UNIVERSITY OF SOUTHERN QUEENSLAND, AUSTRALIA

ANALYSIS AND CLASSIFICATION OF EEG SIGNALS

A Dissertation Submitted by

Siuly

For the Award of

Doctor of Philosophy

July, 2012

Abstract

Electroencephalography (EEG) is one of the most clinically and scientifically exploited signals recorded from humans. Hence, its measurement plays a prominent role in brain studies. In particular, the examination of EEG signals has been recognized as the most preponderant approach to the problem of extracting knowledge of the brain dynamics. EEG recordings are particularly important in the *diagnosis of epilepsy* and in *brain computer interface (BCI)*. In BCI systems, EEG signals help to restore sensory and motor functions in patients who have severe motor disabilities. Analysing EEG signals is very important both for supporting the diagnosis of brain diseases and for contributing to a better understanding of cognitive process.

Although EEG signals provide a great deal of information about the brain, research in classification and evaluation of these signals is limited. Even today the EEG is often examined manually by experts. Therefore, there is an ever-increasing need for developing automatic classification techniques to evaluate and diagnose neurological disorders. Classification techniques can help to differentiate EEG segments and to decide whether a person is healthy. A big challenge is for BCI systems to correctly and efficiently identify different EEG signals of different motor imagery (MI) tasks using appropriate classification algorithms to assist motor disabled patients in communication.

In this dissertation, we aim to develop methods for the analysis and classification of epileptic EEG signals and also for the identification of different categories of MI tasks based EEG signals in BCI's development.

In order to classify epileptic EEG signals, we propose two methods, simple sampling technique based least square support vector machine (SRS-LS-SVM) and clustering technique based least square support vector machine (CT-LS-SVM). The experimental results show that both algorithms perform well in the EEG signal classification and the CT-LS-SVM method takes much less execution time compared to the SRS-LS-SVM technique. The research findings also indicate that the proposed approaches are very efficient for classifying two categories of EEG signals. This

research can help to provide clinical information about patients who have epilepsy, neurological disorders, mental or physiological problems.

In BCI systems, if the MI tasks are reliably distinguished through identifying typical patterns in EEG data, motor disabled people could communicate with a device by composing sequences of these mental states. In this dissertation, for the identification of MI tasks in BCI applications, we developed three methods:

(1) Cross-correlation based logistic regression (CC-LR).
(2) Modified CC-LR with diverse feature sets.
(3) Cross-correlation based least square support vector machine (CC-LS-SVM).

The experimental results have demonstrated the effectiveness of the methods for the identification of MI tasks. These techniques can assist clinical diagnoses and rehabilitation tasks.

Finally we investigated two issues for the MI classification:
(1) Which algorithm performed better.
(2) Which EEG data is more suitable for getting information about MI tasks.
Is it the motor area data or the all-channels data?

To answer these two questions, we considered the three algorithms: the CC-LS-SVM, the CC-LR and the cross-correlation based kernel logistic regression (CC-KLR). Based on the experimental results, we concluded that the CC-LS-SVM algorithm is the best algorithm for the MI tasks EEG signal classification, and the all-channels EEG data can provide better information than the motor area EEG data for the MI tasks classification. Furthermore, the CC-LS-SVM approach can correctly identify the discriminative MI tasks, demonstrating the algorithms superiority in the classification performance over other existing methods.

Certification of Dissertation

I hereby declare that the work presented in this dissertation is my own and is, to the best of my knowledge and belief, original except as acknowledged in the text. This dissertation has not previously been submitted either in whole or in part for a degree at this or any other university.

Signed
(Siuly)
Date:

Signed Date:
(Principal Supervisor)
(Associate Professor Yan Li)

Signed Date:
(Associate Supervisor)
(Associate Professor Peng (Paul) Wen)

Acknowledgements

First of all, I would like to thank almighty Allah, for His guidance and strength.

I would like to express my sincere gratitude and appreciation to my supervisor, Associate Professor Yan Li, for her continuous inspiration, support, guidance, encouragement, advice, patience, and individual feedback throughout the course of my PhD study. She was an approachable supervisor who motivated me to achieve my goals. I feel very grateful and blessed to have worked under her supervision. I would not imagine carrying out this research without her constant support and help. I would also like to express my sincere gratitude and appreciation to my associate supervisor Associate Professor Peng (Paul) Wen, for his continuous encouragement, support, valuable advice, and suggestions.

I gratefully acknowledge the University of Southern Queensland (USQ) for offering me the 2009 USQ Postgraduate Research Scholarship to perform this research successfully, without which, this work would not have been possible.

I would also like to thank Professor Mark Sutherland, Director of Centre for Systems Biology (CSBi) for granting me half-tuition fees support, and supporting me to attend conferences.

I would also like to thank to all of the staff in the Department of Mathematics and Computing for their co-operation. I would also like to express my gratitude and appreciation to my postgraduate colleagues in our research group for their discussions, comments and advice, especially during our group seminars. I would also like to thank all of my friends for their encouragement and supports.

I wish to express my appreciation to my honourable parents, loving brothers and sisters for their support. Last and most importantly, I wish to thank my husband, Dr. Enamul Kabir who stands beside me and encourages me constantly for every step of my work in this thesis. My thanks are also to my son, Shadman Enam Srijon for giving me happiness, joy and enduring love. I would like to dedicate this dissertation to my son, my beloved husband and my respectable parents.

Contents

List of Figures

List of Tables

Publications Based on This Dissertation

1. Siuly, Li, Y., (2012) 'Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain computer interface', *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* DOI (identifier) 10.1109/TNSRE.2012.2184838, in press. (ERA A*, 2010)

2. Siuly, Li, Y. and Wen, P. (2011) 'Clustering technique-based least square support vector machine for EEG signal classification', *Computer Methods and Programs in Biomedicine*, Vol. 104, no. 3, pp. 358-372. (ERA A*, 2010)

3. Siuly, Li, Y. and Wen, P. (2011) 'Identification of Motor Imagery Tasks through CC-LR Algorithm in Brain Computer Interface', *International Journal of Bioinformatics Research and Applications* for publication, 2011 (in press). (ERA B, 2010)

4. Siuly, Li, Y. and Wen, P. (2011) 'EEG signal classification based on simple random sampling technique with least square support vector machines', *International journal of Biomedical Engineering and Technology*, Vol. 7, no. 4, pp. 390-409. (ERA C, 2010)

5. Siuly, Li, Y. and Wen, P. (2010) 'Analysis and classification of EEG signals using a hybrid clustering technique', *The proceedings* of t*he 2010 IEEE/ICME International Conference on Complex Medical Engineering (CME2010).* pp. 34–39.

6 Siuly, Li, Y. and Wen, P. (2009) 'Classification of EEG signals using Sampling Techniques and Least Square Support Vector Machines', *The proceedings of Fourth International Conference on Rough Sets and Knowledge Technology (RSKT 2009),* LNCS 5589 (2009), pp. 375-382.

7. Siuly, Li, Y., Wu, J. and Yang, J. (2011) 'Developing a Logistic Regression Model with Cross-Correlation for Motor Imagery Signal Recognition', *The proceedings of the 2011 IEEEIICME International Conference on Complex Medical Engineering May 22 - 25, Harbin, China*, pp. 502-507.

8. Siuly, Li, Y. and Wen, P., (2011) 'Comparisons between Motor Area EEG and all-Channels EEG for Two Algorithms in Motor Imagery Task Classification', submitted to the *International Journal of Intelligent Systems Technologies and Applications* for publication, February 2012. (ERA B, 2010)

9. Siuly, Li, Y. and Wen, P., (2012) 'Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain computer interface', submitted to the Journal of *Applied Mathematics and Computation,* February 2012. (ERA A, 2010)

CHAPTER 1
INTRODUCTION

The human brain is the control centre of the body. It is responsible for perception, cognition, attention, emotion, memory and action (Carlson, 2002a; Purves et al., 2004). When a person is thinking, reading or watching television different parts of the brain are stimulated. This creates electrical signals, which, together with chemical reactions, let the parts of the body communicate. These electrical signals can be monitored through scientific techniques such as electroencephalography (EEG), electrocorticography (ECoG), magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), positron emission tomography (PET) and single photo emission computed tomography (SPECT). The aforementioned scientific techniques help us to gain a better understanding of what role the human brain plays whilst an individual is performing a number of tasks. EEG is the most used technique to capture brain signals due to its excellent temporal resolution, non-invasiveness, usability, and low set-up costs (Blankertz, et al., 2008; Grosse-Wentrup et al. 2009). An EEG can show what state a person is in, whether in sleep, awake, anaesthetized, because the characteristic patterns of the electrical potentials differ for each of these states.

EEG is becoming increasingly important in the diagnosis and treatment of mental and brain neuro-degenerative diseases and abnormalities. The analysis and classification of EEG signals are essential both for supporting the diagnosis of brain diseases and for contributing to a better understanding of cognitive process. The main purpose of a classification is to separate EEG segments and to decide whether people are healthy or to estimate the mental state of a subject related to a performed task. Usually huge amounts of data are generated by EEG and visual inspection for discriminating EEGs is a time consuming, error prone, costly process and not sufficient enough for reliable information. Hence, developing automatic classification methods for EEG is vital to ensure the proper evaluation and treatment of neurological diseases (Agarwal et al., 1998). In this dissertation, we focus on the classification of EEG signals in the two most important areas, *epilepsy* and *brain*

computer interface (BCI). We propose two algorithms for the classification of epileptic EEG signals and three algorithms for the identification of motor imagery (MI) task based EEG signals in BCI developments.

These proposed methods can distinguish different categories of EEG signals and provide valuable information about brain states. Our proposed methods will be useful for neurologists to identify the brain diseases correctly and efficiently using the typical patterns of EEG signals. The outcomes of this study will also help to improve the quality of life of patients with brain disorders.

1.1 Overview and Motivation of the Study

The electroencephalography (EEG) is a recording of electrical activity along the scalp generated by the cerebral cortex nerve cells of the brain (Niedermeyer and Silva, 2005). The waveforms recorded are thought to reflect the activity of the surface of the brain. These signal's parameters and patterns indicate the health state of the brain. In Chapter 2, a detailed discussion about EEG is provided. EEG is one of the most clinically and scientifically exploited signals recorded from humans. Hence, its quantification plays a prominent role in brain studies. In particular, the assessment of EEG signals has been very popular to recognize the problem of extracting the knowledge of the brain dynamics and thus it has been heavily exploited in the study of the brain phenomena related to motor imagery.

The two main applications of EEG recordings are in *epilepsy* and *brain computer interface (BCI)*. Epilepsy is a common chronic neurological disorder characterized by seizures (Blume et al., 2001). Seizures are defined as sudden changes in the electrical functioning of the brain. A description about epilepsy and epileptic seizures is provided in Chapter 2. Epileptic seizures result from a temporary electrical disturbance of the brain, and epileptic activity can create clear abnormalities on a standard EEG. In BCI systems, EEG signals provide an effective way to help people who have severe motor disabilities to communicate with the outside world just by brain signals. A BCI system directly measures a brain activity associated with user's *intent* and translates the recoded brain activity into corresponding control signals. This translation involves signal processing and pattern recognition techniques. A BCI can only detect and classify specific patterns of an activity through the continuous brain signals that are associated with specific tasks or

events. In recent years, increasing attention has been devoted to the classification of motor imagery (MI) task problems related to BCI applications as MI activities represent an efficient mental strategy to operate a BCI system. A MI task may be seen as a mental rehearsal of a motor act such as movements of hands, feet, fingers and tongue without any overt motor activity (Kayikcioglu and Aydemir, 2010; Thomas et al., 2009). Each MI task is usually treated as a class of data category. A detailed discussion of BCI systems can be found in Chapter 2.

Since EEG signals provide significant contributions in biomedical science, a careful analysis of the EEG records are needed to provide valuable insight and to improve understanding about it. One challenge in the current biomedical research is how to classify time-varying electroencephalographic (EEG) signals as accurately as possible. Several classification methods are reported to identify different neurological diseases and also to recognize diverse mental states of disabled people using the typical patterns of the EEG signals. Currently, the classification of EEG signals in epileptic activities and MI tasks based BCIs are still far from being fully understood. A considerable amount of neuroscience research is still required to achieve this goal. Hence, in this dissertation, we aim to develop methods to classify the brain activities in these two areas.

1.2 Problem Statements

When measuring an EEG we often have large amounts of data with different categories, in particular if the recordings are done over a long time period. To extract information from such a large amount of data, automated methods are needed to analyse and classify the data by appropriate techniques. Although EEG recordings contain valuable information about the function of the brain; the classification and evaluation procedures of these signals have not been well developed. The evaluation of an EEG recording is usually conducted by experienced electroencephalographers who visually scan the EEG records (Kutlu et al., 2009; Subasi & Ercelebi, 2005). Visual inspection of EEG signals is not a satisfactory procedure because there are no standard criteria for the assessments and it is time-consuming process that can often result in errors due to fatigue. Therefore, there is a need to develop automatic systems for classifying the recorded EEG signals. As BCIs aim to translate an activity of the brain into a command to control an external device completing the

task of a communication, it is a big challenge for BCI systems to properly and efficiently recognize the intension's patterns of the brain using appropriate classification algorithms.

The design of an effective classification system is complex given the high variability of the EEG signals in the presence of different subjects and target events (classes). In recent years, a variety of computerized-analysis methods have been developed to extract the relevant information from EEG recordings and identify different categories of EEG data. From the literature, it's observed that most of the reported methods had a limited success rate. Some methods took more time to perform the required computation work and others were very complex for practical applications. Some methods used small sample setting (SSS) data points as the representative of a large number of data points of EEG recordings. Generally they were not representative enough for the EEG signal classification. On the other hand, in most of the cases, the reported methods did not select their parameters using a suitable technique, although the parameters significantly affect the classification performance. To our knowledge, the researchers did not investigate which EEG data (motor area EEG data or all-channels EEG data) is better for providing more information about the MI tasks signal classification.

To explore these issues, in this dissertation, two methods were proposed for the epileptic EEG signals classification and these were also implemented on mental imagery tasks EEG data to evaluate performances. A further three algorithms were introduced for the classification of MI tasks EEG signals in the BCIs development. Finally, we investigated which algorithm and which EEG data are able to provide more information during the classification of MI EEG signals.

1.3 Contributions of the Dissertation

The work presented in this dissertation focuses on how different EEG signals from different brain activities can be classified to analyse different brain disorders. In this dissertation, we have developed two techniques for classifying EEG signals in the epileptic diagnosis and three techniques for the identification of different categories of MI tasks in BCI applications. The main objective of this study is to develop methods and techniques for identifying different EEG signals from different brain

activities. To investigate the performances of those techniques, we also compare our proposed algorithms with other recently reported algorithms. This manuscript describes the work we carried out in order to address the three objectives:

1: Developing methods for the classification of the epileptic EEG signals which improve the classification rate with less execution time.

2: Introducing methods to identify EEG signals during MI tasks.

3: Investigating which algorithm and which EEG data (motor area data or all-channels data) are better for the MI signal classification.

A brief discussion of these three points is provided below.

1: Developing methods for the classification of the epileptic EEG signals

In order to classify the epileptic EEG signals, we developed two new approaches:

(1) Simple random sampling technique based least square support vector machine (SRS-LS-SVM).

(2) Clustering technique based least square support vector machine (CT-LS-SVM).

For the experimental evaluation, these two algorithms were implemented on *EEG epileptic data*. In order to further evaluate the performance of those two methods, the SRS-LS-SVM algorithm was employed on *mental imagery tasks EEG data* and *Ripley data,* and the CT-LS-SVM method on *mental imagery tasks EEG data* and *motor imagery EEG data*.

Method 1 (SRS-LS-SVM): In this method, we utilized a two-stage simple random sampling (SRS) procedure to obtain representative features from original EEG data. Then we used these features as the input to the least square support vector machine (LS-SVM) to classify two classes of the EEG signals. In the two-stage SRS technique, the sample and sub-sample sizes were determined using a sample size calculator of the "Creative Research System" considering 99-100% confidence interval and 99% confidence level. The experimental results demonstrate that the proposed approach has significant advantages over many other existing methods.

Method 2 (CT-LS-SVM): To make a more robust method with less execution time, we developed a clustering technique based least square support vector machine (CT-LS-SVM) method for the classification of epileptic EEG signals.

In this research, we evaluated the performance of the CT-LS-SVM against the SRS-LS-SVM on the basis of the accuracy and the execution time. The experimental results showed that the CT-LS-SVM approach performed better than the SRS-LS-SVM method with respect to accuracy, and the CT-LS-SVM algorithm takes much less time than the SRS-LS-SVM. The classification performances of the CT-LS-SVM method are also higher than the recently reported algorithms. This method helps neurologists and researchers to identify the epileptic EEG signals.

2: Introducing methods to identify EEG signals during MI tasks

In this dissertation, we further developed three algorithms for the classification of MI tasks EEG signals:

(1) Cross-correlation based logistic regression (CC-LR).

(2) Modified CC-LR with diverse feature sets.

(3) Cross-correlation based least square support vector machine (CC-LS-SVM).

The proposed methods were tested on two benchmark datasets, IVa and IVb of BCI Competition III, and the performances of each method were evaluated through a *k*-fold cross-validation procedure.

Method 3 (CC-LR): We developed a CC-LR algorithm combining cross-correlation (CC) technique and logistic regression (LR) classifier. In this algorithm, we used a cross-correlation (CC) technique for the feature exaction, and a logistic regression (LR) classifier for the classification of those MI features. We considered randomly the signal of Fp1 electrode position (in the international 10-20 system) as a reference signal for the CC technique.

Method 4 (modified CC-LR with diverse feature sets): We developed a modified vision of the CC-LR algorithm with three diverse feature sets as the CC-LR algorithm could not provide a clear idea whether the feature set and the reference signal are the optimal choices. In this algorithm, we investigated which features perform better in the MI tasks classification. We also provided an insight into how to select a reference channel for the CC technique considering biological structure of the brain and the international 10-20 system electrode placement system with EEG signals. Then we determined the C3 electrode signal as the reference signal for the CC method in this study.

Method 5 (CC-LS-SVM): To improve the separability of the MI tasks classification, we developed novel algorithm called CC-LS-SVM, tuning hyper parameters of the LS-SVM method. In this approach, we applied a cross-correlogram based feature extraction procedure for the MI tasks signals, and developed a LS-SVM for the classification of the extracted MI features. To verify the effectiveness of the LS-SVM classifier, we replaced the LS-SVM classifier by a logistic regression (LR) classifier and a kernel logistic regression (KLR) classifier, separately, with the same features extracted from the cross-correlation technique. As the parameters of the LS-SVM can significantly affect the classification performance, we used a two-step grid search algorithm for selecting optimal combinations of parameters of the LS-SVM classifier. In order to report the performance for different reference signals, we used two reference signals, Fp1 and C3, and then decided which EEG channel signal is better for reference. We also added a further three features into the existing features to compare performance. This method offers great potential for the development of MI-based BCI analyses which assist clinical diagnoses and rehabilitation tasks.

3: Investigating which algorithm and which EEG data (motor area data or all-channels data) are better for the MI signal classification

In this dissertation, we investigated which algorithm is the best for the MI tasks classification and which EEG data is better for the MI tasks signal classification. We evaluated the performance of the three algorithms; CC-LS-SVM, CC-LR and CC-KLR on two EEG datasets. From the experimental evaluations, we concluded that the CC-LS-SVM is the best algorithm for the classification of the MI tasks EEG signals and the all-channels EEG can provide better information than the motor area EEG for the MI tasks classification.

This dissertation attempts to develop new approaches to identify the EEG signals for different brain activities. It is our hope that these approaches contribute to successful classification techniques, which can be used in both clinical purposes and the brain research.

