed method, as mentioned in Sec. 3.1.3, some statistical parameters are used as features, which are extracted from various band-limited EEG signals obtained from different channels. Similar to the previous Chapter, in this Chapter, effect of using conventional statistical features on classification accuracy is investigated considering same set of statistical parameters namely avg, med, mod, max, min, std, rms, ent, skew and kurt. It is to be noted that the main objective of this Chapter is to demonstrate the efficacy of proposed correlation feature obtained from various band-limited signals. For that purpose, a sample experiment utilizing band-limited signals is considered. It is expected that the use of proposed inter-SBCC feature along with the conventional statistical features will offer better classification performance. In this regard, two different cases similar to the previous Chapter are considered:

1. Use of only statistical features

2. Use of proposed inter-SBCC feature along with statistical feature

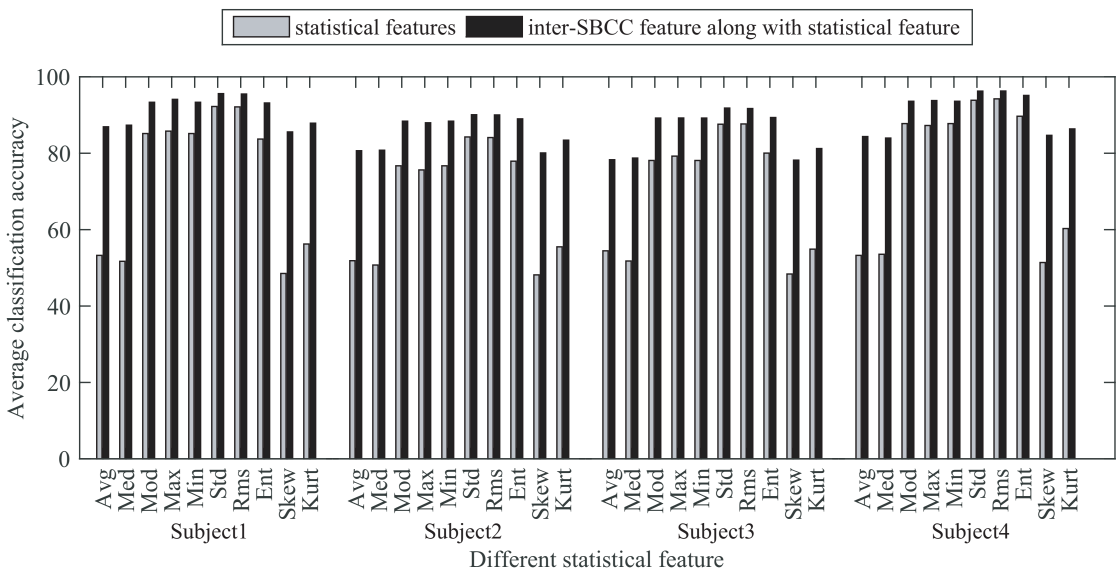

Figure 3.4: Effect of different statistical features of band limited signals on classification accuracy for all four subjects

In Fig. 3.4, classification accuracies considering the previously discussed two cases for the ten statistical features obtained from all four subjects are shown. It is observed that classification accuracy increases if inter-SBCC is combined with channel statistical information of band-limited signals. Statistical features, such as std, rms and entropy of band-limited signals offer better classification performance similar to the feature extraction method described in the previous Chapter. Due to distinctive nature of std, rms and entropy, these three parameters are chosen in the feature vector along with proposed inter-SBCC feature to classify mental tasks. In Fig. 3.5, average classification accuracies of ten different combinations of mental tasks obtained by both SBS and SBS+SBC methods are shown to observe the effect of proposed inter-channel correlation feature on the proposed statistical features. It is found that use of spectral band domain inter-channel correlation feature drastically improves the classification accuracy of statistical features of band-limited signals in all four subjects.

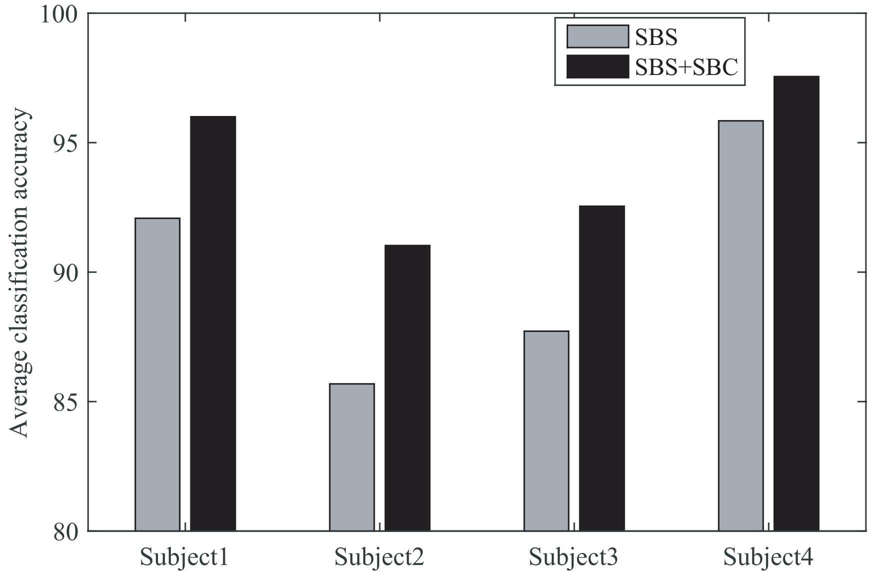

Figure 3.5: Effect of inter-SBCC feature on statistical features in terms of classification accuracy for four subjects

3.2.3 Effect of Kernel in SVM Classifier

The effect of using different kernels in SVM classifier on overall classification performance of the proposed method is thoroughly investigated similar to the feature extraction method described in Chapter 2. In Fig. 3.6, average classification accuracies for 10 different combination of tasks by using three different kernels are plotted. It is found that between linear and quadratic kernel, the latter offers better classification performance. However, it is observed that the classification performances of polynomial kernel are consistently better in comparison to those obtained by linear and quadratic kernels in all cases as observed in the previous Chapter. For that purpose, polynomial kernel of SVM classifier is chosen to classify the tasks in the proposed method.

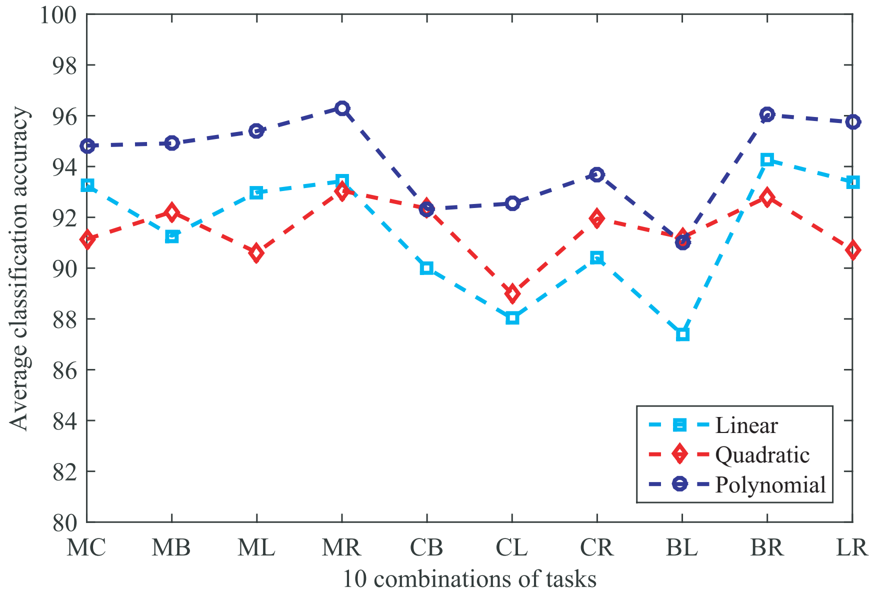

Figure 3.6: Effect of various SVM kernels on proposed spectral band division based method

3.2.4 Effect of variation of number of channel pairs

In the proposed method, all possible pairs of channels are taken into consideration to obtain the proposed inter-channel correlation feature so that maximum channel information can be utilized. However, similar to the previous Chapter, two different experiments are performed to observe the effect of variation of number of channel pairs. In Fig. 3.7, a comparative analysis among these experiments is presented in terms of classification accuracy. In these two experiments, reduced number of channels are utilized and lower classification accuracy compared to the proposed method is found. As a result, it is not possible to select any one particular choice of reduced number of channels to obtain acceptable classification performance in all subjects.