1.4 Structure of the Dissertation

This dissertation consists of nine chapters and each chapter provides important information on our research. The rest of this dissertation is structured as follows:

Chapter 2 provides an overview of EEG signal classification techniques and its background knowledge. Firstly, this chapter briefly introduces the background knowledge and surrounding information of this research on the human brain, the fundamentals of EEG, epilepsy, epileptic seizures and how they affect EEG sits, and the concept of BCI systems. After that, this chapter focuses on the concept of signal classification including its structure and methods.

Chapter 3 introduces a novel method based on a simple random sampling technique and least square support vector machine (SRS-LS-SVM) for the classification of epileptic EEG signals. This classification technique can help to differentiate EEG segments and to decide whether a person is healthy.

Chapter 4 presents a clustering technique based least square support vector machine (CT-LS-SVM) for classifying epileptic EEG signals while reducing the experimental time. In this proposed approach, the clustering technique (CT) is used for feature extraction, and a least square support vector machine (LS-SVM) for the classification purpose. This chapter provides a comparative study between the CT based approach and the SRS based approach with respect to the classification accuracy and execution time.

Chapter 5 focuses on the identification of MI signals for the development of BCI technologies combining cross-correlation (CC) and logistic regression (LR) techniques named by CC-LR. This chapter investigates the performance of the CC technique for a randomly selected reference signal Fp1 and also investigates how more efficient the LR classifier is in classifying the cross-correlated features.

Chapter 6 provides a modified version of the CC-LR algorithm with diverse feature sets for classifying MI tasks EEG signals. This chapter investigates which features are the best for the representation of the distribution of the MI signals. Finally, this

chapter reaches a conclusion on which features are suitable for this algorithm to characterize the distribution of EEG signals. This chapter also provides an insight into how to select a reference channel for the CC technique using EEG signals for getting better information about MI.

Chapter 7 introduces another novel algorithm to improve the separability of the MI tasks combining the CC and LS-SVM methods called by CC-LS-SVM. In this algorithm, the hyper parameters of the LS-SVM method are obtained by a two-step grid search process, which can provide more reliable and consistent results. This chapter shows the performance of the LS-SVM classifier against the other two classifiers, logistic regression (LR) and kernel logistic regression (KLR). The effects of two different reference channels, the electrode position Fp1 and C3, are also investigated in this chapter.

Chapter 8 reports on a comparative study between the motor area EEGs and the all-channels EEGs to identify MI tasks EEG signals in BCI applications. This chapter investigates two issues:

(1) Which algorithm is the best for the MI signal classification.

(2) Which EEG data is better for the MI signal classification. Is it the motor area data or the all-channels data.

This chapter considers C3 channel as a reference channel to calculate each cross correlation sequence in the CC technique.

Chapter 9 provides a summary and findings of the work presented on the issues addressed by this research. This chapter also provides the information of the future work.

CHAPTER 2
OVERVIEW OF EEG SIGNAL CLASSIFICATION AND ITS BACKGROUND KNOWLEDGE

The goal of this dissertation is to develop methods that are capable of classifying different categories of electroencephalography (EEG) signals to help in the evaluation and treatment of neurological diseases. In order to have a broad understanding of *classification*, this chapter mainly provides an overview of classification including its concept, structure and commonly used methods of EEG signal classification.

Before providing the overview of the classification, this chapter introduces the general concepts and background knowledge about EEG, epilepsy and BCIs. A brief description of these terminologies is discussed in Section 2.1. Section 2.2 presents a core idea of the classification of EEG signals. This section also reports different methods used in the literature for the classification of EEG signals. Finally, this chapter gives a brief summary of the EEG and the EEG signal classification for this dissertation.

2.1 Background knowledge related to EEG signals

This section presents the background knowledge related to EEG signals to introduce some terminologies and related information of this research. As EEG is generated from the brain, Section 2.2.1 introduces the anatomical and neurophysiological structure of the human brain. This section particularly focuses on the functions of the main parts of the brain and the creation of the electrochemical currents that are picked up by scalp electrodes and that form the EEG. A review on EEG and its nature is discussed in Section 2.1.2. Section 2.1.3 briefly describes epilepsy and epileptic seizures, and how they affect EEG. Section 2.1.4 provides the necessary information related to the underlying principles of BCIs, including their structure and applications in real life.

2.1.1 Human Brain

The supreme commander of the human body is the brain. it's the central part of the nervous system which governs the functions of various organs in the body. Firstly, in this section we explain the anatomical structures of the brain and their functions. Then we focus on how, why and where the brain generates electrical activities that can be recorded on the scalp. It provides a clear idea to understand the creation of local current flows within the brain that can be captured by EEGs.

2.1.1.1 Brain structures and their functions

Anatomically the brain can be divided into three major parts; cerebrum, cerebellum and brainstem (Gray, 2002) as illustrated in Figure 2.1. Below is a brief explanation of the three aforementioned parts of the brain.

(1) Cerebrum: The cerebrum is the largest and most important part of the human brain and is generally associated with brain functions related to *thoughts*, *movements*, *emotions* and *motor functions*. The outermost layer of the cerebrum is made up of neural tissues known as the cerebral cortex. The cerebrum consists of two hemispheres, such as right and left hemispheres. Each hemisphere can be divided into four lobes: frontal, parietal, occipital and temporal (Purves et al., 2004). These lobes are responsible for a variety of bodily functions.

- **Frontal Lobe** is involved with personality, emotions, problem solving, motor development, reasoning, planning, parts of speech and movement.
- **Parietal Lobe** is responsible for sensation (e.g. pain, touch)**,** sensory comprehension, recognition, perception of stimuli, orientation and movement.
- **Occipital Lobe** is responsible for visual processing.
- **Temporal Lobe** is involved in dealing with the recognition of auditory stimuli, speech, perception and memory.

(2) Cerebellum: The cerebellum is located at the lower back of the head and is also divided into two hemispheres. It is the second largest structure of the brain and contains more than half of the brain neurons. The cerebellum is one of the sensory

areas of the brain that is responsible for *motor control*, *sensory perception* and *co-ordination*. The cerebellum is also associated with *voluntary muscle movements*, *fine motor skills*, *posture* and *balance regulation*.

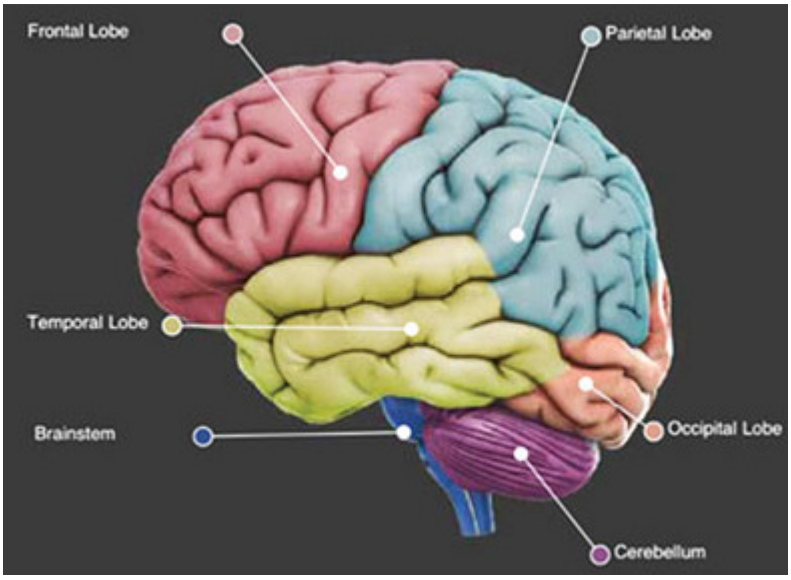


Figure 2.1: Anatomical areas of the brain (Gray, 2002).

(3) Brainstem: The brainstem is located at the bottom of the brain and connects the cerebrum to the spinal cord. The brainstem is like a hard drive of a computer and it is the main control panel of the body. It controls vital functions of the body, including *breathing*, *consciousness*, *movements of the eyes and mouth*, and the *relaying of sensory messages* (*pain*, *heat*, *noise* etc), *heartbeat*, *blood pressure* and *hunger*.

2.1.1.2 Human brains' neurophysiology

The human brain consists of about 100 billion nerve cells called neurons and the electrical charge of the brain is maintained by these neurons. Neurons share the same characteristics and have the same parts as other cells, but the electrochemical aspect lets them transmit electrical signals and pass messages to each other over long distances. Neurons have three basic parts: *cell body (soma)*, *axon* and *dendrites* (Carlson, 2002a; Purves et al., 2004) as shown in Figure 2.2.

The cell nucleus is the heart of the cell, providing it with instructions on what to do. The axon is a long, slender portion of the neuron that connects the nucleus of its own neuron to the dendrite of another. The dendrite is a short section of the neuron with many receptor sites for neurotransmitters that may be sent by a paired axon. Dendrites can be located on one or both ends of the cell. Through the axon-dendrite link, neurons can communicate between each other. This communication is made possible through the action potential.

The action potential is an event where the ion pumps along the outside of an axon rapidly changing the ionic makeup of the axon, allowing an electrical signal to travel quickly through the axon to the next dendrite (Atwood and MacKay, 1989). As a result of this rapid change in ionic charge, a voltage is generated both on the inside and the outside of the cell membrane of the neuron (Carlson, 2002b; Sanei and Chambers, 2007; Purves et al., 2004). These neurons emit a chemical which is sent to another neuron through the synapse, i.e. the gap between the neurons, in order to trigger an activity. The chemicals sent from one neuron to another to trigger it are known as neurotransmitters. The inter-neuron communication system is depicted in Figure 2.2 and Figure 2.3 presents the current flow that contributes to the surface EEG during a net excitatory input.

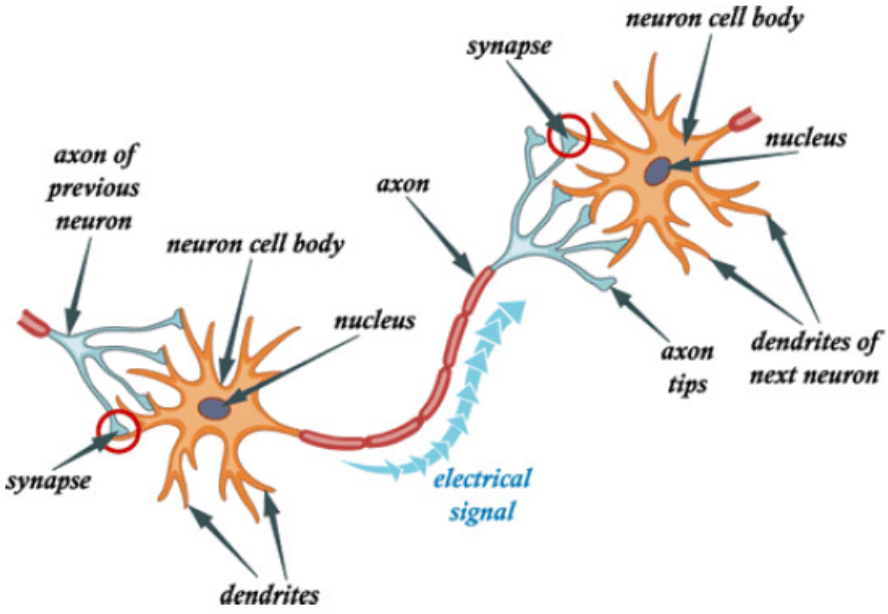


Figure 2.2: A simple structure of neuron (Sanei and Chambers, 2007).

When neurons are activated by means of an electrochemical concentration gradient, local current flows are produced. The electrical activity of neurons can be divided into two subsets; action potentials (AP) and postsynaptic potentials (PSP). If the PSP reaches the threshold conduction level for the postsynaptic neuron, the neuron fires and an AP is initiated (Atwood and MacKay, 1989).

The electrical potentials recordable on the scalp surface are generated by low frequency summed inhibitory and excitatory PSP's from pyramidal neuron cells that create electrical dipoles between the soma and apical dendrites (see Figure 2.3). These PSP's summate in the cortex and extend to the scalp surface where they are recorded as the EEG. Nerve cell AP's have a much smaller potential field distribution and are much shorter in duration than PSPs. AP's therefore do not contribute significantly to either scalp or clinical intracranial EEG recordings. Only large populations of active neurons can generate electrical activity recordable on the scalp (Carlson, 2002b; Sanei and Chambers, 2007; Purves et al., 2004). The voltage, when generated by a single cell, is typically too small to accurately measure with present-day technology.

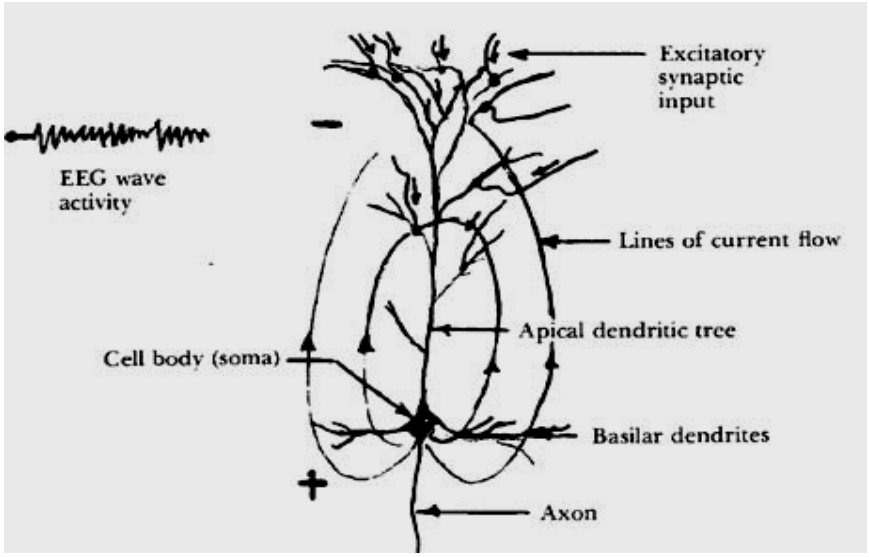


Figure 2.3: A single cortical pyramidal cell showing the current flow that contributes to the surface EEG during a net excitatory input (Carlson, 2002; Atwood and MacKay,1989).

In the EEG measurement, the cerebral cortex is the most relevant structure as it is responsible for higher order cognitive tasks, such as problem solving, language comprehension, movement and processing of complex visual information. Due to its surface position, the electrical activity of the cerebral cortex has the greatest

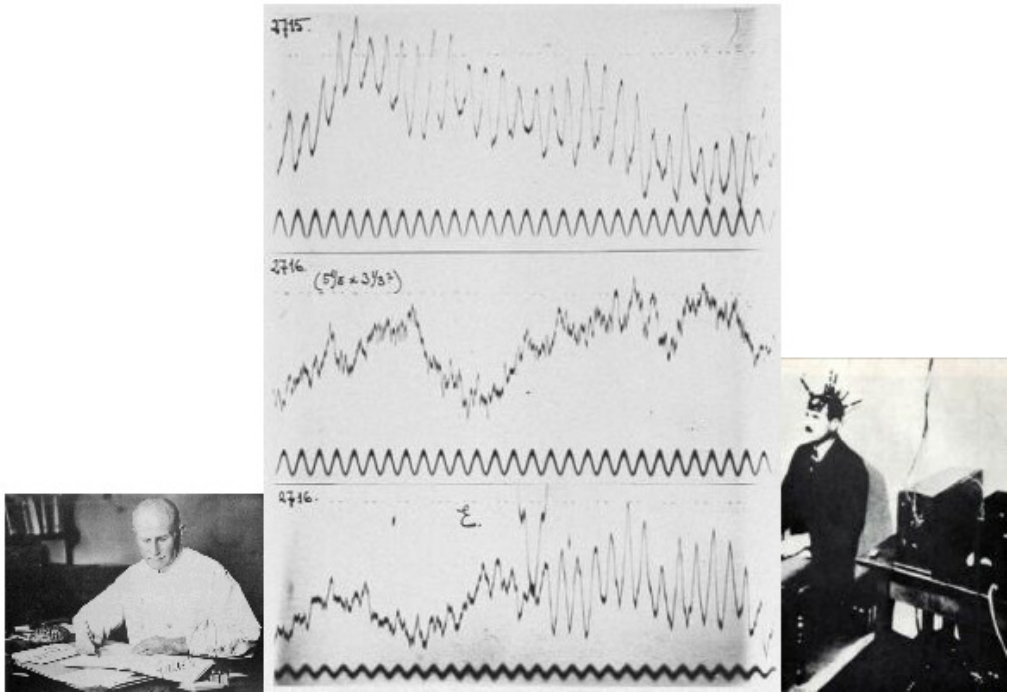
influence on EEG recordings. In the following section, we provide background information of EEG in detail.

2.1.2 Electroencephalography (EEG)

Electroencephalography (EEG) is a measurement of potentials that reflect the electrical activity of the human brain. It is a readily available test that provides evidence of how the brain functions over time. The EEG is widely used by physicians and scientists to study brain functions and to diagnose neurological disorders. The study of the brain electrical activity, through the EEG records, is one of the most important tools for the diagnoses of neurological diseases, such as epilepsy, brain tumour, head injury, sleep disorder, dementia and monitoring depth of anaesthesia during surgery (Hazarika et al., 1997, Adeli et al., 2003) etc. It may also be recommended for the treatment of abnormalities, behavioural disturbances (e.g. Autism), attention disorders, learning problems, language delay etc.

The EEG machine was first introduced to the world by Hans Berger in 1929 (Collura, 1993). Berger, who was a neuropsychiatrist from the University of Jena in Germany, used the German term ‘elektrenkephalogramm’ to describe the graphical representation of the electric currents generated in the brain. He suggested that brain currents changed depending upon the functional status of the brain such as sleep, anesthesia, and epilepsy. This was revolutionary idea that helped create a new branch of medical science called neurophysiology. Figure 2.4 displays the first recording of EEG signals made by Hans Berger. Berger noticed that rhythmic changes (brain waves) varied with the individual’s state of consciousness.

During the EEG test, a number of small discs called electrodes are placed to different locations on the surface of the scalp with temporary glues. Then each electrode is connected to an amplifier (one amplifier per pair of electrodes) and an EEG recording machine. Finally, the electrical signals from the brain are converted into wavy lines on a computer screen to record the results. EEG recordings, depending on their use, can have from 1 to 256 electrodes recorded in parallel, which is called multichannel EEG recordings. One pair of electrodes usually makes up a channel. Each channel produces a signal during an EEG recording.



Hans Berger (1873-1941) Recordings of EEG made by Berger Patient of Berger

Figure 2.4: First recording of EEG signals made by Hans Berger (Berger, 1929).

There are two types of EEG, depending on where the signal is taken in the head: *scalp* or *intracranial*. For the *scalp EEG*, small electrodes are placed on the scalp with good mechanical and electrical contact. Special electrodes implanted in the brain during the surgery result in *intracranial EEG*. On the other hand, the EEG measured directly from the cortical surface using subdural electrodes is called the *electrocorticogram (ECoG)*. The amplitude of an EEG signal typically ranges from about 1 to 100 μV in a normal adult, and it is approximately 10 to 20 mV when measured with subdural electrodes such as needle electrodes. Since the architecture of the brain is non-uniform and the cortex is functionally organised, the EEG can vary depending on the location of the recording electrodes.

The question of how to place the electrodes is important, because different lobes of cerebral cortex are responsible for processing different types of activities. The standard method for the scalp electrode localization is the international 10-20 electrode system (Jasper, 1958a). The “10” and “20” represent actual distances between neighbouring electrodes are either 10% or 20% of the total font-back or right-left distance of the skull. The positions are determined by the following two points; *nasion*, which is the point between the forehead and the nose, level with the

eyes and *inion* which is the bony prominence at the base skull on the midline at the back of the head. Figure 2.5 presents the electrode position on the brain according to the international 10-20 system. Each location uses a letter to identify the lobe and a number to identify the hemisphere location. The letters F, T, C, P and O stand for Frontal, Temporal, Central, Parietal and Occipital, respectively. A "z" refers to an electrode placed on the midline. Even numbers refer to electrode positions on the right hemisphere, whereas odd numbers refer to those on the left hemisphere. Since an EEG voltage signal represents a difference between the voltages at two electrodes, the display of the EEG for the reading EEG machine may be set up in several ways. The placement of the electrodes is referred to as a montage. The EEG can be monitored with the following montages.

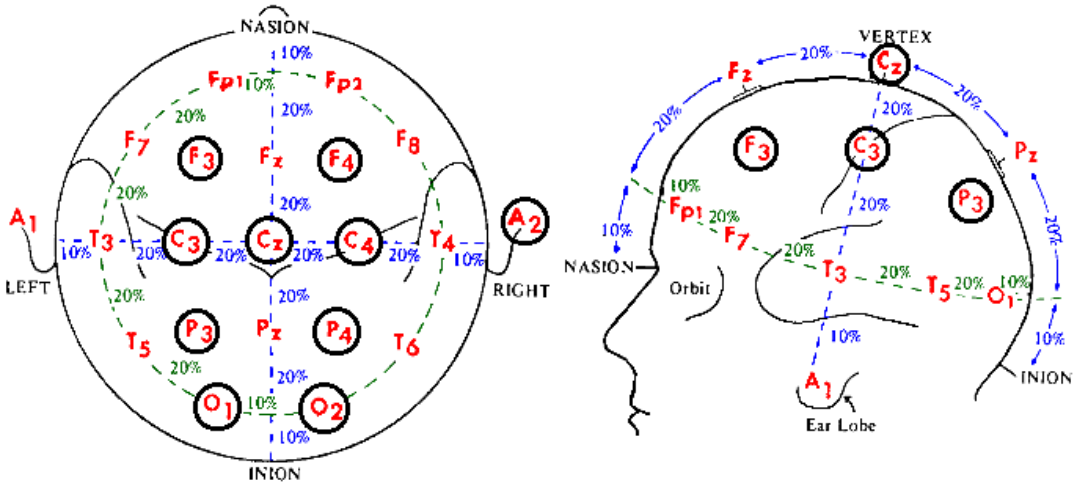


Figure 2.5: The international 10-20 electrode placement system (Jasper, 1958b).

Bipolar montage: One pair of electrodes usually makes up a channel. Each channel (waveform) represents the difference between two adjacent electrodes (Niedermeyer and Lopes Da Silva, 2005; Fisch, 1999). The entire montage consists of a series of these channels. For example, the channel "Fp1-F3" represents the difference in the voltage between the Fp1 electrode and the F3 electrode. The next channel in the montage, "F3-C3," represents the voltage difference between F3 and C3, and so on, through the entire array of electrodes.

Referential montage: Each channel represents the difference between a certain electrode and a designated reference electrode (Niedermeyer and Lopes Da Silva, 2005; Fisch, 1999). There is no standard position for this reference; it is, however, at

a different position than the "recording" electrodes. Midline positions are often used because they do not amplify the signal in one hemisphere versus. the other. Another popular reference is "linked ears", which is a physical or mathematical average of electrodes attached to both earlobes and mastoids.

Average reference montage: The outputs of all of the amplifiers are summed and averaged, and this averaged signal is used as the common reference for each channel (Fisch, 1999).

Laplacian montage: Each channel represents the difference between an electrode and a weighted average of the surrounding electrodes (Fisch, 1999). With digital EEG, all signals are typically digitized and stored in a particular (usually referential) montage; since any montage can be constructed mathematically from any others, the EEGs can be viewed by an EEG machine in any display montage that is desired.

The EEG patterns are very important for understanding brain activities by identifying morphological features or examining frequency bands associated with different mental activities or conscious states. The frequency bands can be divided into five categories. In the next section, we discuss the most common patterns of EEG signals in situations where individuals are in a state of alertness, sleeping, suffering from a brain disorder and experienced extreme emotional.

2.1.2.1 Nature or rhythms of the EEG signals

Frequency is one of the most important for assessing abnormalities in clinical EEGs and for understanding functional behaviours in cognitive research. With billions of oscillating communities of neurons as its sources, the human EEG potentials are manifested as aperiodic unpredictable oscillations with intermittent bursts of oscillations which are typically categorized in specific bands such as 0.5-4 Hz (delta, δ), 4-8 Hz (theta, θ), 8-13 Hz (alpha, α), 13-30 Hz (beta, β) and >30Hz (gamma, γ) (Niedermeyer and Lopes Da Silva, 2005; Fisch, 1999). Figure 2.6 illustrates examples of these EEG rhythms.

Delta wave lies between the range of 0.5 to 4 Hz and the shape is observed as the highest in amplitude and the slowest in waves. It is primarily associated with deep sleep, serious brain disorder and in the waking state.

Theta wave lies between 4 and 8 Hz with an amplitude usually greater than 20 μV. Theta arises from emotional stress, especially frustration or disappointment and unconscious material, creative inspiration and deep meditation.

Alpha contains the frequency range from 8 to 13 Hz, with 30-50m μV amplitude, which appears mainly in the posterior regions of the head (occipital lobe) when the subject has eyes closed or is in a relaxation state. It is usually associated with intense mental activity, stress and tension. Alpha activity recorded from sensorimotor areas is also called mu activity.

Beta is in the frequency range of 13 Hz-30 Hz. It is seen in a low amplitude and varying frequencies symmetrically on both sides in the frontal area. When the brain is aroused and actively engaged in mental activities, it generates beta waves. Beta waves are characteristics of a strongly engaged mind. Beta is the brain wave usually associated with active things, active attentions, and focusing on the outside world or solving concrete problems.

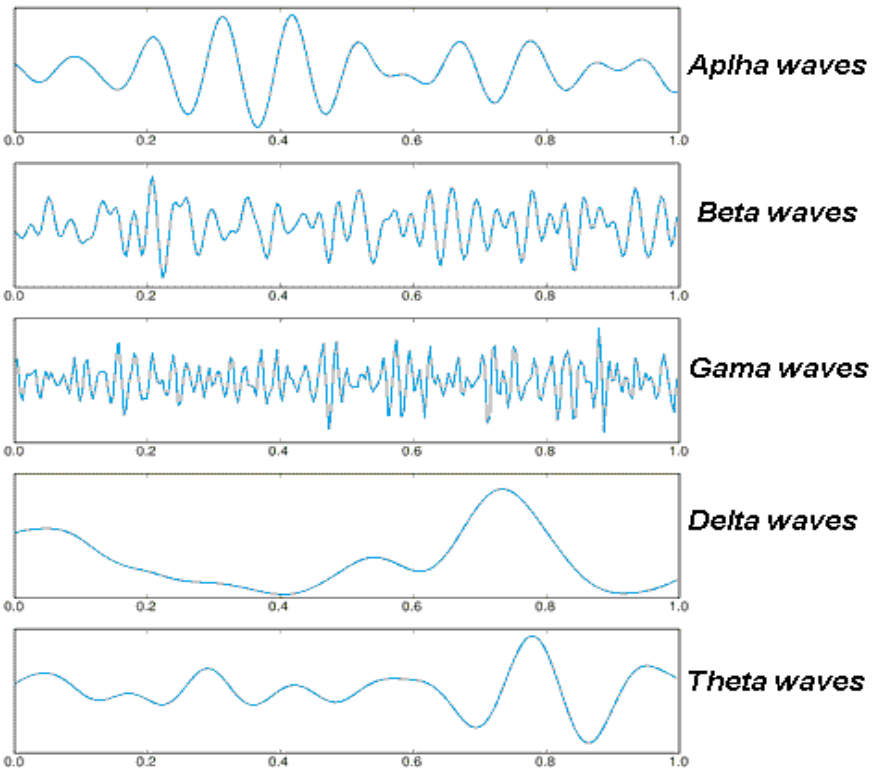


Figure 2.6: Example of different types of normal EEG rhythms (Lotte, 2009).

Gamma waves have the frequency from 30 Hz and up. This rhythm is sometimes defined as having a maximal frequency around 80 Hz or 100 Hz. It is associated with various cognitive and motor functions.

Electrical signals in the EEG that are originated from non-cerebral origin are called artifacts. EEG data is almost always contaminated by such artifacts. The amplitude of artifacts can be largely relative to the size of amplitude of the cortical signals of interest. This is one of the reasons why it takes considerable experience to correctly interpret EEGs clinically.

As one of the main diagnostic applications of EEG is in the case of epilepsy, in this dissertation, two algorithms are developed to classify epileptic EEG signals. To provide an overview about epileptic EEG signals, the following subsection discusses what epilepsy and epileptic seizures are, and what they affect EEG signals.

2.1.3 Epilepsy, epileptic seizures and their effects on EEG signals

Epilepsy is the most prevalent neurological disorder in humans. About 50 million people worldwide have epilepsy (Lima et al., 2009), and nearly two out of every three new cases are discovered in developing countries. More than two million people in the United States have experienced an unprovoked seizure or been diagnosed with epilepsy. Epilepsy is characterized by recurrent seizures (Blume et al., 2001). Seizures are defined as sudden changes in the electrical functioning of the brain, resulting in altered behaviours, such as losing consciousness, jerky movements, temporarily loss of breath and memory loss. These usually happen in the cortex, or outside rim of the brain.

Epilepsy may develop because of an abnormality in brain wiring, an imbalance of nerve signalling chemicals called neurotransmitters, or some combination of these factors. Neurons normally generate electrochemical impulses that act on other neurons, glands, and muscles to produce human thoughts, feelings and actions. In epilepsy, the normal pattern of neuronal activity becomes disturbed, causing strange sensations, emotions, and behaviours, or sometimes convulsions, muscle spasms and loss of consciousness (Quyen et al., 2000). There may be a kind of brief electrical "storm" arising from neurones that are inherently unstable because

of a genetic defect (as in the various types of inherited epilepsy), or from neurones made unstable by metabolic abnormalities such as low blood glucose, or alcohol. Alternatively, the abnormal discharge may come from a localized area of the brain (this is the case in patients with epilepsy caused by head injury or brain tumour). During a seizure, neurons may fire as many as 500 times a second, much faster than normal (1-100μV). In some people, this happens only occasionally; for others, it may happen up to hundreds of times a day.

EEG recordings contain valuable information for understanding epilepsy. Epileptic activity can create clear abnormalities on a standard EEG. Epilepsy leaves its signature in the EEG signals. The detection of seizures occurring in the EEGs is an important component in the diagnosis and treatment of epilepsy. Two categories of abnormal activity can be observed in an EEG signal: *ictal* (during an epileptic seizure) and *inter-ictal* (between seizures). Often, the onset of a clinical seizure is characterized by a sudden change of frequency in the EEG measurement. It is normally within the alpha wave frequency band with slow reduction in frequency but increase in amplitude during the seizure period. It may or may not be spiky in shape. For assisting the diagnosis and treatment of epilepsy or neurological disease, this dissertation aims to develop methods that can identify the epileptic EEG signals during seizure activity and also during seizure-free time.

Recently, EEG has had many applications in BCI technologies in which a person could communicate with other or control devices directly by means of a brain activity without using the normal channels of peripheral nerves and muscles. In the next section, we describe the concept, structure and applications of BCI systems.

2.1.4 Concept of Brain Computer Interfaces (BCIs)

A brain computer interface (BCI) can be defined as a "communication and control channel that does not depend on the brain's normal output channels of peripheral nerves and muscles" (Wolpaw et al., 2002). The messages and commands sent through a BCI are encoded into the user's brain activity. BCI technologies provide a direct interface between a brain and a computer (Vaughan et al., 2003). The ultimate object of a BCI is to provide humans an alternative communication channel, allowing direct transmission of messages from the brain by analysing the brain's mental activities.

A BCI is a technology that finds a new communicative way of using only the brain to command machines. An electrode cap is placed on the head of a user for measuring EEG signals. To command machines a user imagines a specific task such as movement of limbs, composing of words. These tasks affect the patterns of EEG signals. Computers detect and classify these patterns into different tasks in order to control a computer application (such as cursor movement), or control a machine (e.g. wheelchair).

Unlike all other interfaces, BCIs do not require an actual movement, and hence BCIs may be the only means of communications possible for people who have severe motor disabilities. BCIs can also provide the communication and control for other user groups and goals, such as patients with less severe motor disabilities who wish to control an orthosis or wheelchair (Graiman et al., 2008), and healthy users in situations where conventional means of communication are difficult, impractical or inadequate (Allison et al., 2007). BCIs could also help reduce symptoms resulting from stroke, autism, emotional and attention disorder (Kouijzer et al., 2009). There are two types of BCIs: *invasive*, which are based on signals recorded from electrodes implanted over the brain cortex (requiring surgery), and *non-invasive*, based on signals recorded from electrodes placed on the scalp (outside the head) (Wolpaw et al. 2002). In recent research, the non-invasive EEG is the most preferable technique.

2.1.4.1 Architecture of BCI systems

Typically a BCI system requires the following of a closed-up process which generally consists of six steps: brain activity measurement, pre-processing, feature extraction, classification, translation into a command and feedback (Mason and Birch, 2003) shown in Figure 2.7.

1. Brain activity measurement: Measuring brain activity effectively is a critical step for BCI communications. Human intentions modulate the electrical signals which are measured using various types of electrodes and then these signals are digitized. In this thesis, we use EEGs as the measurement of brain activities.

2. Pre-processing: Pre-processing aims to simply subsequent processing operations, improving signal quality without losing information. In this step, the recorded signals are processed to clean and denoise data in order to enhance the relevant information embedded in the signals (Bashashati et al. 2007).

3. Feature extraction: The brain patterns used in BCIs are characterized by certain features. Feature extraction aims at describing the signals by a few relevant values called “features” (Bashashati et al. 2007).

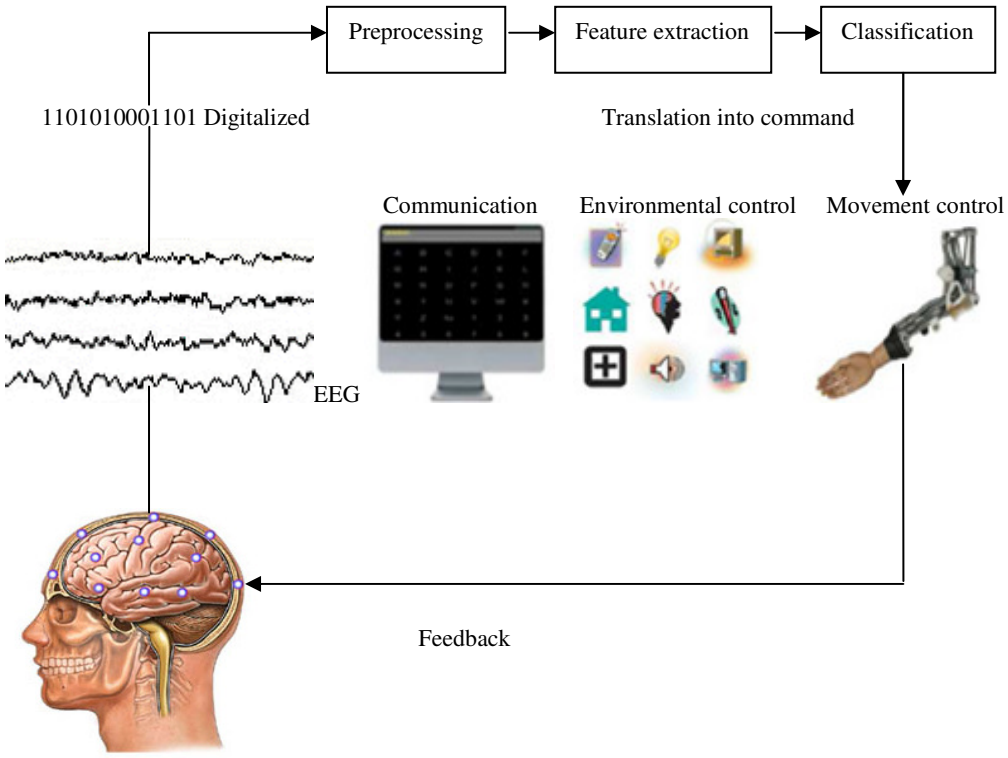


Figure 2.7: A general architecture of a BCI system.

4. Classification: The classification step assigns a class to a set of features extracted from the signals. This class corresponds to the type of mental states identified. This step can also be denoted as “feature translation”.

5. Translation into a command/application: Once the mental state is identified, a command is associated to this mental state in order to control a given application such as a computer or a robot.

6. Feedback: Finally, this step provides the user with feedback about the identified mental state. This aims to help the user control his/her brain activities. The overall objective is to increase the users performances.

A BCI can only detect and classify specific patterns of an activity in continuous brain signals that are associated with specific tasks or events. What a BCI user has to do to produce these patterns is determined by the mental strategy a BCI system employs. The mental strategy is the foundation of any brain computer communication. The

mental strategy determines what the user has to do in order to produce brain patterns that the BCI can interpret. The most common mental strategies are motor imagery (MI) and selective (focused) attention. Motor imagery (MI) is the imagination of a movement without actually performing the movement. On the other hand, BCIs based on selective attention require external stimuli provided by a BCI system. The stimuli can be auditory or somatosensory. In this research, we work on the MI for the BCI systems.

2.1.4.2 Applications of BCI systems

The range of possible BCI applications is very broad. BCIs have been validated with many applications, including spelling devices/virtual keyboard (e.g. patients in a locked-in state), simple computer games, environmental control, neuroprosthesis (e.g. tetraplegic patients), biofeedback therapy (e.g. reduction of epileptic seizures, enhanced stroke rehabilitation, treatment of attention deficit disorders), navigation in virtual reality, and genetic cursor control applications (Blankertz et al. 2007; Pfurtscheller et al 2006b; Sellers and Donchin, 2006). Most of these applications run on conventional computers that host a BCI system and its application as well. An increasing number of systems allow control of more sophisticated devices, including orthoses, prostheses, robot arms and mobile robots (Graimann et al., 2009; Vellistte et al., 2008).

Communication or control based BCI technologies require typical patterns of brain activity which can be consciously generated or controlled by a subject and ultimately clearly distinguishable by a computer system. The system requires a BCI user to concentrate on a mental task in order to produce a characteristic brain pattern that identifies with the desired control, e.g. the imagination of a hand movement. The performances of different mental tasks generate different EEG responses and hence can be translated into a control codebook for the user. In this thesis, we develop three algorithms to identify different mental tasks (e.g. MI task) for the development of communication based BCI systems.

In this dissertation, we are interested in classifying EEG signals e.g. epileptic EEG signals and MI tasks based EEG signals of BCI systems. In the following section, we provide a detailed discussion of the EEG signal classification, including

classification concepts, structure, types and methods used for epileptic EEG data and MI tasks EEG data in BCIs.

2.2 Overview of EEG signal Classification

The classification of EEG signals plays an important role in biomedical research. Classifying EEG signals is very important in the diagnosis of brain diseases and for contributing to a better understanding of cognitive processes. An efficient classification technique assists to distinguish EEG segments, and in the decision making on a person's health. As EEG recordings contain a large amount of data, one key problem is how to represent the recorded EEG signals for further analysis, such as classification. It is, firstly, important to extract useful features from raw EEG signals, and then use the extracted features for classification. The following sections provide detailed outlines for the signal classification in this research area.