3.2.5 Performance Comparison among Various Methods

The classification performance of the proposed method and that of the three available methods reported in [11], [12] is compared. Among these three methods, the

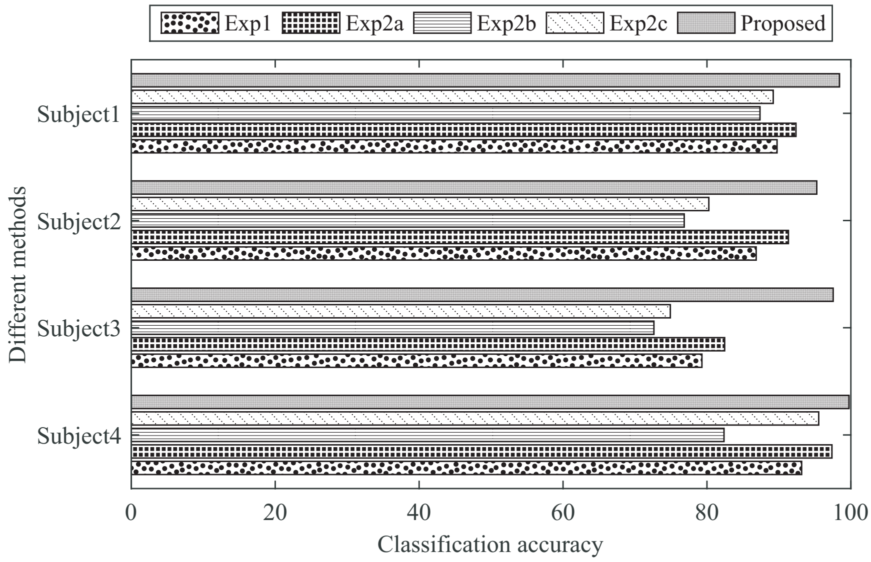

Figure 3.7: Effect of variation of number of channel pairs on proposed inter-SBCC feature

first one utilizes power of spectral bands and asymmetry ratios from four bands (referred to as PAR4) and the second one also utilizes similar power and asymmetry ratios from five bands including the Gamma band (referred to as PAR5) as features. The third one introduces one additional band ($40 - 100$ Hz) along with the five bands utilized in third method and extracts power and asymmetry ratios as features (referred to as PAR6).

For the purpose of performance evaluation, leave one out cross validation technique is carried out in all methods. In Tables 3.1-3.4, the classification accuracies obtained by using four different subjects are separately reported for four methods. It is found that the classification accuracies obtained from different subjects are 87.19% or more in the proposed method. It is found that the classification performance obtained by the proposed method varies from subject to subject, not at a very large scale. Methods 1 to 3 provide relatively low classification accuracy and consistently better classification accuracy is obtained by the proposed method compared to other methods.

3.2.6 Computation Time

Average computational time is measured to extract features from one test signal for four methods namely PAR4, PAR5, PAR6 and proposed method. The whole process of computation is performed using Intel(R) Core(TM) i5-4200M processor

Table 3.1: Overall classification accuracy comparison of proposed SBS+SBC method with existing methods for Subject 1

Task	PAR4 [11]	PAR5 [11]	PAR6 [12]	Proposed
MC	88.42	88.95	91.58	97.37
MB	82.11	83.95	87.37	97.37
ML	86.05	87.37	90.26	98.95
MR	89.47	91.05	93.95	99.47
CB	72.37	78.95	82.63	96.58
CL	65.26	69.74	77.11	91.32
CR	71.58	74.74	80.26	94.21
BL	69.74	71.05	82.89	89.21
BR	83.42	86.58	92.37	99.47
LR	70.26	77.37	81.84	96.05
Avg	77.87	80.97	86.03	96.00
Stddev	8.91	7.65	5.82	3.47

Table 3.2: Overall classification accuracy comparison of proposed SBS+SBC method with existing methods for Subject 2

Task	PAR4 [11]	PAR5 [11]	PAR6 [12]	Proposed
MC	69.47	76.58	83.42	91.32
MB	78.95	86.32	91.05	93.95
ML	70.53	83.42	87.63	93.95
MR	71.32	79.21	90.00	92.37
CB	74.74	80.53	92.11	87.63
CL	64.74	75.53	89.47	90.53
CR	68.68	73.42	87.11	89.47
BL	71.84	79.47	86.84	90.26
BR	76.05	79.74	86.05	90.53
LR	71.58	80.79	84.21	90.26
Avg	71.79	79.50	87.79	91.03
Stddev	4.01	3.74	2.85	1.96

Table 3.3: Overall classification accuracy comparison of proposed SBS+SBC method with existing methods for Subject 3

Task	PAR4 [11]	PAR5 [11]	PAR6 [12]	Proposed
MC	68.42	74.91	79.82	90.88
MB	73.16	74.56	81.40	90.18
ML	71.58	74.39	80.70	92.11
MR	74.74	79.47	87.89	95.79
CB	70.53	72.98	81.05	87.19
CL	72.81	77.19	80.35	92.81
CR	68.60	75.44	81.40	92.98
BL	74.39	74.56	84.21	91.40
BR	73.86	75.96	84.21	94.91
LR	77.89	83.33	87.54	97.19
Avg	72.60	76.28	82.86	92.54
Stddev	2.92	3.05	2.96	2.92

Table 3.4: Overall classification accuracy comparison of proposed SBS+SBC method with existing methods for Subject 4

Task	PAR4 [11]	PAR5 [11]	PAR6 [12]	Proposed
MC	83.95	90.53	97.63	99.74
MB	86.84	90.53	94.74	98.16
ML	86.05	88.42	92.89	96.58
MR	84.21	88.95	93.95	97.63
CB	81.58	82.37	86.58	97.89
CL	78.16	81.58	87.63	95.53
CR	88.68	92.89	96.32	98.16
BL	68.16	77.63	83.95	93.16
BR	94.47	95.26	97.11	99.21
LR	94.47	96.05	97.37	99.47
Avg	84.66	88.42	92.82	97.55
Stddev	7.75	6.09	4.99	2.01

Table 3.5: Feature dimension and feature extraction time comparison for proposed SBS+SBC method with existing methods

Different methods	PAR4 [11]	PAR5 [11]	PAR6 [12]	Proposed
Feature dimension	60	75	90	198
Average time (ms)	52.63	60.02	76.36	34.19

with 2.50 GHz clock speed and 4 GB ram. The feature dimension and the feature extraction time for four methods are listed in Table 3.5.

It is found that despite having more features, the proposed method uses a very small computation time for feature extraction compared to other three methods. For six selected band-limited signals and six channels of each band-limited signal, the feature dimension of the proposed method is $6 \times ({}^{6}\mathbf{C}_2 + 3 \times 6) = 198$. In case of PAR4, PAR5 and PAR6, feature dimensions are $4 \times 3 \times 3 + 4 \times 6 = 60$, $5 \times 3 \times 3 + 5 \times 6 = 75$ and $6 \times 3 \times 3 + 6 \times 6 = 90$ respectively.

3.3 Conclusion

In the proposed mental task classification scheme, inter-channel correlation coefficient of each band-limited signal is utilized to explore the relationship between channels, which is referred to as inter-SBCC method. Moreover, intra-channel features, such as standard deviation, rms and entropy of each band-limited signal are also measured. It is found that consistently better classification accuracy is obtained if proposed inter-SBCC feature is utilized along with conventional statistical features of the band-limited signals. It is observed that adoption of specific band limitation does not provide better classification performance and thus all six bands from each channel are taken into consideration for feature extraction. Effect of selecting different combinations of channels is also investigated and it is observed that considering all combination of channels provide the best classification performance irrespective of the task or the subject. Results obtained from various types of investigation verify that the proposed method outperforms other methods in terms of classification accuracy.

Chapter 4

Mental Task Classification Scheme Utilizing Correlation Coefficient Extracted from Inter-channel Wavelet Domain Signal

In this Chapter, an efficient scheme of extracting features from EEG signal is proposed for mental task classification based on inter-channel relationship in wavelet domain. It is shown that use of wavelet domain inter-channel relationship can drastically improve the classification performance obtained by conventional wavelet statistics. Both multi-level wavelet decomposition and node reconstruction are utilized for proposed inter-channel correlation feature extraction. It is expected that the correlation obtained from different combination of channels will be different for various mental tasks depending on the nature of the stimulus generated in the brain and thus can provide distinctive features. Support vector machine (SVM) classifier is used to carry out classification of five different mental tasks obtained from the same dataset described in Chapter 2. It is found that the proposed scheme can classify mental tasks with a very high level of accuracy compared to that obtained by some existing methods.

4.1 Proposed Method

The proposed mental task classification scheme can be divided into four major steps: wavelet domain analyses, inter-channel relation, feature extraction and classification. These steps are described in detail in the following subsections.

4.1.1 Wavelet domain analyses

Due to random nature of EEG data and interferences introduced during recording, it is very difficult to obtain distinctive characteristics from the time domain EEG data for different types of mental tasks. In view of obtaining better distinguishing behaviour of EEG signal for various mental tasks, one common approach is to divide the EEG data in various frequency bands and then carry out analysis in each band separately. In this regard, generally standard frequency bands corresponding to various states of vigilance (or activity) are considered for band-limited signal generation. These common bands are: delta (< 4 Hz), theta ($4 - 7$ Hz), alpha ($8 - 13$ Hz), beta ($14-20$ Hz), and gamma ($24-37$ Hz) [11]. However, in the current application of mental task classification, it may not be useful to restrict the EEG signal analysis only to these low frequency bands. The main reason behind the fact is that while performing mental tasks, relatively high frequencies remain active. In [12] , for the purpose of mental task classification, frequency band up to 100 Hz is used. As a result, in the proposed method, available all high frequency signals are taken into consideration. In order to decompose the EEG data, like some existing mental task classification techniques [24], wavelet analysis is utilized. It is well known that wavelet decomposition (WD), the most common time-frequency multi-resolution technique, is found very effective in EEG [24]. However, WD based scheme performs decomposition only in the lower frequency bands. As an alternative, wavelet packet decomposition (WPD) can be used where decomposition is performed both in lower and higher frequency regions. Moreover, it offers low computational cost and ease of implementation [22], [34].

A wavelet packet is represented as a function

$$W^{\phi}_{\psi,k}[n] = 2^{-\psi/2}W^{\phi}(2^{-\psi}n - k), \phi = 1, 2, ..., \psi^{m} \tag{4.1}$$

where parameters ϕ, ψ, k and m correspond to modulation, dilation, translation and level of decomposition in wavelet packet tree respectively [22], [34]. The following relationships are utilized to obtain the wavelet W^{ϕ} :

$$W^{2\phi} = \frac{1}{\sqrt{2}}\sum_{-\infty}^{\infty} h(k)W^{\phi}(\frac{n}{2} - k) \tag{4.2}$$

$$W^{2\phi+1} = \frac{1}{\sqrt{2}}\sum_{-\infty}^{\infty} g(k)W^{\phi}(\frac{n}{2} - k) \tag{4.3}$$

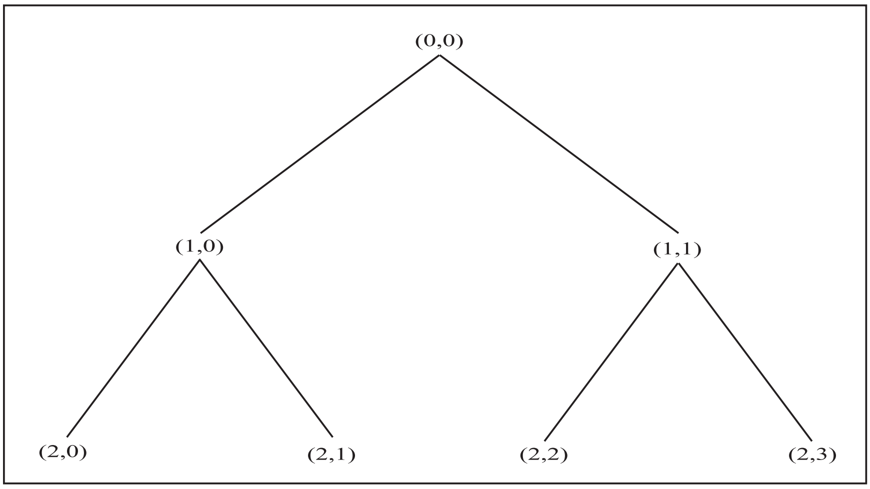

Figure 4.1: Tree decomposition of EEG signal

Here W^{ϕ} is called as a mother wavelet and the discrete filters $h(k)$ and $g(k)$ are quadrature mirror filters associated with the scaling function and the mother wavelet function. The filtering operations in the WPD result in a change in the signal resolution and the sub-sampling operation causes change in the scale. Thus, WPD helps in analyzing the signal at different frequency bands with different resolutions.

The wavelet packet coefficients $c^{\phi}_{\psi,k}$ corresponding to the signal $y[n]$ can be obtained as,

$$c^{\phi}_{\psi,k} = \sum_{-\infty}^{\infty} y[n] W^{\phi}_{\psi,k}[n] \tag{4.4}$$

provided the wavelet coefficients satisfy the orthogonality condition. Wavelet packet coefficients thus obtained at different levels can be used to reconstruct the original signal. However, wavelet packet coefficients at a particular node can also be used to reconstruct signal corresponding to that node, which is termed as WPNR signal in this Chapter. For example, in Fig. 4.1, where two level decomposition is considered, four different WPNR signals can be reconstructed from four nodes namely (2,0), (2,1), (2,2) and (2,3). In Fig. 4.2, a sample EEG signal is considered and shown in Fig. 4.2(a). Corresponding four WPNR signals are termed as WPNR(2,0), WPNR(2,1), WPNR(2,2) and WPNR(2,3) respectively.

One major problem in WPD is the rapid reduction of the length of wavelet coefficients in each decomposition level similar to conventional wavelet analysis. As a result, if features are extracted from multi-level wavelet coefficients, there is a chance of getting deteriorated feature quality due to reduced length of multi-level wavelet coefficients in comparison to main data length. On the contrary, if WPNR signal is considered at each particular node of decomposition level, because of retaining same length (without decreasing the coefficient length in each decomposition level), better statistical characteristics can be obtained. For example, for two level WPD of an N length data, as shown in Fig. 4.1, from four different nodes, by performing wavelet packet node reconstruction, one can obtain four different N length WPNR signals. On the contrary, the length of each WPD coefficient vector corresponding to those four nodes would be N/4. In comparison to N/4 length WPD coefficient vector, N length WPNR signal is expected to provide better statistical characteristics.

A large number of wavelet functions are available in the literature namely Daubechies, Symlets, Coiflet, Biorsplines, ReverseBior, and Discrete Meyer. Since the Daubechies family shows good performance in mental task related EEG signal [34], [22], db4 wavelets of the Daubechies family are utilized for feature extraction in the proposed method. Finally, various wavelet domain analyses are taken into consideration for the purpose of feature extraction, namely wavelet decomposition (WD), wavelet packet decomposition (WPD), wavelet node reconstruction (WNR) and wavelet packet node reconstruction (WPNR). In all the proposed methods, two levels of decomposition are taken into consideration.

4.1.2 Inter-channel relation

The amount of stimulation generated in different areas of brain depends on the nature of tasks performed. For example, visual tasks most likely stimulate the occipital region. Therefore, it is expected that EEG data obtained from different areas of brain while performing a particular task will not be same. This idea may be utilized in order to extract discriminative features for different types of mental tasks. Measuring inter-channel relationship may be an effective approach to compute this correlation feature. For this purpose, in the proposed method, correlation coefficient is computed between a pair of channels.