2.2.1 Concept of classification

The task of *classification* occurs throughout daily life, and essentially means decisions being made based on currently available information. Examples of classification tasks include the *mechanical procedures* used for sorting letters on the basis of machine read postcodes, *assigning individuals* to credit status on the basis of financial and other personal information, and the *preliminary diagnosis* of a patient's disease in order to select immediate treatment while awaiting definitive test results (Brunelli, 2009).

In machine learning and pattern recognition, *classification* refers to an algorithm procedure for assigning a given piece of input data into one of a given number of categories (Duda et al., 2001; Brunelli, 2009). An example would be assigning an e-mail to a "Spam" or "non-spam" section or giving a diagnosis to a patient based on observed characteristics (gender, blood pressure or presence or absence of certain symptoms, etc.). The piece of input data is formally known as an *instance* and the categories are termed *classes*. The *instance* is formally described by a vector of *features,* which together constitute a description of all known characteristics of the *instance*.

The goal of the classification is to assign class labels to the features extracted from the observations of a set of data in a specific problem. An algorithm that implements classification, especially in a concrete implementation, is known as a *classifier*. The term *classifier* sometimes also refers to the mathematical function, implemented by a classification algorithm that maps input data to a category. *Classifie*rs are able to learn how to identify the class of a feature vector thanks to training sets. These sets are composed of feature vectors labelled with their classes of belonging.

This research works on the EEG signal classification. Measuring brain activity through EEG leads to the acquisition of a large amount of data. In order to obtain the best possible performances it is necessary to work with a smaller number of values which describe some relevant properties of the signals. These values are known as "features". Features are generally aggregated into a vector known as a "feature vector" (Lotte, 2009). Thus, feature extraction can be defined as an operation which transforms one or several signals into a feature vector. The feature vector, which is comprised of the set of all features used to describe a pattern, is a reduced dimensional representation of that pattern. Signal classification means to analyse different characteristic features of a signal, and based on those characteristic features, decide to which grouping or class the signal belongs. The resulting classification decision can be mapped back into the physical world to reveal information about the physical process that created the signal.

2.2.2 Types of classification

There are two main divisions of classification: *supervised classification* and *unsupervised classification*. In *supervised classification*, observations of a set of data are associated with class labels. In *unsupervised classification,* observations are not labelled or assigned to a known class (Jain et al., 2000). Descriptions of aforementioned are provided as below.

2.2.2.1 Supervised classification

Supervised classification is preferred in the majority of biomedical research. Most of the classification algorithms deal with a group of data that has some information

about the dataset. In other words, the class label information is given within the dataset for training the classifier. This type of classification belongs to *supervised* learning, in which a supervisor instructs the classifier during the construction of the classification model. *Supervised* procedure assumes that a set of *training data* (the *training set*) has been provided, consisting of a set of instances that have been properly labelled by hand with the correct output (Duda et al., 2001; Brunelli, 2009).

In the *supervised learning* approach, there are pairs of examples in the given training dataset which can be mathematically expressed as $D=\{(x_1, y_1), (x_2, y_2),\ldots\ldots, (x_N, y_N)\}$. Here, $x_1, x_2,\ldots\ldots,x_N$ are the observations and $y_1, y_2,\ldots\ldots,y_N$ are the class labels of the observations. For example, if the problem is filtering spam, then x_i is some representation of an email and y_i is either "spam" or "non-spam". The observations can be any vector, whose elements are selected from a set of features. For practical considerations, we usually have real valued observations and it is easy to assume $x \varepsilon X$. Also, one can choose any type of representation for the class labels. For simplicity, they are usually represented as real numbers that is $y \varepsilon Y$. Therefore, in *supervised classification*, the aim is to find the transformation between the *feature* space X and the *class label* space Y, i.e. $f : X \rightarrow Y$. If the *class space* has a finite number of elements, i.e. $y \varepsilon \{1,2,\ldots\ldots, L\}$ then the problem is considered as a *classification task*. For the case of a binary classification problem, the classes are divided into two categories, such as the target and non-target classes. For clarity and conformity with the literature, these classes are represented as $Y = \{-1,+1\}$ where the negativity represents the non-target case.

Algorithms for the classification depend on the type of label output, on whether learning is *supervised* or *unsupervised*, and on whether the algorithm is *statistical* or *non statistical* in nature. Statistical algorithms can be further categorized as generative or discriminative. The algorithms in *supervised classification* procedure predicting categorical labels are Linear discriminant analysis (LDA), Support vector machine (SVM), Decision trees, Naive Bayes classifer, Logistic regression, K-nearest-neighbor (*k*NN) algorithms, Kernel estimation, Neural networks (NN), Linear regression, Gaussian process regression, Kalman filters etc.

In a typical *supervised classification* procedure, the dataset is divided into two, *training set* and *testing set*. Using the *training set*, a classifier is constructed. Then the performance of the classifier is evaluated using the *testing* set. This

evaluation is sometimes repeated for different parameters of the classifier constructed. By that way the parameters of the classifier is optimized. After that optimization, the classifier is ready to assign class labels to the features with unknown class labels. The goal of the learning procedure is to maximize this test accuracy on a "typical" *testing set*.

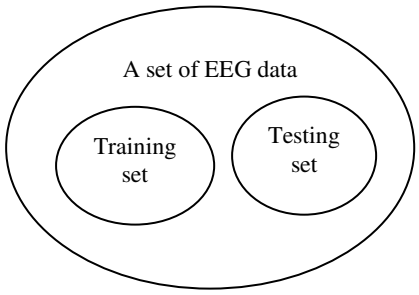


Figure 2.8: Mutually exclusive training and testing set.

Classification normally refers to a *supervised* procedure. In this research, we have used *supervised* procedure in the classification of EEG signals. During the experiment, we divide each EEG dataset into two mutually exclusive groups *training set* and *testing set* as shown in Figure 2.8. The reasoning for having the sets separated can be linked to memorization and generalization. The training set is used to train the classifier, while the testing set is used to evaluate the performance of the classifier.

2.2.2.2 Unsupervised classification

The *unsupervised classification* procedure involves grouping data into classes based on some measure of inherent ability (e.g. the distance between instances, considered as vectors in a multi-dimensional vector space). This procedure assumes training data has not been hand-labelled, and attempts to find inherent patterns in the data that can then be used to determine the correct output value for new data instances (Duda et al., 2001; Brunelli, 2009). In *unsupervised learning*, any information about the class labels of the measurements is not available even for a small set of data.

The common algorithms of *unsupervised classification* are *K*-means clustering, Hierarchical clustering, Principal Component Analysis (PCA), Kernel Principal Component Analysis (Kernel PCA), Hidden Markov Models, Independent Component Analysis (ICA), Categorical mixture model, etc.

A combination of the two classification procedures (*supervised* and *unsupervised*) that has recently been explored is *semi-supervised* learning, which uses a combination of labelled and unlabelled data (typically a small set of labelled data combined with a large amount of unlabelled data.

2.2.3 Structure of classification

According to pattern recognition principles, the *classification* process consists of two stages: *feature extraction* and *classification*. The *feature extraction* stage entails the extraction of the most important features of the signal. The *classification stage* involves the use of the classifier to determine the particular class of a signal based on its extracted features. In the *feature extraction stage*, the characteristics or features upon which the signal is to be classified must be defined. There are number of features that could be extracted from the signal for the purpose of classification. In the *classification stage*, based on the selected signal features, a classifier will determine to which class the signal belongs.

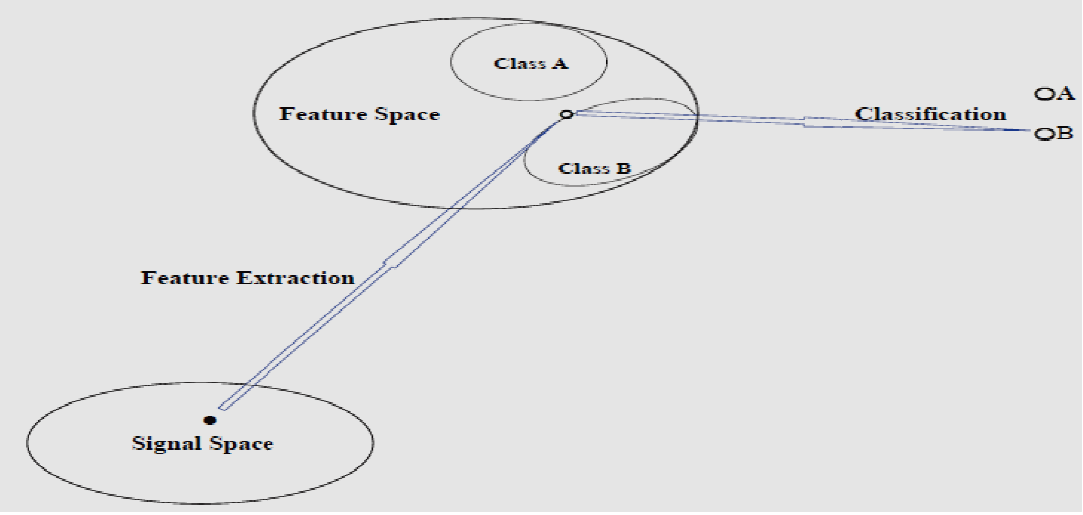


Figure 2.9: Process of signal classification in the biomedical engineering.

The output of the classification system, in which the class membership of the input signal is determined, can then be used to infer what event in the real world process occurred to produce the input signal. The concept of signal classification can be depicted in Figure 2.9.

Figure 2.9 presents a structure of how signals with different categories are classified, extracting features from original data in pattern recognition area. From this figure, it is seen that appropriate features are extracted from the signal space and generate a feature space. In the feature space, the features are divided into two classes (class A and class B). Finally, a classifier attempts to identify the extracted features during the classification.

In this dissertation, we use three methods: *random sampling technique, clustering technique* and *cross-correlation technique* to extract the critical features from the EEG signals, while *least square support vector machine (LS-SVM)*, *logistic regression* and *kernel logistic regression* are the tools used to perform the classification.

2.2.4 Common used methods of the EEG signal classification

In previous studies there were various methods used for the different type of EEG signal classification. This research develops methods for *the epileptic EEG signal classification* and for *the MI based EEG signal classification in BCI systems*. Section 2.2.3.1 reports the methods which were used in the epileptic EEG data for the feature extraction and the classification. The methods for the MI tasks based EEG data of BCI systems are described in Section 2.2.3.2.

2.2.4.1 Methods used in the epileptic EEG signal classification

An accurate feature extraction method is very important to extract good features from original signals that would significantly affect the accuracy of classifying EEG signals. In fact, if the features extracted from EEGs are not relevant and do not accurately describe the EEG signals employed, a classification algorithm which will use such features will have trouble identifying the classes of these features. As a result, the correct classification rates will be very low. From the literature, it is seen that a variety of methods have been used for feature extraction in epileptic EEG data. The feature extraction methods can be classified into four groups: *parametric methods*, *non-parametric methods*, *time-frequency methods* and *eigenvector methods*.

The *parametric or model-based methods* assume that the signal satisfies a generating model with known functional form, and then proceeds by estimating the

parameters in the assumed model. Some popular parametric methods are autoregressive (AR) model (Ubeyli, 2009a), moving average (MA) model (Ubeyli, 2009a), and autoregressive-moving average (ARMA) model (Ubeyli, 2009a). The AR model is suitable for representing spectra with narrow peaks. The MA model provides a good approximation for those spectra which are characterized by broad peaks and sharp nulls. Such spectra are encountered less frequently in applications than narrowband spectra, so there is a somewhat limited interest in using the MA model for spectral estimation. Spectra with both sharp peaks and deep nulls can be modelled by the ARMA model. The practical ARMA estimators are computationally simple and often quite reliable, but their statistical accuracy may be poor in some cases (Kay, 1988; Kay & Marple, 1981; Proakis & Manolakis, 1996; Stoica and Moses, 1997).

The *non-parametric methods* rely entirely on the definitions of power spectral density (PSD) to provide spectral estimates. These methods constitute the ‘‘classical means” for the PSD estimation. Two common *non-parametric methods,* periodogram and the correlogram (Ubeyli, 2009a), provide reasonably high resolution for sufficiently long data lengths, but are poor spectral estimators because their variance is high and does not decrease with increasing data length. The high variance of the periodogram and correlogram methods motivates the development of modified methods that have lower variance at a cost of reduced resolution (Ubeyli, 2009a).

Mappings between the *time and the frequency domains* have been widely used in signal analysis and processing. The methods which are usually used in *time-frequency domain* are fast Fourier transform (FFT) (Welch, 1967), short time Fourier transform (STFT) (Ubeyli, 2009a) and wavelet transform (WT) (Adeli and Dadmehr, 2003; Subasi, 2005; Ubeyli, 2009b). Since Fourier methods may not be appropriate for non-stationary signals, or signals with short-lived components, alternative approaches have been sought. Among the early works in this area is Gabor’s development of the short-time Fourier transform (STFT). The wavelet transform (WT) provides a representation of the signal in a lattice of ‘‘building blocks” which have good frequency and time localization. The wavelet representation, in its continuous and discrete versions, as well as in terms of a multi-resolution approximation is presented in (Akay, 1998; Ubeyli & Guler, 2004).

Eigenvector methods are used for estimating frequencies and powers of signals from noise-corrupted measurements. These methods are based on an eigen

decomposition of the correlation matrix of the noise–corrupted signal. Even when the signal-to-noise ratio (SNR) is low, the eigenvector methods produce frequency spectra of high resolution. The eigenvector methods, such as Pisarenko, multiple signal classification (MUSIC), and minimum-norm, are best suited to signals that can be assumed to be composed of several specific sinusoids buried in noise (Proakis & Manolakis, 1996; Stoica & Moses, 1997; Ubeyli & Guler, 2003).

These feature extraction methods have been combined with different types of classifiers in the last a few years. Examples include Adaptive neuro-fuzzy inference system (Guler and Ubeyli, 2005), support vector machine (SVM) (Makinac, 2005; Burges, 1998; Guler et al. 2007; Chandaka et al. 2009; Silver, 2006; Fan, 2006), least square support vector machine (LS-SVM) (Ubeyli, 2010; Siuly et al. 2011c; Siuly et al. 2011d; Siuly et al. 2010; Siuly et al 2009; Hanbay, 2009) and artificial neural network (ANN) (Guler, 2005; Subasi, 2007; Ubeyli, 2008; Jahankhani, et al. 2006; Subasi and Ercelebi, 2005; Guo et al. 2009), multilayer perceptron neural network (MLPNN) (Guler and Ubeyli, 2007), recurrent neural network (RNN) (Guler et al., 2005), relevance vector machine (RVM) (Lima et al, 2009), probabilistic neural network (PNN) (Guler and Ubeyli, 2007), mixture of experts (MEs) (Ubeyli, 2009a), modified mixture of experts (MMEs) (Ubeyli, 2009a), etc.

The performance of a classifier depends greatly on the characteristics of the data to be classified. There is no single classifier that works best on all given problems. Various empirical tests have been performed to compare classifier performance and to find out the characteristics of data that determine classifier performance. The measures of accuracy and confusion matrix are very popular used to evaluate the quality of a classification system. More recently, receiver operating characteristic (ROC) curves have been used to evaluate the trade-off between true- and false-positive rates of classification algorithms. This research mainly uses accuracy to assess the performance of the proposed methods. The confusion matrix and ROC curves are also used to evaluate the performance.

2.2.4.2 Methods used in the MI based EEG signal classification in the BCI systems

In BCI applications, several methods have been studied and employed for the feature extraction from the MI based EEG signals. These include autoregressive (AR)

(Schloglel al., 2002; Pfurtschelleret al., 1998; Burke et al., 2005; Guger et al., 2001), fast Fourier transform (FFT) (Polat et al., 2007), common spatial patterns (CSP) (Blanchard et al., 2003; Lemm, et al. 2005), spatio-spectral patterns (Wu et al., 2008), wavelet coefficients (Qin et al., 2005; Ting et al, 2008). Feature extraction methods based on Self Organizing Maps (SOM) using auto-regressive (AR) spectrum (Yamaguchi, et al., 2008) and inverse model (Qin et al., 2004; Kamousi et al., 2005; Congedo et al., 2006) have been studied to discriminate the EEG signals recorded during the right and left hand motor imagery. In the decomposition of EEG multiple sensor recordings, the feature selection used were PCA and ICA (Sanei, and Chambers, 2007). All movement-related potentials are limited in duration, frequency and spatial information of EEG data (Congedo et al., 2006; Sanei, and Chambers, 2007). The combination of time-frequency (TF) and Linear Discriminant Analysis (LDA) technique can be used (Bian et al., 2010). The feature extraction method based on discrete Wavelet transform has been employed in (Kousarrizi et al., 2009) to control the cursor movement via EEGs. The variance and mean of signals decomposed by a Haar mother wavelet served as the inputs to the classifiers (Kousarrizi et al., 2009). In other studies, db40 wavelet packet decomposition was used to select features of EEG signals to control a four-direction motion of a small ball on the computer screen (Bian et al., 2010).

There are five different categories of classifiers; *linear classifiers*, *neural networks*, *non-linear Bayesian classifiers*, *nearest neighbour classifiers* and *combinations of classifiers* that have been studied in a BCI system design (Lotte et al., 2007; Md Norani et al., 2010). *Linear classifiers* are discriminant algorithms that use linear functions to distinguish classes. Linear Discriminant Analysis (LDA) otherwise known as Fisher's Linear Discriminant Analysis (FLDA) and Support Vector Machine (SVM) are the most popular techniques used to separate the data representing different classes by using hyperplanes (Lotte et al., 2007; Duda et al., 2001). FLDA has the ability to distinguish signals from a related movement activity with a classification accuracy of 81.63% from a single trial (Kaneswaran et al., 2010). In recognizing P300 potentials obtained from spelling a word, FLDA (an accuracy of 95.75%) outperforms Least Squares Analysis (LSA) and Stepwise Linear Discriminant Analysis (SWLDA) (Congedo et al., 2006). However, LDA and SVM classifiers have other limitations. The main limitation of LDA is its linearity, which can cause poor outcomes when it deals with complex nonlinear EEG data (Lotte et

al., 2007; Garcia et al., 2003). SVM are known to have good generalization properties, but have low speed of execution (Lotte et al., 2007).

The most widely used *neural networks (NNs)* for BCI systems, is the multilayer perceptron (MLP), a common approximator that is sensitive to overtraining, especially with such noisy and nonstationary data as EEGs (Lotte et al., 2007; Md Norani et al., 2010). Other types of NNs used in BCIs are the Gaussian classifier (Millan et al., 2004), Learning Vector Quantization (LVQ) neural network (Pfurtscheller et al., 1993), fuzzy ARTMAP Neural Network (Palaniappan et al., 2002), Dynamic Neural Networks such as the Finite Impulse Response Neural Network (FIRNN) (Haselsteiner, and Pfurtscheller, 2000), Time-Delay Neural Network (TDNN) or Gamma dynamic Neural Network (GDNN) (Barreto et al., 1996), RBF Neural Network (Hoya et al., 2003), Bayesian Logistic Regression Neural Network (BLRNN) (Penny et al., 2000), Adaptive Logic Network (ALN) (Kostov et al., 2000) and Probability estimating Guarded Neural Classifier (PeGNC) (Felzer and Freisieben, 2003). Recently, Barbosa et al (Ming et al., 2009) studied the performance of Probabilistic Neural Network Delta band (PNN-DB), MLP Neural Network with Driven Pattern Replication (MLP-DPR), Modular Multi-net system (MMN) and Hierarchical Model (HM) to find the best method to control a mobile robot. They discovered that the HM with statistical implementation could produce the best result with an accuracy of 91%.

There are two types of *Non-linear Bayesian classifiers* used in BCI systems; Bayes quadratic and Hidden Markov Model (HMM) (Md Norani et al., 2010). Both classifiers produce non-linear decision boundaries. The advantages of these classifiers are that they are generative and reject uncertain samples more efficiently than discriminative classifiers (Lotte et al., 2007). Nearest Neighbour classifiers are also used in BCIs, for example, *k* Nearest Neighbour (*k*NN) and Mahalanobis Distance. *k*NN assigns an unseen point the dominant class among its k nearest neighbors within the training set. *k*NN has failed in several BCI experiments due to being very sensitive to the curse-of-dimensionality (Lotte et al., 2007; Felzer and Freisieben, 2003). However, it may perform efficiently with low-dimensional feature vectors (Lotte et al., 2007). Mahalanobis Distance classifier has been used to detect the imagination of hand movement tasks and the accuracy produced by this classifier is 80% (Ming et al., 2009).

Classifiers can be combined to reduce the variance and thus increase the classification accuracy. Boosting, voting and stacking are the classifier combination strategies used in BCI applications (Lotte et al., 2007). Boosting consists of several classifiers in cascade where the errors committed by the previous classifier are focussed by each classifier (Lotte et al., 2007). In voting, several classifiers are used with each of them assigning an input feature vector to a class. Due to its simplicity and efficiency, voting is the most popular one and has been combined with LVQ NN, MLP or SVM (Lotte et al., 2007). Stacking uses several classifiers which are called level-O classifiers to classify the input feature vectors. The output of each of these classifiers serves as the input to a meta-classifier (or level-I classifier) which is responsible for making the final decision. In BCI research, stacking has been used as level-O classifiers in Hidden Markov Models (HMM), and as a meta-classifier in SVM (Lee and Choi, 2003).

Other classification methods used in the recent BCI research include Particle Swarm Optimisation (PSO), Fisher classifier and Fuzzy logic. The PSO has been incorporated in (Satti et al., 2009) work to select subject specific-frequency band for an efficient tuned BCI system. The Fisher classifier was used by Bian et al. (2010) to recognize SSEVP signals generated from controlling a small ball movement on a computer screen. An investigation on the performance of Fuzzy Logic in detecting four different imagery tasks revealed that this technique could only provide an accuracy of 78% with a slow computation time (Saggiol et al., 2009).

From the literature, it is seen that there are numerous signal processing techniques employed for the feature extraction and the classification stages, but there are still some limitations to be considered. The drawbacks of these methods are that they could not produce enough accuracy for this field and do not work well when the data size is very large. Most of them require a lengthy training time. These limitations can be overcome by some future enhancements. To overcome these problems this study aims to introduce methods for the classification of epileptic EEG data and also for the identification of the MI based EEG data in BCI systems.

2.3 Summery

This chapter provides an overview of EEG signal classification and also provides necessary background knowledge related to the EEG. Firstly, this chapter presents an

outline of the human brain, the fundamentals of EEG, epileptic effect on EEGs and the basic concept of BCI technologies. In the EEG measurement, the cerebral cortex is the most relevant structure as it is responsible for higher order cognitive tasks. Due to its surface position, the electrical activity of the cerebral cortex has the greatest influence on EEG recordings. Then this chapter discusses about the classification of EEG signals and also reviews which methods were used for the EEG signal classification in the previous study. From the literature review of the EEG signal classification, it can be concluded that there are still some limitations associated with the existing methods. Hence developing new classification algorithms are needed for a reliable diagnosis and treatment of some neurodegenerative diseases.

In the next chapter, a novel method based on a simple random sampling technique and least square support vector machine (SRS-LS-SVM) is introduced to classify epileptic EEG signals. This approach is also tested to identify different categories of mental imagery tasks EEG signals in BCI applications.

CHAPTER 3
TWO-STAGE RANDOM SAMPLING WITH LEAST SQUARE SUPPORT VECTOR MACHINE

The development of classification techniques for the analysis of EEG signals is essential in the area of biomedical research. Classification techniques can help in the diagnosis and evaluation of brain diseases by differentiating EEG segments. This chapter introduces a new classification method based on simple random sampling technique with least square support vector machine (SRS-LS-SVM) to classify two-class of the EEG signals. This study aims to establish a method to determine an optimal classification scheme and also to infer clues about the extracted features by the combination of a simple random sampling (SRS) technique and a least square support vector machine (LS-SVM). The SRS technique is employed in two stages to extract features from the original data reducing the dimensionality of the data. Then the LS-SVM is implemented on the extracted feature vectors for the classification of the EEG signals. In this chapter, we proposed a new algorithm that can distinguish EEG segments from different pairs of two-class EEG signals in a multiclass EEG dataset.

The contents of this chapter have been published in the *international journal of biomedical engineering and technology* (Siuly et al., 2011d) and also in the *Proceedings of rough sets and knowledge technology* (Siuly et al., 2009).

3.1 Introduction

There is an ever-increasing need for developing automatic systems to evaluate and diagnose neurological disorder diseases and to prevent the possibility of the analyst missing information. In order to meet the growing need, in this chapter, a novel method has been developed to extract the relevant information from EEG recordings.

From pattern recognition point of view (Duda et al., 2001), one key problem is how to represent the large amount of recorded EEG signals for further analysis, such as classification. It is, firstly, important to extract useful features from raw EEG signals and then, use the extracted features for classification. In the literature,

numerous signal processing techniques are often employed for the feature extraction and the classification stage. The main drawback of these methods is that they do not work well when the data size is very large. To tackle the problem, this chapter introduces a new classification algorithm combining a simple random sampling (SRS) technique and a least square support vector machine (LS-SVM) to classify two-class EEG signals.

The idea of using the SRS for feature extraction is completely new in pattern recognition. The SRS is a very popular sampling technique in statistics to select representatives of a population to determine characteristics or features of the whole population as suggested by Van Dalen (Cochran, 1977; Islam, 2007). During this sampling process, each individual has the same probability of being chosen. Due to randomization, it is free from conscious and unconscious bias that researchers may introduce while selecting a sample. As simple random samples are a representative part of a population, it is a natural expectation that selecting samples from EEG signals may improve the accuracy of classification. This expectation is achieved in this chapter where the SRS is used in two stages to select representatives of EEG signals from the original data. Using the SRS technique, we compress many data points into fewer parameters, which are termed features. These features represent the behaviours of the EEG signals, which are particularly significant for recognition and diagnosing purposes. In this work, nine statistical features, such as *minimum*, *maximum*, *mean*, *median*, *mode*, *first quartile*, *third quartile*, *inter-quartile range* and *standard deviation* are extracted from EEG segments by the SRS technique before the classification. Those features are the most representative values to describe a set of EEG signals and are used as inputs in a LS-SVM for classification.

Recently, LS-SVMs are becoming increasingly popular as a powerful new tool for data classification and function estimation and has been applied to a variety of domains, such as bioinformatics (Brown et al., 2000) and pattern recognition (Burges, 1998), with a great success. Due to some modifications in the original support vector machine (SVM), the problem generated by the LS-SVM can be solved with a set of linear equations, which is less complex than the quadratic programming used in SVMs (Suykens et al., 2002). In this chapter, the LS-SVM with a radial basis function (RBF) kernel is employed for the classification of different pairs of two-class EEG signals by using sampling features as the inputs. This study aims to establish a method to classify two-class EEG signals using the SRS technique for

feature extraction and the LS-SVM, which is applied to the extracted features for classification.

The proposed method is tested on EEG epileptic data (EEG time series, 2005) and on a mental imagery tasks EEG database (Chiappa & Millán, 2005) to classify different pairs of two-class EEG signals. The method achieves 100% classification accuracy for the EEG epileptic data for healthy subjects with eyes open (Set A) and for epileptic patients during seizure activity (Set E), which is compared with two most recently reported methods. Furthermore, we use our method on a non-EEG data, a synthetic two-class problem from Ripley data (Ripley, 1996), to verify the effectiveness of this methodology. With this dataset, we obtain impressive results again. The performance of the proposed method is evaluated with respect to the sensitivity, specificity, classification accuracy and receiver operating characteristic (ROC) curve. A significant improvement in accuracy is achieved by using the LS-SVM on extracted features for the three databases.

3.2 Related Work

Up to now, several techniques have been proposed on the classification of EEG signals in different literature and diverse classification accuracies have been reported in the last decade for the EEG epileptic data (EEG time series, 2005) and the mental imagery tasks EEG data (Chiappa & Millán**,** 2005). Brief descriptions of the previous research are provided below.

Guler et al. (2005) assessed the diagnostic accuracy of recurrent neural networks (RNNs) using Lapunov exponents trained with the Levenberg-Marquardt algorithm on the EEG epileptic database for healthy subjects with eyes open (Set A), epileptic patients during seizure free-interval from epileptogenic zone (Set D), and epileptic patients during seizure (Set E). The total classification accuracy of that model was 96.79%. Jahankhanni et al. (2005) used a wavelet transform for feature extraction and neural networks for EEG signal classification on the EEG epileptic database of healthy subjects with eyes open (Set A) and epileptic patients during seizure activity (Set E) and they achieved 98% classification accuracy.

Subasi (2007) introduced a mixture of an expert model with a double-loop expectation-maximization algorithm for detection of epileptic seizures. He employed a wavelet transform for feature extraction on the same two sets of the EEG epileptic

data and achieved an overall accuracy of 94.5%. Polat et al. (2007) used an approach for the classification of epileptiform EEG using a hybrid system based on a decision tree classifier and the fast Fourier Transform. Their method obtained a 98.72% classification accuracy using a 10-fold cross-validation for the same data sets of the epileptic EEG data. Sun (2007a) repoted an improved random subspace method and its application to the EEG signal classification. He employed his proposed method on the mental imagery tasks EEG data for BCI Competition III (Data set V). From the experimental results, it was approved that the proposed method outperformed the random subspace method, and is robust on the choice of the number of nearest neighbors. The results also showed that the performance improvement of the improved random subspace method is larger than that of the random subspace method. Sun et al. (2007b) carried out an experimental evaluation of ensemble methods for the EEG signal classification on the mental imagery tasks data in brain-computer interface (BCI) Competition III (Data set V). The findings of that method were helpful in guiding the choice of classification algorithms for BCI applications.

Ubeyli (2008) used the mixture of an expert system employing wavelet coefficients for the classification of three EEG signals (healthy subjects with eyes open (Set A), epileptic patients during seizure free-interval from epileptogenic zone (Set D), and epileptic patients during seizure (Ser E) of the EEG epileptic database. The total accuracy obtained was 93.17% by this method.

Guo et al. (2009) introduced relative wavelet energy and artificial neural networks for the classification of EEG signals. They implemented their approach on the EEG epileptic database for healthy volunteers with eyes open (Set A) and epileptic patients during seizure activity (Set E). Their method attained a 95% classification accuracy. Chandaka et al. (2009) also utilized a cross-correlation aided SVM classifier for classifying EEG signals of the healthy subjects with eyes open (Set A) and epileptic patients during seizure (Set E) on the EEG epileptic data. An overall classification of 95.96% was reported in their paper.

Ubeyli (2010) reported a combination of the model-based methods and the least square support vector machine for the analysis of EEG signals. Three model-based methods, Burg autoregressive (AR), moving average (MA) and autoregressive moving (ARMA) were used for the feature extraction and the LS-SVMs were implemented for the classification of those features. The method achieved 99.56% total classification accuracy on the same dataset (Set A and Set E).

All of these work proposed different approaches to classify EEG signals through different models. Until now, no study has been reported in literature related to the SRS technique for the feature extraction on EEG signals. In this chapter, we propose a new approach based on the SRS technique with the LS-SVM denoted by SRS-LS-SVM to classify different pairs of two-class EEG segments, which is described in the following section.

3.3 Methodology

In this chapter, we develop a combined algorithm that employs the SRS technique for features extraction and LS-SVM for the EEG signal classification. The block diagram of the proposed method is depicted in Figure 3.1. This figure shows the different steps of the proposed method for the EEG signal classification.

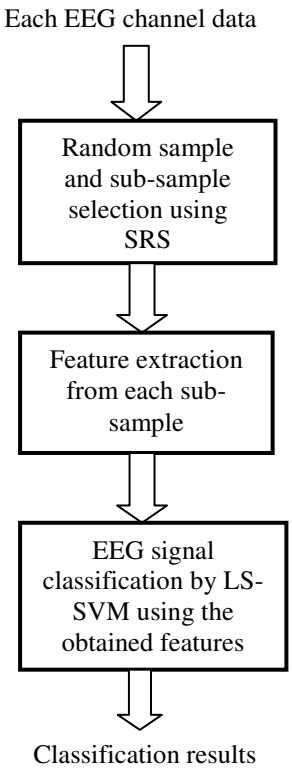


Figure 3.1: Block diagram of the SRS-LS-SVM method for EEG signal classification.

The proposed method mainly consists of three parts: (i) random sample and sub-sample selection using the SRS technique; (ii) feature extraction from each random

sub-sample and (iii) LS-SVM for the classification of EEG signals. These parts are described in detail in the following three subsections.

3.3.1 Random sample and sub-sample selection using SRS technique

Simple random sampling (SRS) is the purest form of the probability sampling. A random sample is obtained by choosing elementary units in such a way that each unit in the population has an equal chance of being selected. In this procedure, a random sample is free from sampling bias. For a given population, if the sample size is adequately taken then it represents the characteristics of the population.

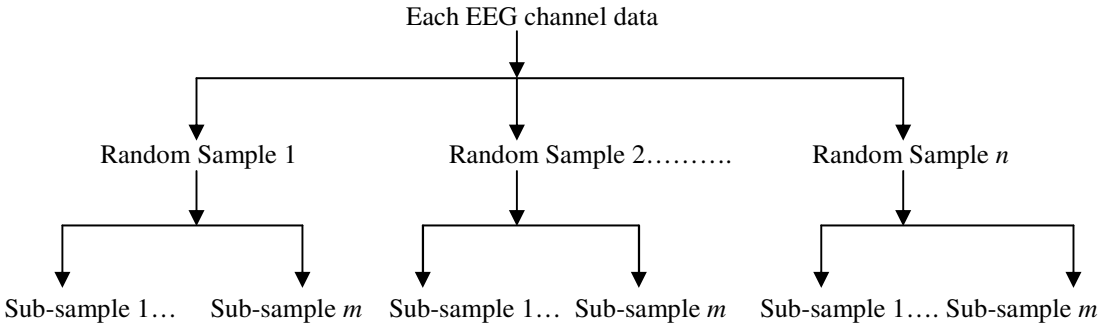


Figure 3.2: Random samples and sub-sample selection diagram using SRS technique.

There are different types of sampling techniques used in statistics (Cochran, 1977; Islam, 2007). The applications of these techniques depend on the structure of a population. In this chapter, the SRS technique is applied in two stages to select random samples and sub-samples from each EEG channel data file and finally different features are evaluated from each sub-sample set to represent the distribution of the EEG signals. These features reduce the dimensionality of the data discussed in Section 3.3.2. Figure 3.2 illustrates how different random samples and sub-samples are selected from each EEG data file. In this study, random samples and sub-samples are selected by the SRS method in the following two steps:

Step 1: n random samples of suitable sizes are selected from each EEG channel data, where n is the number of random samples and $n \geq 2$.

Step 2: *m* random sub-samples with appropriate sizes are then selected from each random sample, which are obtained in the first step. Here *m* is the number of random sub-samples and $m \geq 2$.

For any applications, the number of sample (*n*) and sub-sample (*m*) selections are chosen based on empirical approach. The consistency of the results will be improved if the number of samples is sufficiently large. In each step, the sample and sub-sample sizes are determined using a sample size calculator of the "Creative Research System" available in (Sample size calculator-online), considering a 99-100% confidence interval and a 99% confidence level. After Step 2, different statistical features, namely *minimum*, *maximum*, *mean*, and *median*, *mode*, *first quartile*, *third quartile* and *standard deviation* are calculated from each sub-sample set. They are discussed in the next subsection.

3.3.2 Feature extraction from different sub-samples

The goal of the feature extraction is to pull out features (special patterns) from the original data for a reliable classification. Feature extraction is the most important part of the pattern recognition because the classification performance will be degraded if the features are not chosen well (Hanbay, 2009). The feature extraction stage must reduce the original data to a lower dimension that contains most of the useful information included in the original vector. It is, therefore, necessary to find out the key features that represent the whole dataset, depending on the characteristics of the dataset.

In this chapter, nine statistical features are extracted from each sub-sample data points as they are the most representative values to describe the distribution of the EEG signals. The features are the *minimum*, *maximum*, *mean*, *median*, *mode*, *first quartile*, *third quartile*, *inter-quartile range* and *standard deviation* of the EEG data. Out of above nine features, *minimum*, *maximum*, *first quartile*, *second quartile* (also called *median*) and *third quartile* are together called a five number summary (De Veaux et al., 2008; Islam, 2004). A five number summary is sufficient to represent a summary of a large data. It is well known that a five number summary from a database provides a clear representation about the characteristics of a dataset.

Again a database can be symmetric or skewed. It is investigated that some of the EEG data files used in this study have symmetric distributions and others are

skewed distributions. For a symmetric distribution, an appropriate measure for measuring the centre and variability of the data are the *mean* and the *standard deviation*, respectively (De Veaux et al., 2008; Islam, 2004). For skewed distributions, the median and the *inter-quartile range* (IQR) are the appropriate measures for measuring the centre and spread of the data. On the other hand, *mode* is the most frequent value that is also a measure of locations in a series of data. Like *mean* and *median*, the *mode* is used as a way of capturing important information about a data set.

For these reasons, we consider these nine statistical features as the valuable parameters for representing the distribution of the EEG signals and also brain activity as a whole in this study. The obtained features are employed as the input for the LS-SVM. The description of the LS-SVM is provided in the following subsection.

3.3.3 Least square support vector machine (LS-SVM) for classification

Least square support vector machine (LS-SVM) is a relatively new powerful tool in the field of biomedical, which is employed basically for the classification purpose. The LS-SVM was originally proposed by Suykens and Vandewalle (1999) and corresponds to a modified version of a support vector machine (SVM) (Vapnik, 1995). Recently LS-SVM has drawn a great amount of attention into how to solve problems of pattern recognition by employing a kernel function. It solves a set of linear equations instead of a quadratic programming problem and all training points are used to model the LS-SVM. This approach significantly reduces the cost in complexity and computation time for solving the problem. The formulation of LS-SVM is briefly introduced as follows.

Consider a training set $\{x_i, y_i\}_{i=1,2,\ldots,N}$ where x_i is the ith input features vector of d-dimension and y_i is the class label of x_i, which is either +1 or –1. In the feature space, the classification function of the LS-SVM (Suykens et al., 2002; Guo et al., 2006) can be described as

$$y(x) = sign[w^T \phi(x) + b]. \tag{3.1}$$

where w is the weight vector, b is the bias term and $\phi(x)$ is the nonlinear mapping function that maps the input data into a higher-dimensional feature space. The weight vector, *w,* and bias term, *b,* need to be determined. In order to obtain *w*

and *b*, the following optimization problem to be solved is as follows (Suykens et al., 2002; Guo et al., 2006)

$$Min \quad J(w,b,e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{N} e_i^2. \tag{3.2}$$

Subject to the equality constraint

$$y_i[w^T \phi(x_i) + b] = 1 - e_i, \ i=1, 2,..., N. \tag{3.3}$$

Here γ is the regularization parameter, e_i is the classification error variable and J is the cost function which minimizes the classification error.