Correlation coefficient is one kind of measuring tool to quantify relationship between two or more signals. In this Chapter, it is utilized to obtain inter-channel correlation of i-th and j-th channel. Instead of directly using EEG data, correlation coefficient is obtained from various representations of wavelet data obtained from different channels. The inter-channel correlation coefficient $R_w(i, j)$ obtained from

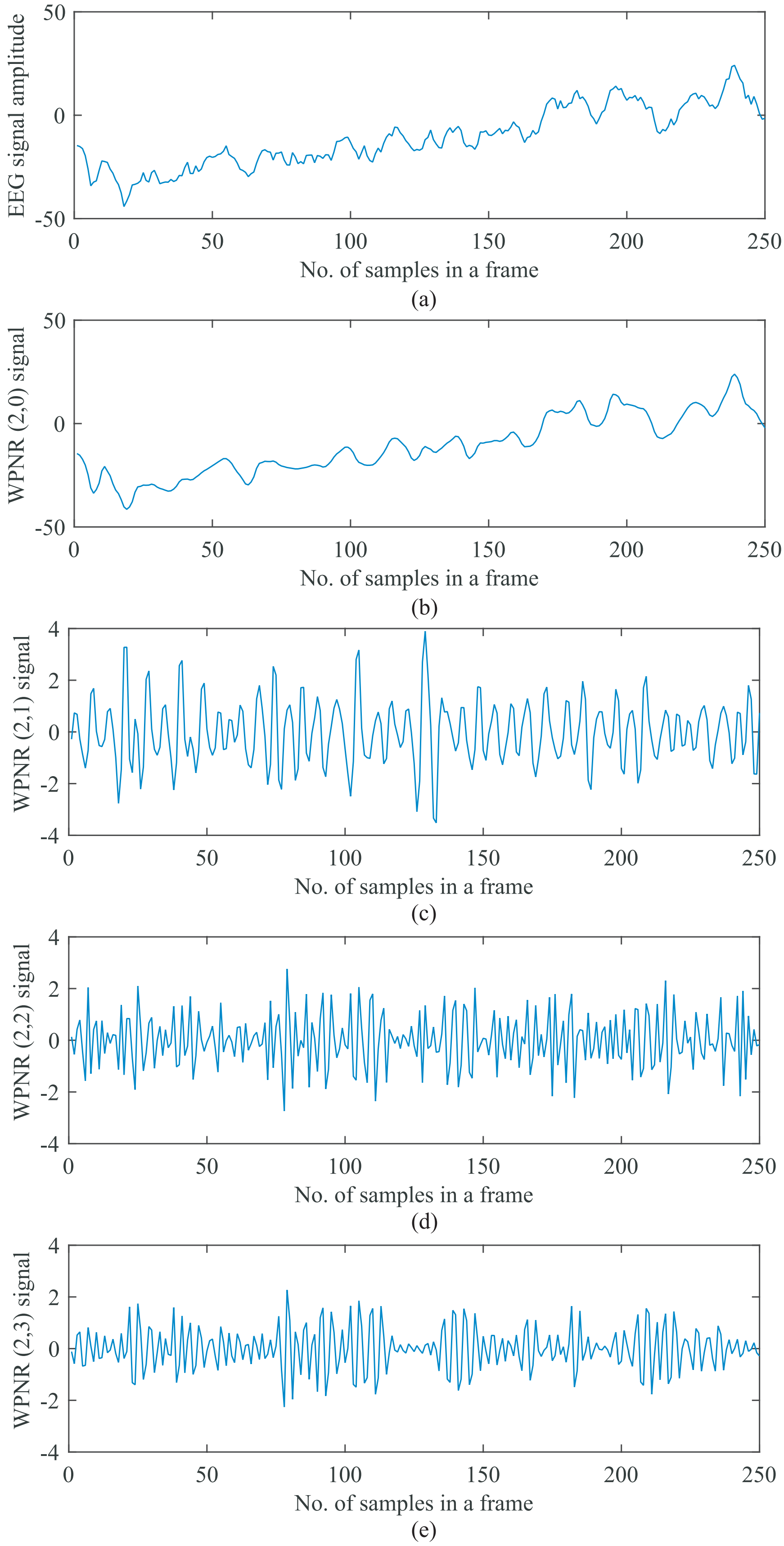

Figure 4.2: EEG signal and it's wavelet packet reconstructed signals

i and j-th electrode can be estimated as

$$R_w(i,j) = \frac{C_w(i,j)}{\sqrt{C_w(i,i)C_w(j,j)}} \tag{4.5}$$

where $C_w(i,j)$ is the (i,j)-th component of the covariance matrix $\mathbf{C_w}$ of the channel wavelet represented signals y_w^i and y_w^j, each consists of N samples. It is expressed as

$$\mathbf{C_w} \quad = \quad \begin{bmatrix} cov\langle y_w^i[n], y_w^i[n]\rangle & cov\langle y_w^i[n], y_w^j[n]\rangle \\ cov\langle y_w^j[n], y_{\psi,i}^i[n]\rangle & cov\langle y_w^j[n], y_w^j[n]\rangle \end{bmatrix} \tag{4.6}$$

The covariance of $y_w^i[n]$ and $y_w^j[n]$ denoted by $cov\langle y_w^i[n], y_w^j[n]\rangle$ is calculated considering the following formula

$$cov\langle y_w^i[n], y_w^j[n]\rangle = \frac{1}{N-1}\sum_{n=1}^{N}(y_w^i[n]-\mu_i)^\star(y_w^j[n]-\mu_j) \tag{4.7}$$

Here μ_i and μ_j indicate the mean of wavelet represented signals obtained from i-th and j-th channels, respectively and $\star$ denotes the complex conjugate. In the proposed method, all possible pair of i-th and j-th channels are taken into consideration to obtain inter-channel correlation coefficient of wavelet data, which is expected to provide maximum utilization of channel information. For six channels, if two of them are used to obtain inter-channel correlation coefficient, then total ${}^6\mathbf{C}_2 = 15$ combinations are plausible and all combinations namely 'C3-C4', 'C3-P3', 'C3-P4', 'C3-O1', 'C3-O2', 'C4-P3', 'C4-P4', 'C4-O1', 'C4-O2', 'P3-P4', 'P3-O1', 'P3-O2', 'P4-O1', 'P4-O2' and 'O1-O2' are taken into consideration. The reason behind choosing all possible channel pairs is to avoid selection of subset of channel pairs, which depends on empirical knowledge and intuition. Moreover, choosing all possible channel pairs also provides maximum channel information.

There are various advantages of utilizing correlation coefficient as feature. It is found that the effect of different types of external noises is reduced after cross correlation. Moreover, it provides bounded values. For the purpose of investigating the feature quality of correlation coefficient obtained for various tasks, a sample experiment considering multiplication and rotation task is performed. Inter-channel correlation coefficient is obtained from WPNR signal, which is termed as IC-WPNRCC. In Fig. 4.3, the box plot corresponding to IC-WPNRCCs obtained for fifteen different combinations of channels is presented. The boxplot indicates various statistical information, such as median, first and third quartiles and outliers of IC-WPNRCC.

There are thirty boxplots, each boxplot represents IC-WPNRCC measured from a particular combination of channel for a particular type of task.