The Lagrangian can be defined for Equation (3.2) (Suykens et al., 2002; Guo et al., 2006) as

$$L(w,b,e;\alpha) = J(w,b,e) - \sum_{i=1}^{N} \alpha_i \{ y_i[w^T \phi(x_i) + b] - 1 + e_i \}. \tag{3.4}$$

Here α_i denotes Lagrange multipliers (which can be either positive or negative due to equality constraints). According to the conditions of Karush-Kuhn-Tucker (KKT) (Fletcher, 1987), we partially differentiate L and obtain formulas as follows:

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i \phi(x_i),$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0,$$

$$\frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i, \quad for\ i=1,..., N,$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \phi(x_i) + b + e_i - y_i = 0, \ for\ i=1,....., N.$$

Now we get

$$w = \sum_{i=1}^{N} \alpha_i \phi(x_i) = \sum_{i=1}^{N} \gamma e_i \phi(x_i) \tag{3.5}$$

Putting the value of w from equation (3.5) in equation (3.1), the following result is obtained:

$$y(x) = \sum_{i=1}^{N} \alpha_i \phi(x_i)^T \phi(x) + b = \sum_{i=1}^{N} \alpha \left\langle \phi(x_i)^T, \phi(x) \right\rangle + b \tag{3.6}$$

By eliminating e_i and w, the solution is given by the following set of linear equations:

$$\begin{pmatrix} 0 & 1_v^T \\ 1 & K + \frac{1}{\gamma} I \end{pmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{3.7}$$

Where $y = [y_1; ...; y_N], 1_v = [1; ...; 1], \alpha = [\alpha_1; ...; \alpha_N]$ and $K = K(x, x_i) = \phi(x)^T \phi(x_i)$, $i = 1, ..., N$. $K(x, x_i)$, which is called the inner-product kernel should satisfy the case of Mercer's condition.

It is seen from equation (3.7), all Lagrange multipliers (support vectors) are nonzero, which means that all training data contribute to the solution. After applying of the Mercer condition, the decision function of the LS-SVM is then constructed for the classification as follows:

$$y(x) = sign\left(\sum_{i=1}^{N} y_i \alpha_i K(x, x_i) + b \right) \tag{3.8}$$

Where α_i and b are the solutions to the linear system and $y(x)$ is the LS-SVM output (estimated class) for the input vector x. Further explanation of the input vector x of the LS-SVM is described in Section 3.5.1. There are different types of kernel function, for example linear kernel, polynomial kernel, radial basis function (RBF) and multi-layer perception (MLP) kernel. In our work, the RBF kernel, $K(x, x_i) = \exp\left(-\frac{\| x - x_i \|^2}{2\sigma^2} \right)$ is used as the kernel function because most of the biomedical researchers consider this function as an ideal one.

Thus, the LS-SVM model has two hyper parameters, γ and σ, that need to be determined a priori. In this chapter, the above LS-SVM algorithm is applied for classifying two-class EEG signals represented by the extracted features obtained by the SRS. The LS-SVM implementation is carried out in MATLAB environment (version 7.7, R2008b) using the LS-SVMlab toolbox, available in (LS-SVMlab toolbox (version 1.5)-online).

3.3.4 Performance evaluation

There are various types of methods for performance evaluations. In this study, the stability of the performance of the LS-SVM classifier is assessed based on different

statistical measurements, such as sensitivity, specificity, classification accuracy and a receiving operating characteristic (ROC) curve. The definition of sensitivity, specificity and accuracy (Guo et al., 2009) and the brief description of the ROC curve (Fawcett, 2006) are explained in the following:

- Sensitivity: the number of true positive decisions divided by the number of actual positive cases;
- Specificity: the number of true negative decisions divided by the number of actual negative cases;
- Classification accuracy: the number of correct decisions divided by the total number of cases.
- ROC curve: it is a very useful tool for visualizing, organizing and selecting a classifier based on its performance (Fawcett, 2006). A ROC curve plots the sensitivity (true positive rate) on the X-axis and the 1-specificity (false positive rate) on the Y-axis. The area under the ROC curve is an important value to evaluate the performance of a binary classifier and its value is always between 0 and 1. If the area of the ROC curve is 1, it indicates that the classifier has a perfect discriminating ability. If the area equals 0.5, the classifier has no discriminative power at all and no suitable classifier should have an area under this curve less than 0.5 (Fawcett, 2006).

3.4 Experimental Data

In this study, an EEG epileptic dataset and a mental imagery tasks EEG dataset for BCI Competition III (Data set V) are used for the proposed method to classify different pairs of two-class EEG signals. In order to further test the validity of the proposed algorithm, the synthetic two-class problem from Ripley data (1996) is employed. These three databases are described in the following three subsections.

3.4.1 EEG epileptic data

The EEG epileptic data developed by the Department of Epileptology, University of Bonn, Germany described in (Andrzejak et al., 2001), is publicly available in (EEG time series, 2005). The whole database consists of five EEG data sets (denoted A-E), each containing 100 single-channel EEG signals of 23.6s from five separate classes.

Each signal was chosen after visual inspection for artefacts, such as causes of muscle activities or eye movements. All EEG recordings were made with the same 128-channel amplifier system, using an average common reference. The recoded data was digitised at 173.61 samples per second using 12-bit resolution. Band-pass filter settings were 0.53-40 Hz (12dB/oct). Set A and Set B were collected from surface EEG recordings of five healthy volunteers with eyes open and eyes closed, respectively. Sets C, D and E were created from the EEG records of the pre-surgical diagnosis of five epileptic patients.

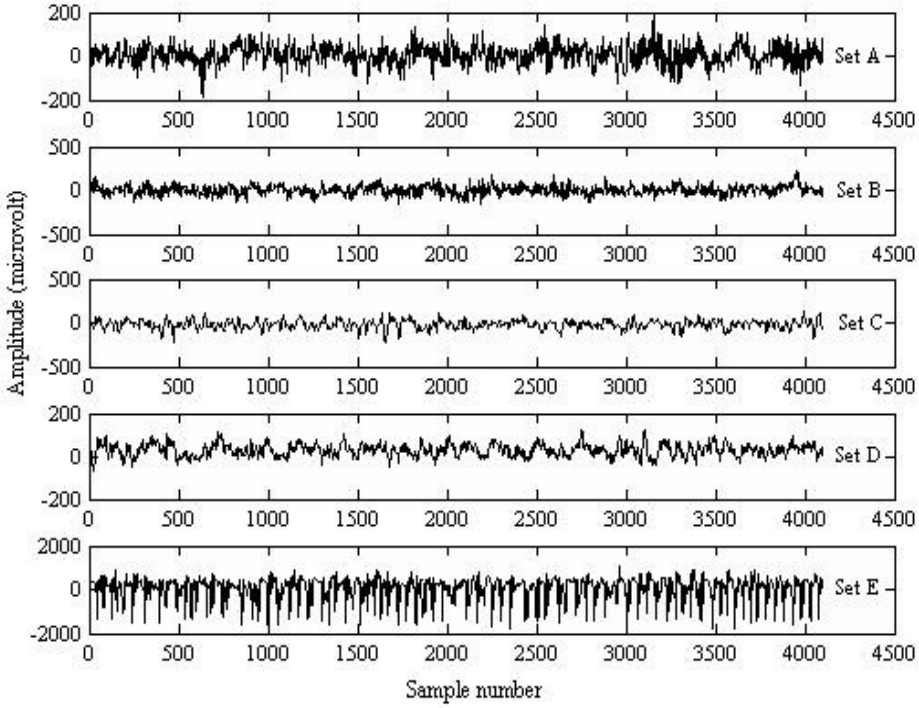


Figure 3.3: Example of five different sets of EEG signals taken from different subjects.

Signals in Set C and Set D were recorded in seizure-free intervals from five epileptic patients from the hippocampal formation of the opposite hemisphere of the brain and from within the epileptogenic zone, respectively. Set E contains the EEG records of five epileptic patients during seizure activity. Figure 3.3 depicts some examples of five EEG signals (Set A to Set E). The amplitudes of those EEG recordings are given in microvolts (μV).

3.4.2 Mental imagery tasks EEG data

The data set V, for brain computer interface (BCI) Competition III, contains EEG recordings from three normal subjects during three kinds of mental imagery tasks, which are the imagination of repetitive self-paced left hand movements (class 1), the imagination of repetitive self-paced right hand movements (class 2), and generation of different words beginning with the same random letter (class 3) (Millán**,** 2004 and Chiappa & Millán, 2005)**.** Figure 3.4 shows the exemplary EEG signals for left hand movements (class 1), right hand movements (class 2) and word generation (class 3) taken from Subject 1 for this dataset. In these tests, subjects sit in a normal chair, with relaxed arms resting on their legs. For a given subject, there are four non-feedback sessions recorded on the same day, each lasting four minutes or so with breaks of 5-10 minutes in between.

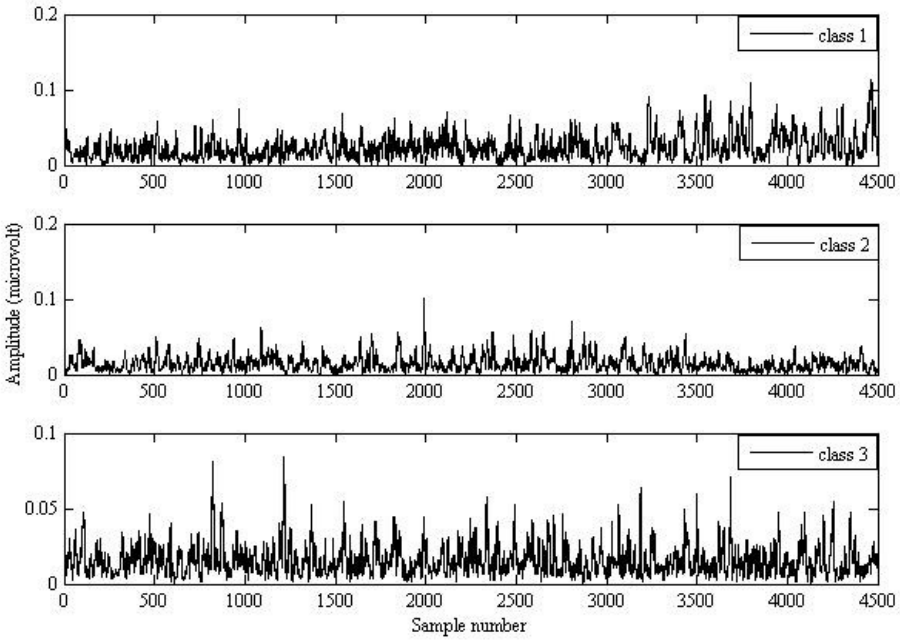


Figure 3.4: Exemplary EEG signals for left hand movements (class 1), right hand movements (class 2) and word generations (class 3) taken from Subject 1.

The subjects performed a given task for about 15 seconds and then switched randomly to the next task at the operator's request (Chiappa & Millán, 2005). The sampling rate of the raw EEG potential signals is 512 Hz. The signals are first spatially filtered by means of a surface Laplacian. Then, every power spectral density in the band of 8-30 Hz is estimated using the last second of data with a frequency

resolution of 2 Hz for the 8 centro-parietal channels (closely related to the current mental tasks) C3, Cz, C4, CP1, CP2, P3, Pz and P4. EEG recordings of 12 frequency components are obtained from each of the 8 channels, producing a 96 dimensional vector.

3.4.3 Two-class synthetic data

This data contains synthetic two-class problem from Ripley data, which was used in (Ripley, 1996). The dataset is publicly available in (Ripley data-online). The well-known Ripley dataset widely used as a benchmark consists of two classes. Each pattern has two real-valued co-ordinates and a class that can be 0 or 1. Each class corresponds to a bimodal distribution that is an equal mixture of two normal distributions (Ripley, 1996). Covariance matrices are identical for all the distributions and the centre is different. The training set consists of 250 patterns (125 patterns in each class) and the test set consists of 1000 patterns (500 patterns in each class). This data is interesting because there is a big overlap between both classes and the number of the test data is much larger than the number in the training pattern.

3.5 Experimental Results and Discussions

As mentioned before in Section 3.4, three datasets are applied for the proposed approach in this chapter. The proposed approach is employed to classify two-class EEG signals to the EEG epileptic database and the mental imagery tasks EEG database (Data set V for BCI competition III), separately. The two-class synthetic data from Ripley (1996) is applied to judge the efficacy of the proposed algorithm. Matlab version 7.7.0 (2008b) is used for all implementations. The experimental results of the three datasets are discussed below.

3.5.1 Results for the EEG epileptic dataset

The EEG epileptic data consists of five sets (Set A-Set E) having 100 single-channels of EEG signals (refer to Section 3.4.1 for details). Each data set contains 100 data files and each data file holds one channel EEG data, which has 4096 data points. In our proposed method, we consider n =10 (the number of random samples), m (the

number of random sub-samples)=5, random sample size=3287 and random sub-sample size=2745 for each EEG channel data of a set (a class) from the epileptic data. Thus 10 random samples of sizes 3287 are selected from each EEG channel data file and 5 random sub-samples of sizes 2745 are chosen from each random sample by the SRS technique. Then nine statistical features (*minimum*, *maximum*, *mean*, *median*, *mode*, *first quartile*, *third quartile*, *inter-quartile range* and *standard deviation*) are calculated from each sub-sample. Thus we obtain a feature vector set of size [1×45] from 5 random sub-samples of each random sample. Hence, for 10 random samples of an EEG channel data file, we acquire a feature vector set of size [10×45] and consequently we obtain a feature vector set of size [1000×45] for 100 channel data files of a data set. Therefore, we obtain a feature vector set of size [2000×45] for two data sets and [5000×45] for five data sets. Finally the feature vector set is employed as the input in the LS-SVM for the training and testing purposes.

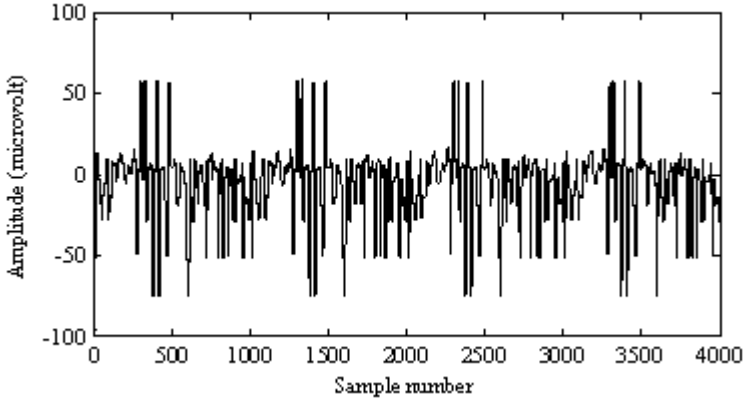


Figure 3.5: Exemplary mean feature points obtained by the SRS from healthy subjects with eyes open (Set A).

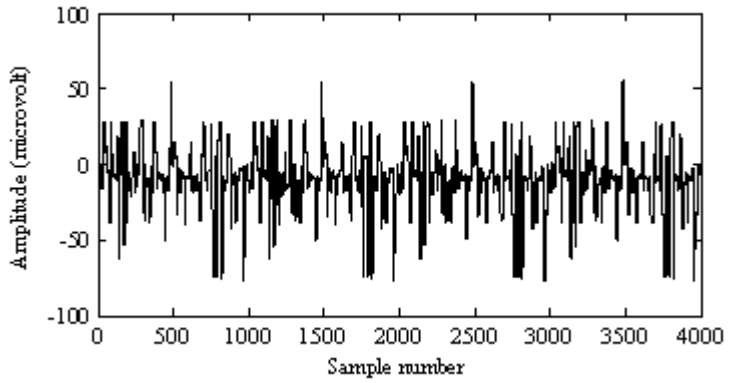


Figure 3.6: Exemplary mean feature points obtained by the SRS from epileptic patients during seizure-free intervals within the hippocampal formation of the opposite hemisphere of the brain (Set C).

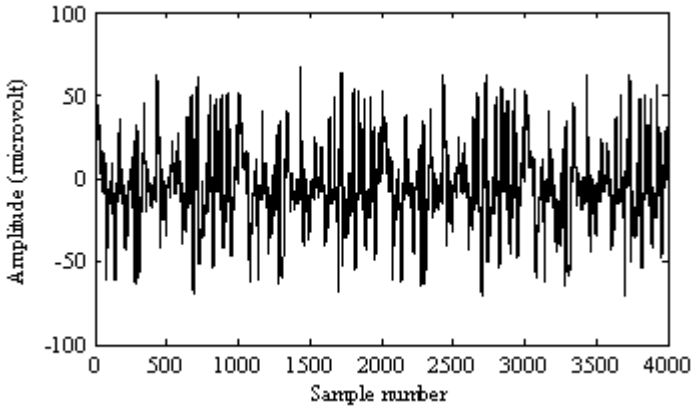


Figure 3.7: Exemplary mean feature points obtained by the SRS from epileptic patients during seizure activity (Set E).

Figures 3.5-3.7 show examples of the mean feature points of EEG recording from the healthy subjects with eyes open (Set A), epileptic patients during seizure-free interval within hippocampal formation of the opposite hemisphere of the brain (Set C) and epileptic patients during seizure activity (Set E), respectively, by the proposed sampling method. In Figures 3.5-3.7, it is noted that only 4000 feature points out of 45000 are used for each set of the EEG epileptic database to show typical results of the proposed SRS. Set A, Set C and Set E from the EEG epileptic data are used as representatives. These figures indicate that the representative patterns of the original data are detected by the SRS technique.

In this study, we obtain 5000 vectors (1000 vectors from each data set) of 45 dimensions (the dimensions of the extracted features) for the five EEG data sets. Five experiments are performed for the EEG epileptic dataset. In each experiment, we use a pair of two-class, which has 2000 vectors with 45 dimensions taking 1000 vectors from each class. From each class, we use the first 500 vectors for the training and the rest 500 vectors for the testing. Thus, for each pair of two-class, we obtain 1000 vectors of 45 dimensions as the training set and 1000 vectors with the same dimensions as the testing set. In equation (3.8), we employ the training set as x and the class label of the training set as y. In this study, the training vectors are applied to train the LS-SVM classifier, where the testing vectors are used to verify the accuracy and the effectiveness of the trained LS-SVM for the classification of two-class of EEG signals. The LS-SVM with RBF kernel function is employed as an optimal kernel function over different kernel functions that were tested.

There are two important parameters (γ, σ^2) in the LS-SVM, which should be appropriately chosen for achieving a desirable performance. The values of the two parameters significantly affect the classification performance of the LS-SVM. In order to achieve the best results, LS-SVM are trained with different combinations of the parameters, γ and σ^2. The proposed method is conducted with different pairs of the five EEG data sets in the EEG epileptic data and the best classification result is obtained for the pair of Set A and Set E when γ=1 and $\sigma^2 = 1$ with zero misclassification rate in both the training and testing results. For the pairs of the Sets B and E, Sets C and E, Sets A and D and Sets D and E, we achieved optimal results using different combinations of the parameters γ and σ^2, which were (10,1), (10,10), (1,10) and (2,1), respectively.

The experimental results for five pairs of the EEG epileptic data sets are shown in Table 3.1. Here it is observed that the highest classification accuracy obtained is 100% for healthy subjects with eyes open (Set A) and epileptic patients during seizure activity (Set E). It is known that seizures produce abnormal electronic signals in the brain and there are large variations among the recorded EEG values in Set A and Set E. Due to the nature of the large differences, it is relatively easier to classify Set A and Set E as demonstrated by the 100% of the classification accuracy in the experiment. For Set B and Set E, 99.50% classification accuracy is obtained. A classification accuracy of 96.40% is achieved for Set C and Set E. As it can be seen from Table 3.1, the classification accuracy is 88.00% for Sets A and D. The accuracy is not significant as the EEG data from Set A and Set D is more analogous to each other.