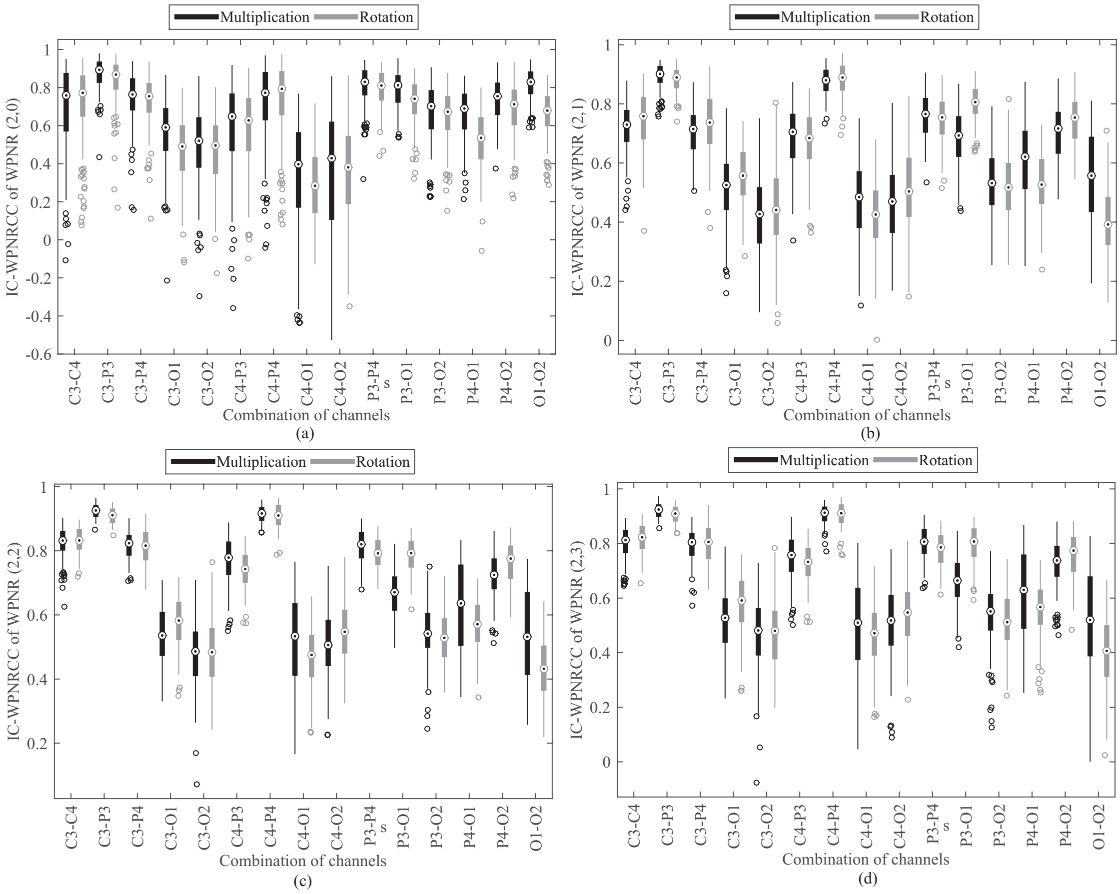

Figure 4.3: IC-WPNRCC obtained from WPNR signals of Subject 1

The presence of outliers decrease gradually from boxplot presented in Fig. 4.3(a) to boxplot presented in Fig. 4.3(d). The boxplot in Fig. 4.3(a) represents the IC-WPNRCC feature quality obtained from WPNR(2,0) signal which contains very low frequency information as shown in Fig. 4.2(b). Similarly, boxplot in Fig. 4.3(b)-Fig. 4.3(d) represent the IC-WPNRCC feature quality obtained from WPNR(2,1) to WPNR(2,3) signal. As discussed before, in mental task application, relatively high frequencies remain active. Therefore, comparatively better feature quality is obtained in high frequency WPNR signal than that obtained by low frequency WPNR signal as shown in Fig. 4.3.

4.1.3 Feature extraction

In the proposed method, at first, the raw EEG signal is preprocessed with a 60 Hz notch filter. To decompose EEG data corresponding to a channel, various wavelet domain analyses are taken into consideration, namely WD, WPD, WNR and WPNR as discussed in Sec. 4.1.1. Along with the proposed inter-channel correlation feature, a set of statistical parameters, namely root mean square(RMS), standard deviation and entropy are computed from all four wavelet representations. A method utilizing statistical features extracted from the WD is referred to as WDS and when inter-channel correlation coefficient is extracted from WD, it is termed as WDC. Similarly, WPDS and WPDC for WPD, WNRS and WNRC for WNR and WPNRS and WPNRC for WPNR can be obtained. Hence there are four different approaches to be used for feature extraction by the proposed method, namely

1. WDS+WDC
2. WPDS+WPDC
3. WNRS+WNRC
4. WPNRS+WPNRC.

For each of the four proposed methods, number of correlation coefficients obtained is $N_w \times^{N_c} \mathbf{C}_2$. Here, N_c indicates number of channels and N_w indicates number of wavelet represented signal. The number of features obtained from statistical parameters (std, rms and entropy) of each wavelet represented data for a test frame is $N_w \times (N_c + N_c + N_c)$. Finally the total feature dimension of the proposed method is $N_w \times ({}^{N_c}\mathbf{C}_2 + N_c + N_c + N_c)$.

4.1.4 Classification

Unless otherwise specified, polynomial kernel of SVM classifier is utilized to obtain classification accuracy in the proposed method. In all cases, leave-one-out cross validation scheme is employed to generate classification result, where each frame is tested one by one. During the testing of a frame, all the remaining frames are used for training. The overall accuracy is calculated based on the classification results obtained in all the frames using (2.15).

4.2 Simulation Results and Discussion

In this Section, performance of various feature extraction methods is investigated considering proposed four methods and some available existing methods. Moreover, performance of various feature extraction methods is investigated considering classification accuracy obtained under different conditions, such as utilizing different statistical parameters as feature, use of various EEG channel locations and effect of different kernel functions in SVM classifier. A comparative analysis on classification performance between the proposed method and some other methods is also performed.

In the proposed method, instead of directly using channel data, corresponding wavelet signals are used to extract inter-channel correlation feature and statistical parameters (std, rms and entropy). Polynomial kernel of SVM classifier is employed in leave one out cross-validation manner to obtain classification accuracy. The classification task is carried out considering two types of mental tasks at a time, as conventionally done by other researchers [11], [12]. In this way, ten different combinations of the five types of mental tasks, as mentioned in Sec. 2.1, are possible. In what follows, detail results and analyses are presented.

4.2.1 Effect of Different Statistical Feature

In the proposed method, as mentioned in Sec. 4.1, some statistical parameters are used as features, which are extracted from the various wavelet representations of the channel EEG signals. Similar to the previous two chapters, in this Chapter, effect of using conventional statistical features on classification accuracy is investigated considering same set of statistical parameters, namely avg, med, mod, max, min, std, rms, ent, skew and kurt. It is to be noted that the main objective of this Chapter is to demonstrate the efficacy of proposed correlation feature obtained from various wavelet domain signals. For that purpose, a sample experiment utilizing WPNR signal is considered. It is expected that the use of proposed IC-WPNRCC feature along with the conventional statistical features will offer better classification performance. In this regard, two different cases similar to the previous two chapters are considered:

1. Use of only statistical features
2. Use of proposed IC-WPNRCC feature along with statistical feature

In Fig. 4.4, classification accuracies considering the previously discussed two cases for the ten statistical features obtained for all subjects are shown. It is observed that

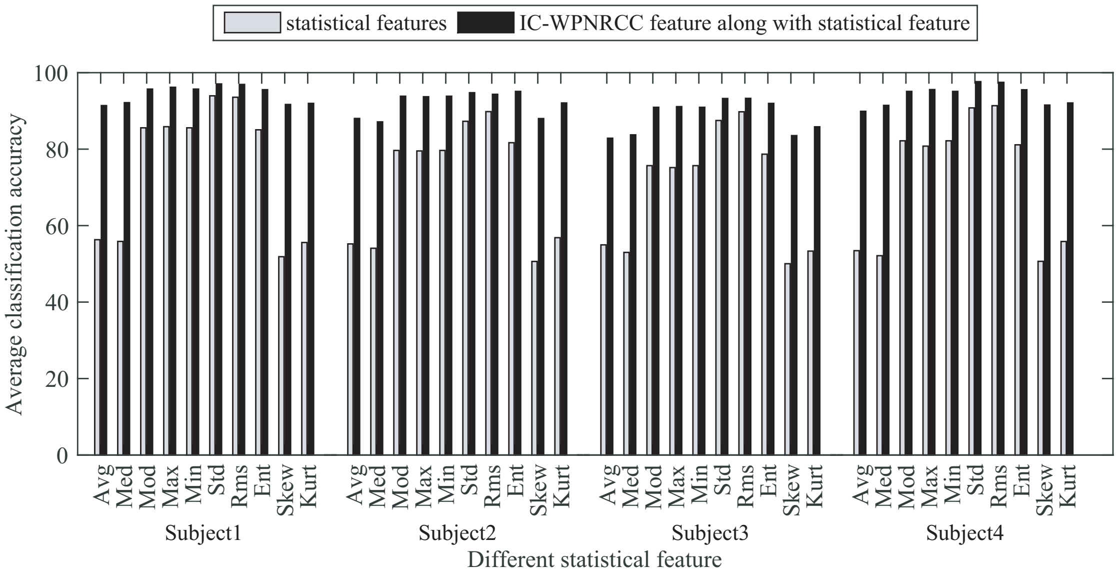

Figure 4.4: Effect of different statistical features of WPNR signals on classification accuracy for all four subjects.

classification accuracy increases if IC-WPNRCC is combined with channel statistical information of WPNR signals. Statistical features, such as std, rms and entropy of WPNR signals offer better classification performance similar to the two proposed methods described in the previous two chapters. Due to distinctive nature of std, rms and entropy, these three parameters are chosen in the feature vector along with proposed inter-WPNCC feature to classify mental tasks.

4.2.2 Effect of Kernel in SVM Classifier

The effect of using different kernels in SVM classifier on overall classification performance of the proposed method is thoroughly investigated similar to the previous two methods. In Fig. 4.5, average classification accuracies for 10 different combination of tasks by using three different kernels are plotted. It is to be noted that the features are extracted from WPNR signals. It is observed that the classification performances of polynomial kernel are consistently better in comparison to those obtained by linear and quadratic kernels in all cases as observed in the previous two chapters. For that purpose, polynomial kernel of SVM classifier is chosen to classify the tasks in the proposed method.