Table 3.1: Experimental results and the area values under ROC curve for two-class pairs of the EEG signals for the EEG epileptic database.

Different data sets	Sensitivity (%)	Specificity (%)	Classification accuracy (%)	Area under ROC curve
Set A and Set E	100.00	100.00	100.00	1.00000
Set B and Set E	99.80	99.20	99.50	1.00000
Set C and Set E	98.00	94.80	96.40	1.00000
Set A and Set D	94.00	82.00	88.00	0.96812
Set D and Set E	88.00	100.00	94.00	1.00000
Mean/average	95.96	95.20	95.58	-
Standard deviation	5.0604	7.6890	4.4869	-

On the other hand, the method obtains 94.00% classification accuracy for Set D and Set E. It is also noted from Table 3.1 that the area values under ROC curve for all

pairs of EEG data sets is 1 except for the pair of Set A and Set D. Hence it is apparent that the proposed approach has a high discriminating capability to classify EEG signals and produces excellent results for classifying EEG brain signals between Set A and Set E. As shown in Table 3.1, the average value and standard deviation of classification accuracies for the different combinations of the EEG data sets were obtained as 95.58% and 4.4869, respectively. The results demonstrate that the method proposed in this paper is a very promising technique for the EEG signal classification.

The proposed approach is capable of classifying the EEG signals for Set A and Set E with 100% classification accuracy. The result indicates that the proposed method has significantly improved in performance compared to the two most recent methods, LS-SVM and model-based methods by Ubeyli (2010) and cross-correlation aided SVM by Chandaka et al. (2009). The performance comparison of the present method with the two most recently reported methods for Set A and Set E are shown in Table 3.2.

Table 3.2: Comparison of performance of our proposed method with two most recently reported methods for Set A and Set E of the EEG epileptic database.

Different methods	Sensitivity (%)	Specificity (%)	Classification accuracy (%)
SRS technique and LS-SVM (proposed)	100.00	100.00	100.00
LS-SVM and model-based methods (Ubeyli, 2010)	99.50	99.63	99.56
Cross-correlation aided SVM (Chandaka et al., 2009)	92.00	100.00	95.96

It is shown in Table 3.2 that Ubeyli in 2010 obtained 99.56% classification accuracy when she applied the LS-SVM and model-based methods on Set A and Set E (the same data sets used in this paper) for EEG signal classification. At the same time, Chandaka et al. (2009) used a cross-correlation aided SVM approach to classify the EEG signals for the same data sets and reported the classification accuracy as 95.96%. By contrast, our proposed method reaches 100% classification accuracy for the same pair of data sets. The results demonstrate that our approach can classify more accurately the EEG signals of all epileptic and healthy subjects using the extracted features from the SRS technique.

The Receiving operating characteristic (ROC) curve is drawn in Figure 3.8 on the testing vector set for the EEG data sets of healthy subjects with eyes open (Set A) and epileptic patients during seizure activity (Set E). The ROC curve presents an

analysis of the sensitivities and specificities when all possible sensitivity/specificity pairs for the full range of experiments are considered. A good test is the one for which sensitivity (true positive rate) rises rapidly and 1-specificity (false positive rate) hardly increases at all until sensitivity becomes high (Ubeyli, 2008). From Figure 3.9, it is seen that the area value of the ROC curve is 1, which indicates that the LS-SVM model has effectively classified the EEG signals using the extracted features from Sets A and E. Therefore, it is obvious that the sampling features well represent the EEG signals and the LS-SVM classifier trained on these features achieves a high classification accuracy.

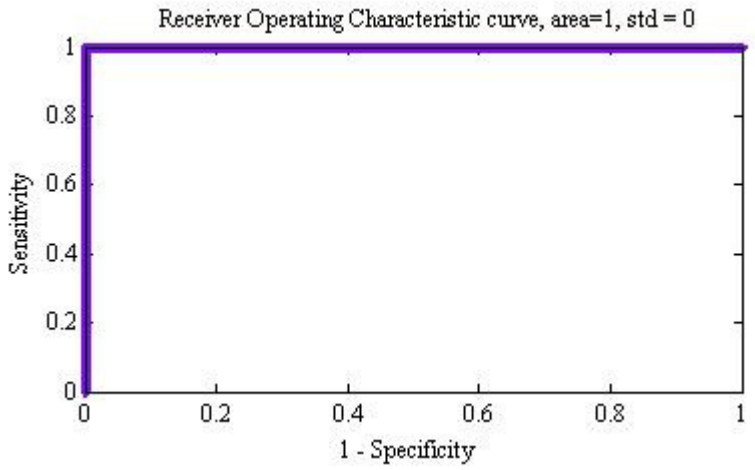


Figure 3.8: ROC curve for healthy subjects with eye open (Set A) and epileptic patients during seizure activity (Set E) in the EEG epileptic data.

3.5.2 Results for the mental imagery tasks EEG dataset

The mental imagery tasks EEG database contains EEG recordings from three normal subjects during three kinds of mental imagery tasks. The EEG data were recorded for a subject with four non-feedback sessions on the same day. The number of recorded samples in four sessions for each subject is given in Table 3.3. According to the data description (see Section 3.4.2), each of the eight channels has 12 frequency components and an EEG sample obtained from the eight channels is a 96 dimensional vector. In this study, we use the first three session's EEG recorded data of three kinds of the mental imagery tasks for each subject. Thus 10528 vectors of 96 dimensions are obtained for Subject 1, 10400 vectors for each of Subject 2 and Subject 3 with the same dimensions as the original data in the experiment.

Table 3.3: Number of recorded values in four sessions from the mental imagery tasks EEG data (Data set V for BCI competition III).

Subjects	Session 1	Session 2	Session 3	Session 4
1	3488	3472	3568	3504
2	3572	3456	3472	3472
3	3424	3424	3440	3788

In the proposed approach (see Section 3.3.1), we take n=10, m=5 for each channel EEG data of all subjects from the mental imagery tasks. For Subject 1, the random sample and sub-sample sizes of each channel EEG data are determined as 2490 and 2166 for class 1; 2862 and 2442 for class 2; and 3318 and 2767 for class 3, respectively. For Subject 2, we get the random sample and sub-sample sizes as 2513 and 2183, respectively, for class 1; 2829 and 2418 for class 2; and 3246 and 2716 for class 3. For Subject 3, the random sample and sub-sample sizes are obtained 2829 and 2418 for class 1; 2851 and 2434 for class 2 and 2851 and 2434 for class 3, respectively. Here, it is noted that random samples and sub-sample sizes are different for each class of a subject as the number of observations of all classes are not the same in the mental imagery tasks EEG data. The nine features (see in Section 3.3.2) are then calculated from each sub-sample of a class in a subject. Finally the LS-SVM algorithm is trained on these features for classifying EEG signals.

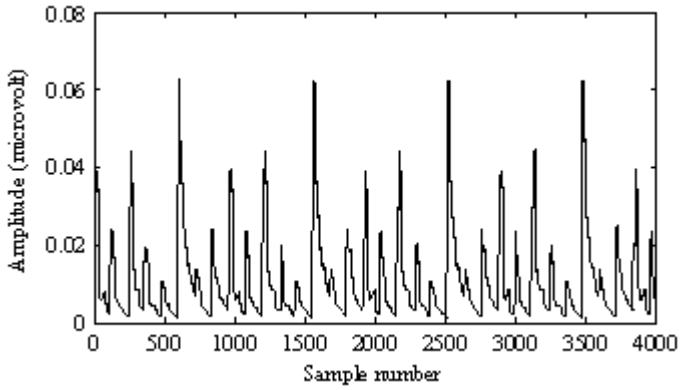


Figure 3.9: Exemplary mean feature points obtained by the SRS from left hand movements (class 1) of Subject 1.

Figure 3.9 shows the extracted mean feature points of EEG recorded data for imagination of repetitive self-paced left hand movements (class 1) of Subject 1 from the mental imagery tasks data by the proposed SRS technique. It is seen from Figure

3.9 that 4000 feature points out of 43200 of the mental imagery tasks EEG dataset are presented to display representative outcomes of the proposed SRS technique.

From the mental imagery tasks EEG data, we select 960 vectors of 45 dimensions from each class of a subject. For the dataset, nine experiments are carried out using a pair of two-class EEG data. Each experiment uses 1920 vectors of 45 dimensions for a two-class data with 960 vectors from each class as the training set. The performance is also evaluated based on this vector set.

After the feature extraction, a LS-SVM classifier has been trained using these extracted features. In this database, the best classification results are found for all pairs when γ=10000 and σ^2 =1. From Table 3.4, it is observed that classification accuracies for different pairs of different subjects are obtained ranging from 93% to 100%.

Table 3.4: Experimental results for different pairs of two-class EEG signals for the mental imagery tasks EEG data (Data set V for BCI competition III).

Subject	Pair of Classes	Sensitivity (%)	Specificity (%)	Classification accuracy (%)	Area under ROC curve
1	class 1 and class 2	97.19	98.75	97.97	0.99792
	class 1 and class 3	97.50	98.12	97.81	0.99682
	class 2 and class 3	94.37	93.02	93.70	0.98870
2	class 1 and class 2	100.00	100.00	100.00	1.00000
	class 1 and class 3	99.79	99.58	99.69	0.99990
	class 2 and class 3	99.79	99.58	99.69	0.99995
3	class 1 and class 2	100.00	99.79	99.90	0.99997
	class 1 and class 3	99.90	100.00	99.95	0.99999
	class 2 and class 3	99.90	99.90	99.90	1.00000
Average /Mean		98.72	98.75	98.73	-
Standard deviation		1.9717	2.2399	2.0739	-

The average classification accuracy and standard deviation for all classification accuracies of all subjects are achieved at 98.73% and 2.0739, respectively. The area values under the ROC curve for all pair's classification are close to 1. The experimental results, therefore confirm that the obtained features through the SRS technique actually represent the most important information in the recorded EEG data and our approach is powerful for EEG signal classification.

3.5.3 Results for the two-class synthetic data

In this section, we discuss the experimental results of a non-EEG dataset, which is a two-class synthetic data from Ripley (1996). There are two sets of data, training and

testing sets (see more in Section 3.4.3). The original Ripley data structure and the feature vectors obtained by the SRS are presented in Table 3.5.

In this experiment, the proposed method is implemented on the 250 given training data with two dimensions to find training feature vector set. Choosing n=10, m=5, the random sample size=246 and the random sub-sample size=242 in the experiment. Then the mentioned nine features are computed from each sub-sample.

Table 3.5: The original two-class synthetic data from Ripley (1996) and the extracted feature vectors by the SRS technique.

Classes	Original data		Features vectors of 45 dimensions	
	Training data	Testing data	Training vector set	Testing vector set
class 1	125	500	10	10
class 2	125	500	10	10
Total	250	1000	20	20

Thus we obtain 10 vectors of 45 dimensions as a training vector set. Similarly, we employ the algorithm to the 1000 given testing data of two dimensions and take n=10, m=5 with now a random sample size of 943 and a random sub-sample size of 892. The nine features are extracted from each sub-sample. As the result, we get 10 vectors of 45 dimensions, which are used as the test vectors in the experiment. The optimal classification results were achieved when we set γ=1 and σ^2 =1 in the training and testing.

Classification results of the algorithm are displayed by a confusion matrix in Table 3.6. In the confusion matrix, each cell consists of the number of vectors classified for the corresponding combinations of the predicted and actual outputs.

Table 3.6: Confusion matrix for Ripley data (1996).

		Predicted value	
		class 1 (vectors)	class 2 (vectors)
Actual outcome	class 1	10	0
	class 2	0	10

As it can be seen in Table 3.6, an overall 100% classification accuracy is obtained by the SRS-LS-SVM approach. The correct classification rate is 100% for class 1 and 100% for class 2. According to the confusion matrix, no misclassification is occurred using the proposed method.

Table 3.7 presents sensitivity, specificity, classification accuracy and the area

value under the ROC curve of the LS-SVM classifier. The method results in 100% sensitivity, 100% specificity and 100% classification accuracy on the Ripley data set. The area under ROC curve is 1 for the dataset, which confirms a perfect classification ability of the approach.

The outcomes of this dataset also prove that the proposed method can be successfully used in any classification area.

Table 3.7: Experimental results for the Ripley data (1996).

Statistical parameters	Value
Sensitivity	100%
Specificity	100%
Classification Accuracy	100%
Area under ROC curve	1.000

The experiential results from the above three databases used in this study demonstrate that the SRS is able to effectively extract the features from the original data, which is very important for a successful classification by the LS-SVM. They also demonstrate that the SRS-LS-SVM is a very promising approach for pattern classification.

3.6 Conclusions

This chapter presents the development of a novel signal classification algorithm for classifying two categories of EEG signals. The proposed method introduces the SRS technique for feature extraction and a LS-SVM classifier with a RBF kernel function for the classification of any two-class of EEG signals using sampling features as the inputs. The experimental study is conducted with different pairs of two-class EEG signals on an EEG epileptic database and a mental imagery tasks EEG database for BCI Competition III (Data set V), separately. The method achieves a 95.96% average sensitivity, 95.20% average specificity and 95.58% average classification accuracy for the EEG epileptic data. From the mental imagery tasks EEG database, we obtain an average of 98.72% sensitivity, 98.75% specificity and 98.73% classification accuracy using different pairs of two-class EEG signals from three subjects. We are able to achieve 100% classification accuracy on the EEG epileptic database for the pair of healthy subjects with eyes open and epileptic patients during seizure activity. For the same pair of EEG epileptic data, the classification accuracy of Ubeyli's

method (2010) was reported at 99.56% and Chandaka et al.'s method (2009) was 95.96%. It is found that our results are the highest classification accuracy achieved for that pair of EEG data to date, to the best of our knowledge. To determine the effectiveness of the method on non-EEG data, the proposed algorithm is also applied to the synthetic two-class problem from Ripley data set (1996). The sensitivity, specificity and classification accuracy rate were found to be a 100% for the values of the dataset. The results demonstrate that the proposed methodology is superior. The experimental results also indicate that the SRS is efficient for extracting features representing the EEG signals. The LS-SVM classifier has the inherent ability to solve a pattern recognition task for the sampling features.

CHAPTER 4
CLUSTERING TECHNIQUE BASED LEAST SQUARE SUPPORT VECTOR MACHINE

In Chapter 3, we have developed and studied the two-stage SRS technique based on the LS-SVM (called SRS-LS-SVM) for the classification of EEG signals. The experimental results show that the algorithm works well to solve a pattern recognition task. However, the SRS system does not use all sample points to make representative features for classification. Sometimes valuable sample points can be omitted during sampling which may degrade accuracy. Again, the SRS-LS-SVM algorithm takes more time during experiments due to replications of random samples in the SRS technique.

To overcome the problems, this chapter proposes a clustering technique based LS-SVM algorithm named as CT-LS-SVM for the separation of EEG signals. Decision making is performed in two stages. In the first stage, the clustering technique (CT) has been used to extract representative features from EEG data. In the second stage, the LS-SVM is applied to the extracted features to classify two-class EEG signals. In this chapter, we investigate the performance of the CT-LS-SVM algorithm in the EEG signal classification with respect to accuracy and executions (running) time of experiments and also compare them with the SRS-LS-SVM algorithm. Again, the performance of the proposed method is compared with other existing methods in the literature. The CT-LS-SVM approach helps distinguish two categories of EEG signals and provide clinical information about patients who are potentially having epilepsy or neurological disorders or mental or physiological problems with different treatments.

The contents of this chapter was published in *Computer Methods and Programs in Biomedicine* (Siuly et al., 2011c) and also in the *Proceedings of the 2010 IEEE/ICME International Conference on Complex Medical Engineering (CME2010)* (Siuly et al., 2010).

4.1 Introduction

In recent years, there has been an increasing interest in the applications of machine learning techniques to classify EEG signals, which assist physicians with an appropriate diagnosis of brain diseases. Advanced signal classification techniques for the analysis of EEG signals is essential for developing and understanding the current biomedical research. The classification techniques generally work in two stages, where features are extracted from raw EEG data in the first stage and then the obtained features are used as the input for the classification process in the second stage. It is important to note that features are the compressed parameters that characterize the behaviour of the original data. Feature extraction is the most important part of the pattern recognition process because the classification performance will be degraded if the features are not chosen well. In the present study, the clustering technique (CT) algorithm is used to extract statistical features representing EEG signals from the original data. The LS-SVM is employed for the classification of two-class EEG signals using the obtained features as the input. Then the proposed approach with respect to the classification accuracy and the execution time is compared to the SRS-LS-SVM algorithm (Siuly et al. 2011d). We also compare our experimental results with other reported methods for three databases.

In this chapter, the CT approach is proposed as a new concept for feature extraction from the EEG data. In this procedure, each set of EEG channel data is divided into n (n=16) mutually exclusive groups named clusters with a specific time duration. Each cluster is again partitioned m (m=4) into sub-clusters over a specific time period and nine statistical features, such as *minimum*, *maximum*, *mean*, *median*, *mode*, *first quartile*, *third quartile*, *inter-quartile range* and *standard deviation* are extracted from each sub-cluster, representing EEG signals. The same features are used in the SRS-LS-SVM algorithm in Chapter 3. These features are applied to the LS-SVM classifier as the input for classifying two-class EEG signals. The proposed approach has two main advantages compared to the SRS-LS-SVM. The first advantage is that this method uses all data points for experiments. The second advantage is that by using the CT technique, it takes much less time to run the program. The proposed approach is very simple and thus flexible for the EEG signal classification.

The proposed method is tested on an EEG epileptic EEG dataset (EEG time series, 2005) and a mental imagery tasks EEG database (Chiappa and J.R. Millán, 2005) used in Chapter 3. A new data set, motor imagery EEG data (BCI Competition III, 2005) is added to evaluate the performance of the proposed method in this chapter. Different numbers of features extracted from different time durations are used in the classification. In this chapter, the division of the feature vectors is considered two ways for performance calculation e.g. two groups and 10-fold cross validation process. When feature vector sets are divided into two groups for training and testing, the method achieves an average sensitivity, specificity and classification accuracy of 94.92%, 93.44% and 94.18% respectively, for the EEG epileptic data; 83.98%, 84.37% and 84.17% respectively, for the motor imagery EEG data; and 64.61%, 58.77% and 61.69%, respectively, for the mental imagery tasks EEG data. We apply a 10-fold cross validation method (Abdulkadir, 2009) to evaluate the classification accuracy of the proposed method and we obtain an overall classification success rate of 94.12% for the epileptic data, 88.32% for the motor imagery data and 61.14% for the mental imagery tasks data. This method attains a 99.90% classification accuracy for healthy subjects with eyes open (Set A) and epileptic patients during seizure activity (Set E) in the EEG epileptic dataset. We compare the results with other existing methods in the literature. The experimental results show that the proposed approach takes much less running time than the SRS-LS-SVM technique. The performance of the proposed method is evaluated with respect to the sensitivity, specificity and classification accuracy.

4.1.1 Objective of the research

The main objective of the proposed CT-LS-SVM approach is to develop a system that can distinguish two categories of EEG signals and to investigate whether the CT method is suitable for the feature extraction from EEG data. This algorithm uses all data points of every EEG signals. The two most important criteria of choosing a best method, which are accuracy and computational efficiency, are used in this study for assessment. This study presents a method for two-class EEG signal classification that improves the classification accuracy and reduces the computational time during program running. We investigate the recital of our approach in terms of accuracy and computational efficiency and also compare them with the SRS-LS-SVM method. In

this chapter, we evaluate the performance of the proposed method in two ways; (i) dividing feature set into two groups as training and testing set and (ii) using the 10-fold cross-validation method.