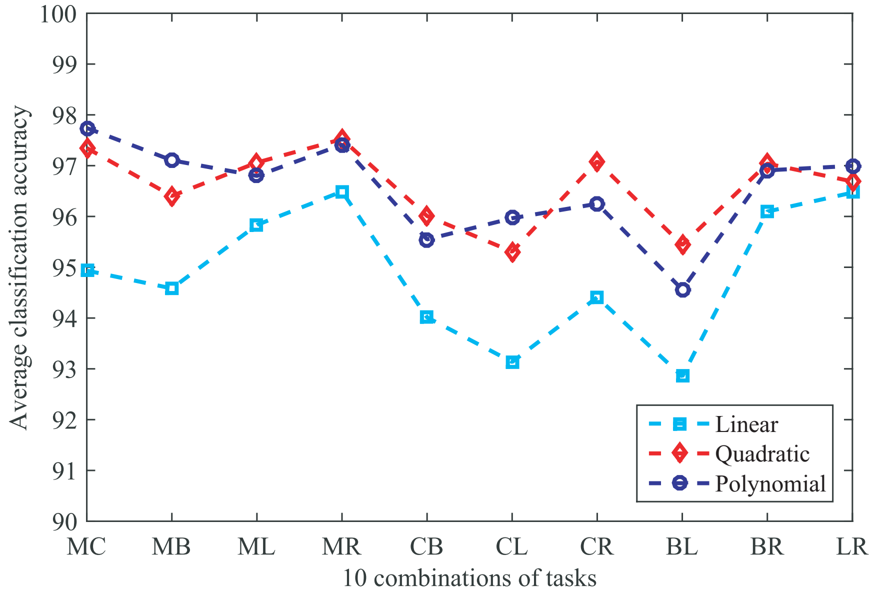

Figure 4.5: Classification accuracy obtained from WPNR signals considering different SVM kernels

4.2.3 Effect of variation of number of channel pairs

In the proposed method, all possible pairs of channels are taken into consideration to obtain the proposed inter-channel correlation feature so that maximum channel information can be utilized. However, similar to the previous two chapters, two different experiments is performed to observe the effect of variation of number of channel pairs. In Fig. 4.6, a comparative analysis among these experiments is presented in terms of classification accuracy. It is to be noted that these experiments are performed on WPNR signals. In these two experiments, reduced number of channels are utilized and lower classification accuracy compared to the proposed method is achieved. As a result, it is not possible to select any one particular choice of reduced number of channels to obtain acceptable classification performance in all subjects.

4.2.4 Effect of using inter-channel correlation feature

One of the objectives is to observe the effect of using proposed inter-channel correlation feature along with conventional statistical features on classification accuracy. In this regard, in Table 4.1, average classification accuracies obtained by various wavelet based proposed features as mentioned in Sec. 4.1.3 are reported to observe the effect of utilizing inter-channel correlation features. It is clearly observed from

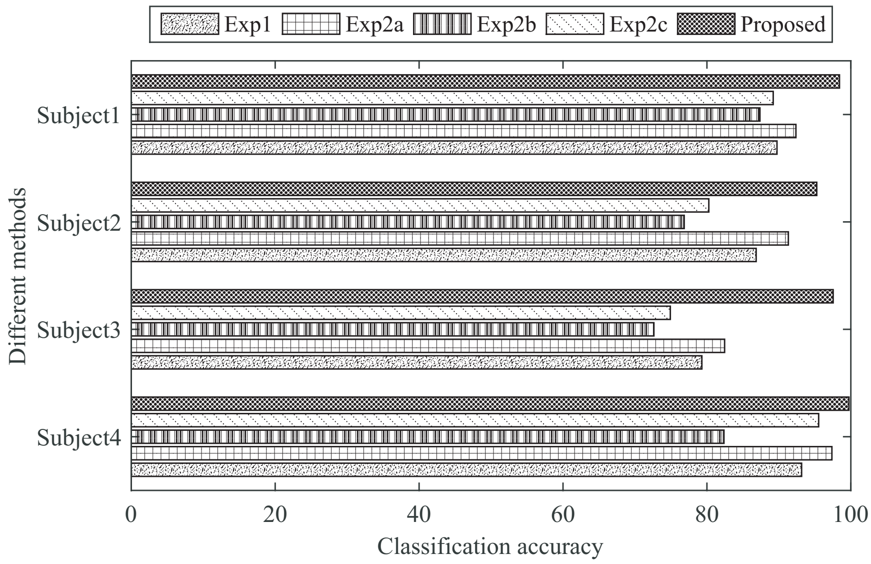

Figure 4.6: Effect of variation of number of channel pairs on proposed IC-WPNRC feature in terms of classification accuracy

Table 4.1: Effect of using inter-channel correlation feature on statistical features in terms of classification accuracy

	WD		WPD		WDNR		WPDNR	
	WDS	WDS+ WDC	WPDS	WPDS+ WPDC	WDNRS	WDNRS+ WDNRC	WPDNRS	WPDNRS+ WPDNRC
Sub-1	93.42	97.16	94.74	97.50	93.42	97.39	95.08	97.82
Sub-2	85.53	94.87	86.39	94.42	89.32	95.11	89.21	95.16
Sub-3	86.95	93.32	88.33	94.39	88.35	93.98	90.98	95.30
Sub-4	92.18	97.55	92.47	97.34	92.55	97.79	93.24	97.84

Table 4.1, use of wavelet domain inter-channel correlation feature drastically improves the classification accuracy than that obtained by statistical features in all four cases.

Classification accuracy of wavelet packet decomposed data is found higher than that of wavelet decomposed data. In all four proposed methods, if proposed inter-channel correlation feature is utilized along with statistical features, classification accuracy increases a lot. For example, for four proposed methods, average classification accuracies obtained from four subjects increase 6.2%, 5.43%, 5.16%, 4.4% respectively as shown in Fig. 4.7. It is found that the classification accuracies obtained from four proposed methods are almost similar. However,highest classification accuracy is obtained in WPNRS+WPNRC method.

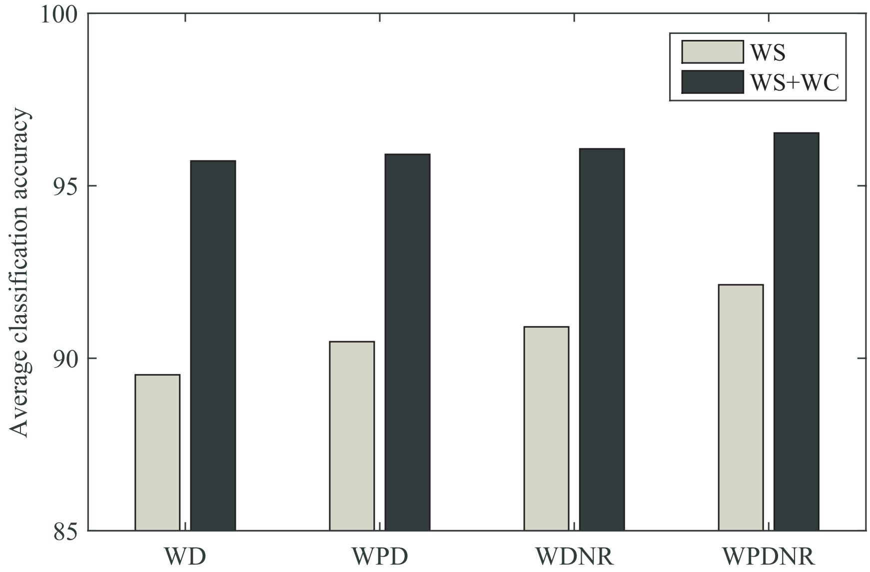

Figure 4.7: Effect of using inter-channel correlation feature on statistical features in terms of classification accuracy

4.2.5 Performance Comparison among Various Methods

With a view to investigate the classification performance, the proposed methods are compared with three existing methods, referred to as PAR5, PAR6 and WF8, in terms of classification accuracy. Among these three methods, the first two methods extract features utilizing power of spectral bands and asymmetry ratios [11], [12]. In PAR5 method, features are extracted from the traditionally used bands, namely delta (< 4 Hz), theta ($4 - 7$ Hz), alpha ($8 - 13$ Hz), beta ($14 - 20$ Hz) and gamma band ($23 - 37$ Hz). In PAR6 method, one additional band ($40 - 100$ Hz) along with these five bands are proposed. The last method utilizes wavelet features for classification purpose [24]. In WF8 method, eight features are extracted from each wavelet decomposed data, namely RMS, variance, Shannon entropy, Lempel-Ziv complexity measure, central frequency, maximum frequency, skewness and kurtosis.

For the purpose of performance evaluation, leave one out cross validation technique is carried out in all methods. In Tables 4.2-4.5, the classification accuracies obtained by using four different subjects are separately reported for four methods. In all cases, it is observed that the proposed feature extraction method outperforms other existing methods reported in this Chapter in terms of classification accuracy. The classification accuracy obtained for BL combination of Subject 4 is found 90.53% which is the lowest classification accuracy obtained among all combination of tasks

Table 4.2: Comparison of proposed WPNRS+WPNRC method with existing methods for Subject 1 in terms of classification accuracy

Task	PAR5 [11]	PAR6 [12]	WF8 [24]	WPNRS+WPNRC
MC	88.95	91.58	95.53	98.42
MB	83.95	87.37	95.26	98.68
ML	87.37	90.26	97.63	99.74
MR	91.05	93.95	98.42	99.74
CB	78.95	82.63	88.16	97.89
CL	69.74	77.11	84.74	95.53
CR	74.74	80.26	84.21	93.95
BL	71.05	82.89	80.26	97.63
BR	86.58	92.37	98.16	100.00
LR	77.37	81.84	90.79	96.58
Avg	80.97	86.03	91.32	97.82
Stddev	7.65	5.82	6.64	1.97

irrespective of subjects. However, if BL combination of mental task is excluded, the average classification accuracy of Subject 4 becomes 98.65% which is comparatively higher than any other subjects. In some combinations of mental tasks, it is observed that existing methods offer competitive classification performance with respect to proposed method. For example, in case of BR combination, almost similar classification accuracy is obtained in case of Subject 1 and Subject 4 for WF8 and proposed method. It is also found that for various combination of mental tasks, classification accuracy varies a lot in each reported existing method. For example, in PAR5 method, for Subject 1 and Subject 4, the standard deviation of classification accuracies for various subjects are found 7.65% and 6.09% compared to 1.97% and 2.80% of the proposed method. It is found that the classification performance obtained by the proposed method varies from subject to subject, but not at a very large scale. For Subject 2, the standard deviation obtained from different combination of mental tasks is found 1.17% which is the least among all four subjects. It is clearly observed that the proposed method offers consistently satisfactory classification accuracy in all cases irrespective of subjects and combination of mental tasks.

4.2.6 Computation Time

The average feature extraction time for a test frame is computed for the proposed methods described in Sec. 4.1.3 and the existing methods Sec. 4.2.5. The whole process of computation is performed using Intel(R) Core(TM) i5-4200M processor

Table 4.3: Comparison of proposed WPNRS+WPNRC method with existing methods for Subject 2 in terms of classification accuracy

Task	PAR5 [11]	PAR6 [12]	WF8 [24]	WPNRS+WPNRC
MC	76.58	83.42	80.26	95.26
MB	86.32	91.05	85.79	97.63
ML	83.42	87.63	85.00	95.00
MR	79.21	90.00	82.37	95.26
CB	80.53	92.11	78.95	95.26
CL	75.53	89.47	85.53	96.05
CR	73.42	87.11	80.26	93.95
BL	79.47	86.84	87.63	95.53
BR	79.74	86.05	81.84	93.95
LR	80.79	84.21	81.32	93.68
Avg	79.50	87.79	82.89	95.16
Stddev	3.74	2.85	2.90	1.17

Table 4.4: Comparison of proposed WPNRS+WPNRC method with existing methods for Subject 3 in terms of classification accuracy

Task	PAR5 [11]	PAR6 [12]	WF8 [24]	WPNRS+WPNRC
MC	74.91	79.82	80.70	97.54
MB	74.56	81.40	80.35	93.16
ML	74.39	80.70	80.70	93.51
MR	79.47	87.89	89.65	98.07
CB	72.98	81.05	78.42	91.93
CL	77.19	80.35	84.21	94.39
CR	75.44	81.40	88.07	97.37
BL	74.56	84.21	80.70	94.56
BR	75.96	84.21	87.19	94.21
LR	83.33	87.54	90.88	98.25
Avg	76.28	82.86	84.09	95.30
Stddev	3.05	2.96	4.51	2.29

Table 4.5: Comparison of proposed WPNRS+WPNRC method with existing methods for Subject 4 in terms of classification accuracy

Task	PAR5 [11]	PAR6 [12]	WF8 [24]	WPNRS+WPNRC
MC	90.53	97.63	95.53	99.74
MB	90.53	94.74	97.37	98.95
ML	88.42	92.89	94.21	98.95
MR	88.95	93.95	94.21	96.58
CB	82.37	86.58	96.05	97.11
CL	81.58	87.63	85.00	97.89
CR	92.89	96.32	94.21	99.74
BL	77.63	83.95	84.74	90.53
BR	95.26	97.11	99.21	99.47
LR	96.05	97.37	96.84	99.47
Avg	88.42	92.82	93.74	97.84
Stddev	6.09	4.99	4.94	2.80

Table 4.6: Feature dimension and average time for feature extraction for existing methods

Different methods	PAR5 [11]	PAR6 [12]	WF8 [24]
Feature dimension	75	90	192
Average time (ms)	60.02	76.36	732.38

Table 4.7: Feature dimension and average time for feature extraction for wavelet domain methods

Different methods	WDS	WDS+ WDC	WPDS	WPDS+ WPDC	WNRS	WNRS+ WNRC	WPNRS	WPNRS+ WPNRC
Feature dim.	54	99	72	132	54	99	72	132
Average time	17.09	25.26	56.29	70.23	29.74	37.87	86.97	103.93

with 2.50 GHz clock speed and 4 GB ram. The feature dimension and the feature extraction time for all eleven methods are listed in Tables 4.6-4.7.

The average feature extraction time increases due to use of proposed inter-channel correlation feature along with statistical features. For example, for four proposed methods, average feature extraction time of a test frame increase 8.17ms, 13.94ms, 8.13ms, 16.96ms respectively than that obtained in case of utilizing only statistical features. However, this slight increase in feature extraction time is very much applicable in real life application considering the drastic improve in classification accuracy obtained due to use of the proposed inter-channel correlation feature. It is observed that the feature extraction time for the proposed WDS+WDC method is very small compared to the existing PAR5 and PAR6 method despite having a larger feature dimension. All four proposed methods in various wavelet representations, namely WD, WPD, WNR and WPNR use a very small computation time for feature extraction compared to recently reported WF8 method. One of the reasons for such a small computation time is the smaller feature dimension of the proposed methods compared to WF8. For six channels and four wavelet packet data obtained from each channel signal utilizing level two decomposition, the feature dimension of the proposed WPDS+WPDC and WPNRS+WPNRC method is $4 \times 3 \times 6 + 4 \times {}^6\mathbf{C}_2 = 132$. Similarly, feature dimension of the proposed WDS+WDC and WNRS+WNRC method is $3 \times 3 \times 6 + 3 \times {}^6\,\mathbf{C}_2 = 99$, where three wavelet coefficient vectors are obtained from each channel signal utilizing level two decomposition. For four wavelet coefficient vectors and similar number of channels for each wavelet coefficient vector, the feature dimension of the WF8 is $4 \times 8 \times 6 = 192$. In case of PAR5 and PAR6, feature dimensions are $5 \times 3 \times 3 + 5 \times 6 = 75$ and $6 \times 3 \times 3 + 6 \times 6 = 90$ respectively.

4.3 Conclusion

In this Chapter, an efficient feature extraction scheme based on wavelet domain inter-channel correlation is proposed. Consistent classification performance is obtained irrespective of mental tasks and irrespective of subjects in comparison to some of the existing methods. It is to be noted that use of correlation feature drastically improves classification accuracy in all cases, namely WD, WPD, WNR and WPNR. Although feature dimension increases due to use of inter-channel correlation feature along with intra-channel statistical features (std, rms and entropy) , it is still implementable in real life application and acceptable considering the achievement in classification accuracy. Classification performance for various feature extraction methods are investigated considering polynomial kernel of SVM classifier. Results obtained from various types of investigation verify that the proposed mental task classification scheme is capable of classifying EEG signals with high classification accuracy.

Chapter 5

Conclusion

5.1 Concluding Remarks

In this Thesis, three different approaches for mental task classification utilizing inter-channel relationship of EEG signal are presented. Inter-channel correlation coefficient of each decomposed EEG signal is utilized to explore the relationship between channels, which is referred to as inter-IMFCC, inter-SBCC and IC-WPNRCC. Moreover, intra-channel features, such as standard deviation, rms and entropy of each decomposed signal are also measured. Finally, both inter-channel features and intra channel features of each decomposed EEG signal are utilized to form feature vector and a quite satisfactory classification performance is achieved. Effect of using conventional statistical features on classification accuracy is investigated considering ten widely used higher and lower order statistical parameters. Moreover, effect of selecting different combinations of channels is investigated and it is observed that considering all combinations of channels provide the best classification performance irrespective of the decomposition techniques. Classification performance for various feature extraction methods are listed considering polynomial kernel. Results obtained from various types of investigation verify that the proposed mental task classification schemes are capable of classifying EEG signals with high classification accuracy.

5.2 Contributions of This Thesis

The major contribution of the thesis are summarized below:

- One of the main contributions of this work is to show the effective use of inter-channel relationship in mental task classification. It is shown that use of cross correlation between various pair of channel data (or decomposed data) can drastically improve the classification accuracy. It is considered that for different types of task, different channels corresponding to different parts of the brain are actuated. Measuring inter-channel relationship in some efficient spectro-temporal domains plays a significant role to cover the spatial and temporal relationship between different channels. Correlation coefficient is utilized to measure the inter-channel relationship of decomposed EEG signal. It is to be noted that in determining the inter-channel relationship, no prior channel selection is required, rather all possible channel combinations are found to provide better results.

- Next it is shown that inter-channel relationship can be better exploited in case of decomposed data domain. In this case, various time-frequency domain decompositions are tested and inter-channel features are proposed utilizing spectral band decomposition, EMD and wavelet decomposition.

- Statistical analysis of the proposed correlation features namely inter-IMFCC, inter-SBCC and IC-WPNRCC is carried out. It is observed that extracted features offer high within class compactness and between class separability.

- The performance of the proposed methods have been investigated based on leave one out cross validation scheme along with the state-of-the art comparison methods. In all classification cases, proposed methods have the superior accuracy with the state-of-the art comparison methods. Results obtained from various types of investigation verify that the three proposed mental task classification methods are capable of classifying EEG signals with consistently high classification accuracy.

5.3 Scopes for Future Work

In this Thesis, three effective feature extraction methods for mental task classification are developed. However, there are some scopes for future research. In this research, we use a popular EEG database of mental tasks, which consists of five class

EEG data. The proposed methods can classify those mental tasks with highest accuracy using different decomposition techniques. In future, the proposed feature extraction methods can be utilized in other applications of EEG based system, such as classification of motor-imagery tasks, alcoholic and non-alcoholic persons, fatigue and non-fatigue condition, drowsiness etc.

Bibliography

[1] Y. Wang, X. Gao, B. Hong, C. Jia, and S. Gao, "Brain-computer interfaces based on visual evoked potentials," *IEEE Engineering in medicine and biology magazine*, vol. 27, no. 5, 2008.

[2] T. Kayikcioglu and O. Aydemir, "A polynomial fitting and k-nn based approach for improving classification of motor imagery BCI data," *Pattern Recognition Letters*, vol. 31, no. 11, pp. 1207–1215, 2010.

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

[4] M. Grosse-Wentrup, C. Liefhold, K. Gramann, and M. Buss, "Beamforming in noninvasive brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1209–1219, 2009.

[5] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[6] S. Siuly and Y. Li, "Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 526–538, 2012.

[7] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields.* Oxford University Press, USA, 1995.

[8] H.-I. Suk and S.-W. Lee, "A novel Bayesian framework for discriminative feature extraction in brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 286–299, 2013.

[9] X. Yong, R. K. Ward, and G. E. Birch, "Sparse spatial filter optimization for EEG channel reduction in brain-computer interface," in *Proc. ICASSP*, 2008, pp. 417–420.

[10] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial patterns with generic learning for EEG signal classification," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2009, pp. 6599–6602.

[11] R. Palaniappan, "Utilizing gamma band to improve mental task based brain-computer interface design," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 3, pp. 299–303, 2006.

[12] L. Zhang, W. He, C. He, and P. Wang, "Improving mental task classification by adding high frequency band information," *J. Med. Syst.*, vol. 34, no. 1, pp. 51–60, 2010.

[13] R. Ameri, A. Pouyan, and V. Abolghasemi, "EEG signal classification based on sparse representation in brain computer interface applications," in *2015 22nd Iranian Conference on Biomedical Engineering (ICBME).* IEEE, 2015, pp. 21–24.

[14] N.-Y. Liang, P. Saratchandran, G.-B. Huang, and N. Sundararajan, "Classification of mental tasks from EEG signals using extreme learning machine," *Int. J. Neural Syst.*, vol. 16, no. 01, pp. 29–38, 2006.

[15] C. W. Anderson, E. A. Stolz, and S. Shamsunder, "Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 3, pp. 277–286, 1998.

[16] D. Vidaurre, C. Bielza, and P. Larrañaga, "Classification of neural signals from sparse autoregressive features," *Neurocomputing*, vol. 111, pp. 21–26, 2013.

[17] D. P. Burke, S. P. Kelly, P. de Chazal, R. B. Reilly, and C. Finucane, "A parametric feature extraction and classification strategy for brain-computer interfacing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 1, pp. 12–17, 2005.

[18] Q. Wang and O. Sourina, "Real-time mental arithmetic task recognition from EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 2, pp. 225–232, 2013.

[19] Y. Kumar, M. Dewal, and R. Anand, "Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network," *Signal, Image and Video Processing*, vol. 8, no. 7, pp. 1323–1334, 2014.

[20]

[21] L. Guo, D. Rivero, and A. Pazos, "Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks," *Journal of neuroscience methods*, vol. 193, no. 1, pp. 156–163, 2010.

[22] X. Li, X. Chen, Y. Yan, W. Wei, and Z. J. Wang, "Classification of EEG signals using a multiple kernel learning support vector machine," *Sensors*, vol. 14, no. 7, pp. 12 784–12 802, 2014.

[23] A. Gupta and R. Agrawal, "Relevant feature selection from EEG signal for mental task classification," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2012, pp. 431–442.

[24] A. Gupta, R. Agrawal, and B. Kaur, "Performance enhancement of mental task classification using EEG signal: a study of multivariate feature selection methods," *Soft Computing*, vol. 19, no. 10, pp. 2799–2812, 2015.

[25] M. T. F. Talukdar, S. K. Sakib, C. Shahnaz, and S. A. Fattah, "Motor imagery EEG signal classification scheme based on entropy of intrinsic mode function," in *Electrical and Computer Engineering (ICECE), 2014 International Conference on*. IEEE, 2014, pp. 703–706.

[26] M. Hariharan, V. Vijean, R. Sindhu, P. Divakar, A. Saidatul, and S. Yaacob, “Classification of mental tasks using stockwell transform,” *Computers & Electrical Engineering*, vol. 40, no. 5, pp. 1741–1749, 2014.

[27] K. Polat and S. Güneş, “Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform,” *Applied Mathematics and Computation*, vol. 187, no. 2, pp. 1017–1026, 2007.

[28] W. J. Williams, H. P. Zaveri, and J. C. Sackellares, “Time-frequency analysis of electrophysiology signals in epilepsy,” *Engineering in Medicine and Biology Magazine, IEEE*, vol. 14, no. 2, pp. 133–143, 1995.

[29] H. P. Zaveri, R. B. Duckrow, N. C. De Lanerolle, and S. S. Spencer, “Distinguishing subtypes of temporal lobe epilepsy with background hippocampal activity,” *Epilepsia*, vol. 42, no. 6, pp. 725–730, 2001.

[30] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, “Epileptic seizure detection in EEGs using time–frequency analysis,” *Information Technology in Biomedicine, IEEE Transactions on*, vol. 13, no. 5, pp. 703–710, 2009.

[31] P. P. Acharjee and C. Shahnaz, “Multiclass epileptic seizure classification using time-frequency analysis of EEG signals,” in *7th International Conference on Electrical & Computer Engineering (ICECE)*, 2012, pp. 260–263.

[32] M. Rahman and S. Fattah, “Mental task classification scheme utilizing correlation coefficient extracted from inter-channel intrinsic mode function,” *BioMed Research International*, in press.

[33] Z. A. Keirn and J. I. Aunon, “A new mode of communication between man and his surroundings,” *IEEE Trans. Biomed. Eng.*, vol. 37, no. 12, pp. 1209–1214, 1990.

[34] D. Wang, D. Miao, and C. Xie, “Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection,” *Expert Systems with Applications*, vol. 38, no. 11, pp. 14 314–14 320, 2011.