4.2 Proposed Methodology

In the literature, numerous techniques have been used to obtain representations and to extract features of interest for classification purposes. Until now, to the best of our knowledge, there is no study that is related to the CT approach for feature extraction of EEG signals. In this chapter, we propose a new algorithm of the CT-LS-SVM for classifying EEG signals. The block diagram of the proposed CT method based on the LS-SVM for EEG signals classification is shown in Figure 4.1.

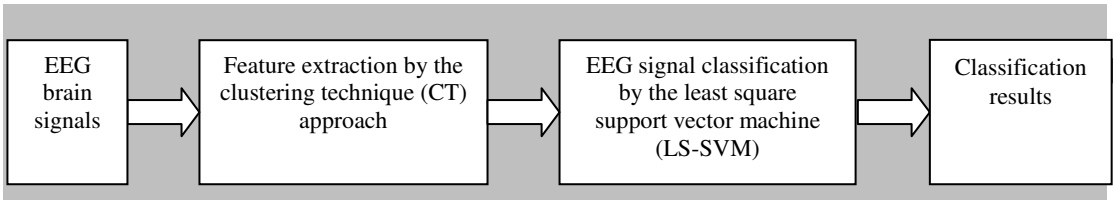


Figure 4.1: Block diagram of the proposed methodology for EEG signal classification.

The first block is the input of EEG brain signals and the second block is the feature extraction using the CT approach, which is responsible for data reduction and to capture most representative features from the original EEG patterns. The obtained features are used for classification through the LS-SVM classifier in the third block. The classification result is obtained in the fourth block. The following subsections describe this method in details.

4.2.1 Clustering technique (CT) for feature extraction

The design of the clustering technique (CT) is completely new in pattern recognition for feature extraction. As EEG signals are aperiodic and non-stationary, we divide the EEG signal of each channel into groups (clusters) and sub-groups (sub-clusters) with a specific time period. To characterize brain activities from the recording, several features are computed from each segmented sub-group. These features allow the representation of each segment as a point in the input vector space. In this

chapter, the CT method is proposed for feature extraction from the original EEG database. This approach is conducted in three stages, and determines different clusters, sub-clusters and statistical features extracted from each sub-cluster.

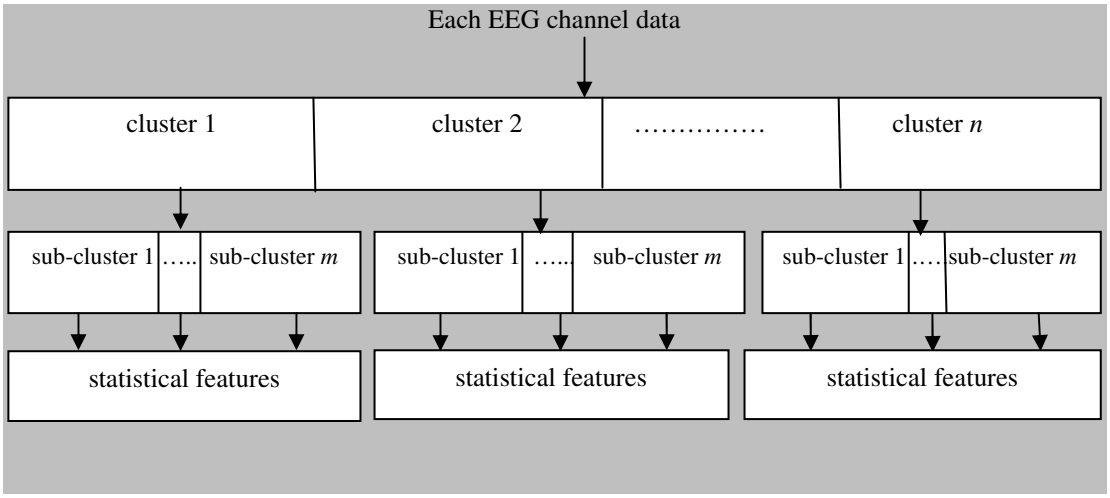


Figure 4.2: Clustering technique diagram for obtaining different clusters, sub-clusters and statistical features.

Figure 4.2 depicts the procedure of the CT method on how different clusters, sub-clusters and statistical features are obtained from the EEG channel data in three stages. These stages are discussed in the following three sub-sections.

4.2.1.1 Stage 1: Determination of clusters

In this technique, each EEG channel data is considered as a population. This population is divided into n groups with a specific time duration, which are called clusters (see Figure 4.2). Here n is the number of clusters and $n \geq 1$. For any applications, the numbers of clusters (n) are determined empirically over time.

4.2.1.2 Stage 2: Determination of sub-clusters

In this stage, each cluster is partitioned into m sub-clusters based on a specific time period. In this approach, m is the number of sub-clusters and $m \geq 1$, where the value of m is determined on an empirical basis over time. The sizes of each cluster and sub-cluster are automatically defined if a time period is fixed.

4.2.1.3 Stage 3: Statistical feature extraction

The following nine statistical features of each sub-cluster of each EEG channel data are used as the valuable parameters for the representation of the characteristics of the original EEG signals.

(I) *Minimum (X_{Min})*

(II) *Maximum (X_{Max})*

(III) *Mean (X_{Mean})*

(IV) *Median (X_{Me})*

(V) *Mode (X_{Mo})*

(VI) *First quartile (X_{Q1})*

(VII) *Third quartile (X_{Q3})*

(VIII) *Inter-quartile range (X_{IQR})*

(IX) *Standard deviation (X_{SD}).*

The feature set is denoted as $\{X_{Min,}\, X_{Max,}\, X_{Mean,}\, X_{Me,}\, X_{Mo,}\, X_{Q1,}\, X_{Q3,}\, X_{IQR,}\, X_{SD}\}$. The same feature set is also used in the previous chapter (Chapter 3) for the SRS-LS-SVM algorithm. The reasons of choosing these statistical features are explained in Section 3.3.2 of Chapter 3. The obtained features are employed as the input for the LS-SVM for the EEG signal classification.

4.3 Experimental Data

In this Chapter, three databases are used to assess the method. They are obtained from three different sources, one for the EEG epileptic data (EEG time series, 2005) and one for the mental imagery tasks EEG data (Chiappa and J.R. Millán, 2005) and one for the motor imagery EEG data (BCI Competition III, 2005). In Chapter 3, EEG epileptic data and the mental imagery tasks EEG data are already discussed in Sections 3.4.1 and 3.4.2, respectively. A briefly description of the motor imagery EEG data is provided as below.

Dataset IVa from BCI Competition III (BCI Competition III, 2005; Blankertz et al. 2006) named motor imagery EEG data was recorded from five healthy subjects (labelled **aa**, **al**, **av**, **aw**, **ay**) who performed right hand motor imagery (class 1) denoted by 'RH' and right foot motor imagery (class 2) denoted by 'RF'. The subjects sat in comfortable chairs with their arms resting on armrests. This data set

contains MI EEG data from the four initial sessions without feedback. The EEG signals were recorded from 118 electrodes according to the international 10-20 system. There were 280 trials for each subject, namely 140 trials for each task per subject. During each trial, the subject was required to perform either of the two (right hand and right foot) MI tasks for 3.5 seconds. A training set and a testing set consisted of different sizes for each subject. Among 280 trials, 168, 224, 84, 56 and 28 trials composed the training set for subject **aa**, **al**, **av**, **aw**, **ay**, respectively, and the remaining trials composed the test set. This study uses the down-sampled data at 100 Hz where the original sampling rate is 1000 Hz.

4.4 Implementation of the proposed CT-LS-SVM algorithm

In this chapter, the proposed method is implemented with the EEG epileptic data, the mental imagery tasks EEG data and the motor imagery EEG data, separately. As discussed in Section 3.4.1 of Chapter 3, the epileptic EEG data has five sets that are Set A to Set E, and each set contains 100 channel data. Every channel consists of 4096 data points with 23.6 seconds. In this method, channel data of each dataset is divided into 16 groups where each group is called cluster and each cluster consists of 256 data points in 1.475 seconds. Then every cluster is again partitioned into 4 sub-clusters and each sub-cluster contains 64 observations of 0.3688 seconds.

The nine statistical features, i.e. $\{X_{Min}, X_{Max}, X_{Mean}, X_{Me}, X_{Mo}, X_{Q1}, X_{Q3}, X_{IQR}, X_{SD}\}$, are calculated from each sub-cluster. Thus we obtain a feature vector set of size [1×36] from 4 sub-clusters of each cluster. Hence, for 16 clusters of an EEG channel data file, we acquire a feature vector set of size [16×36] and consequently we obtain a feature vector set of size [1600×36] for 100 channel data files of a data set. Therefore we obtain a feature vector set of size [3200×36] for two data sets. The feature vector set make an input matrix of size [3200×36] (e.g. input matrix, x=[3200 ×36]) from any two-class signals and are used for the LS-SVM algorithm in equation (3.8). The LS-SVM is trained and tested with these features to classify EEG signals.

In the mental imagery tasks EEG data, each of three normal subjects was performed with three tasks considered as class 1 (left hand movements), class 2 (right hand movements) and class 3 (word generation) (discussed in Section 4.3).

Table 4.1: The number of recorded values in four sessions from the mental imagery tasks EEG data.

Subject	Session 1	Session 2	Session 3	Session 4
1	3488	3472	3568	3504
2	3572	3456	3472	3472
3	3424	3424	3440	3788

The number of recorded samples in four sessions for each subject is given in Table 4.1. Each of the eight channels has 12 frequency components and an EEG sample segment obtained from the eight channels consists of a 96 dimensional vector. In this study, we use the first three sessions' EEG recorded data for each subject. Thus 10528 vectors of 96 dimensions are obtained for Subject 1, 10400 vectors for each of Subject 2 and Subject 3 with the same dimensions as the original data in the experiments. In this application, every channel data from a subject are divided into 16 clusters and each cluster is again partitioned into 4 sub-clusters. In Subject 1, the cluster sizes and sub-cluster sizes of class 1, class 2 and class 3 are 183, 216, 259, and 45, 53, 64, respectively. The cluster sizes of class 1, class 2 and class 3 of Subject 2 are 185, 213, and 252, respectively. The sub-cluster sizes of class 1, class 2 and class 3 of this subject are 46, 53 and 63. For Subject 3, the sizes of the clusters for class 1, class 2 and class 3 are 213, 215 and 215, and the sub-cluster size is 53 for each of the three classes. It is important to note that the number of the observations and the EEG recorded time are not equal for each class.

Table 4.2: The number of clusters and sub-clusters and time period for each cluster and sub-cluster of a class for a subject for the mental imagery tasks EEG data.

Subject	Class	Number of cluster	Number of sub-cluster	Number of channels	Time period for each cluster (seconds)	Time period for each sub-cluster (seconds)
1	class 1	16	4	8	7.5000	1.8750
	class 2	16	4	8	8.4375	2.1094
	class 3	16	4	8	10.3125	2.5781
2	class 1	16	4	8	5.6250	1.4063
	class 2	16	4	8	6.5625	1.6406
	class 3	16	4	8	5.6250	1.4063
3	class 1	16	4	8	9.3750	2.3438
	class 2	16	4	8	9.3750	2.3438
	class 3	16	4	8	10.3125	2.5781

The number of the clusters and sub-clusters and time period for each cluster and sub-cluster of a class for a subject are summarized in Table 4.2. Then the nine statistical features are extracted from each sub-cluster and these features are used as the input to the LS-SVM for classification.

As mentioned in Section 4.3, the motor imagery EEG data used in this study consisted of two tasks denoted as two classes: right hand (denoted by 'RH') and right foot (denoted by 'RF') motor imageries. Each of five healthy subjects performed these two tasks in every trial. The size of each class is different for each subject. Each EEG channel data is divided into 16 clusters with a specific time period and then each cluster is partitioned into 4 sub-clusters. Table 4.3 presents the number of the clusters and sub-clusters and a distributed time period for each cluster and sub-cluster of a class as the EEG recording time for each class is not equal for all subjects. For Subject 1, the cluster and sub-cluster sizes are 5086 and 1271, respectively, for RH, and 6863 and 1715 for its RF. In Subject 2, the number of samples for clusters and sub-clusters are 6569 and 1642 for RH, and 7775 and 1943 for RF. For Subject 3, cluster sizes for RH and RF are 2840 and 2556 and sub-cluster sizes are 710 and 639, respectively. In Subject 4, the cluster and sub-cluster sizes of RH and RF are 2095, 523 and 1549, 387, respectively. For Subject 5, the sizes of clusters and sub-clusters are 904 and 226 for RH and 936 and 234 for RF. From each sub-cluster, the same statistical features that were used for the other two sets of data are calculated. These features vector sets are divided into two groups as the training and testing vector sets which are used as the input of the LS-SVM algorithm. A detail description of the LS-SVM is given in Section 3.3.3 of Chapter 3.

Table 4.3: The number of clusters and sub-clusters and time period for each cluster and sub-cluster of a class for a subject for the motor imagery EEG data.

Subject	Class	Number of cluster	Number of sub-cluster	Number of channels	Time period for each cluster (seconds)	Time period for each sub-cluster (seconds)
1	RH	16	4	118	5.0868	1.2717
	RF	16	4	118	6.8639	1.7160
2	RH	16	4	118	6.5699	1.6425
	RF	16	4	118	7.7754	1.9439
3	RH	16	4	118	2.8409	0.7102
	RF	16	4	118	2.5567	0.6392
4	RH	16	4	118	2.0957	0.5239
	RF	16	4	118	1.5493	0.3873
5	RH	16	4	118	0.9046	0.2261
	RF	16	4	118	0.9365	0.2341

In this study, the stability of performance of the LS-SVM classifier is assessed based on different statistical measurements, such as sensitivity, specificity and classification accuracy. A brief description of these measurements is provided in Section 3.3.4 of Chapter 3. In the experiments, we utilize three databases into two ways. Firstly we divide feature vector sets into two groups as the training and testing sets. For each of the three databases, the training vectors are applied to train the LS-SVM classifier, where the testing vectors are used to verify the accuracy and the effectiveness of the trained LS-SVM for the classification of two-class of EEG signals. Sensitivity, specificity and classification accuracy of the proposed method are therefore calculated from the testing set. Secondly, the 10-fold cross validation method (Abdulkadir, 2009) is used to evaluate the accuracy of the classification method. With the 10-fold cross validation method, the whole feature vector set is divided into 10 mutually exclusive subsets of a equal size and the present method is repeated 10 times. Each time, one of the 10 subsets is used as a test set and other 9 subsets are put together to form a training set. After 10 times repeating of the method, the obtained accuracy of each trial, are averaged. This is named as 10-fold cross validation accuracy. The performance is evaluated on the testing set for all the datasets.

4.5 Experimental Results and Discussions

In this study, we investigate the potentials of applying the CT algorithm for obtaining representative features from all EEG channel data and these features are used as inputs to the LS-SVM algorithm. The RBF kernel function is employed for the LS-SVM as an optimal kernel function over different kernel functions that were tested. The LS-SVM has two important parameters γ and σ^2, which should be appropriately chosen for achieving the desired performance. In order to obtain the best results, the LS-SVM is trained with different combinations of the parameters γ and σ^2. The proposed method is conducted on different pairs of two-class EEG signals with the epileptic EEG data, the mental imagery data tasks EEG data, and the motor imagery EEG data.

The optimal classification results are obtained for Case I, Case II, Case III and Case IV (described in Section 4.5.2) when $\gamma = 10$ and $\sigma^2 = 4$ in the epileptic

EEG data. Case V achieves the best result when $\gamma = 1$ and $\sigma^2 = 10$ for this database. In the mental imagery tasks EEG data (see Section 4.5.1), the optimal classification results are obtained for each pair of Subjects 1 and 2 when $\gamma = 1000000$ and $\sigma^2 = 10000$. For Subject 3, the most favorable results are obtained for $\gamma = 1000$ and $\sigma^2 = 1000.$ In the motor imagery EEG data (see Section 4.5.3), the best possible classification results are achieved for Subjects 1, 2, and 3 when $\gamma = 70$ and $\sigma^2 = 5$. We obtained the optimal results for Subject 4 when $\gamma = 10$ and $\sigma^2 = 10$, and for Subject 5 when $\gamma = 1000$ and $\sigma^2 = 100$. All experiments are performed using the MATLAB software package version 7.7 (R2008b) and run on a 1.86 GHz Intel(R) Core(TM)2 CPU processor machine with 1.99 GB of RAM. The operating system on the machine was Microsoft Windows XP Professional Version 2002. The classification results for the three datasets are presented in the following sections.

4.5.1 Classification results for the mental imagery tasks EEG data

With the mental imagery tasks EEG data, 1536 vectors of 36 dimensions for each class of a subject are obtained using the CT approach from the original data where 1000 vectors of 36 dimensions are utilized as the training set and 536 vectors of the same dimensions for the testing set from each class. For this dataset, nine experiments are carried out using different pairs of two-class EEG data. Each experiment uses 3072 vectors of 36 dimensions for two-class data with 1536 vectors from each class. The LS-SVM classifier is trained with the training set and performances are assessed with the testing set for different pairs of the two-class data. In this chapter, each experiment is considered as a case. Cases 1 to 3 are composed for Subject 1, Cases 4 to 6 are created for Subject 2 and Cases 7 to 9 are compiled for Subject 3, which are given below:

Case 1: class 1 versus class 2

Case 2: class 1 versus class 3

Case 3: class 2 versus class 3

Case 4: class 1 versus class 2

Case 5: class 1 versus class 3

Case 6: class 2 versus class 3

Case 7: class 1 versus class 2

Case 8: class 1 versus class 3

Case 9: class 2 versus class 3

Table 4.4 presents the performance comparison of the proposed CT-LS-SVM method versus the SRS-LS-SVM method for different pairs of two-class EEG signals from the mental imagery tasks EEG data. For most of the cases, the proposed approach achieves a higher classification accuracy, compared to the SRS-LS-SVM method. The average classification accuracy is calculated using all accuracy values for all cases. The average classification accuracy of the proposed CT-LS-SVM method is 61.69% while the SRS-LS-SVM technique obtained a 60.15% average classification accuracy in the mental imagery tasks EEG data. For each subject, the sensitivity of almost all cases increases more in the proposed method compared to the SRS-LS-SVM. The average sensitivity is 64.61% for the CT-LS-SVM method while the SRS-LS-SVM has 58.17%. The average specificity of the proposed method is 58.77% while the SRS-LS-SVM has 62.14% for the same dataset. Considering the results shown in Table 4.4, one can observe that the proposed method is more capable to classify the two-class EEG signals than the SRS-LS-SVM.

It is observed that the experimental results of the SRS-LS-SVM method for the same dataset are not exactly same as Chapter 3 because the SRS-LS-SVM algorithm is implemented with different data setting in this chapter. In Chapter 3, the SRS-S-SVM algorithm was applied on a training set to train the algorithm and the same training set is also used to evaluate performance of the method.

Table 4.4: Performance comparison of the proposed CT-LS-SVM versus the SRS LS-SVM for different pairs of two-class EEG signals from the mental imagery tasks EEG data.

Subject	Different cases	CT-LS-SVM (proposed)			SRS-LS-SVM		
		Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
1	Case 1	73.13	59.33	66.23	61.61	67.26	64.43
	Case 2	80.60	71.27	75.93	69.35	66.67	68.01
	Case 3	62.31	62.50	62.41	50.30	58.93	54.61
2	Case 4	58.21	59.70	58.96	60.71	70.54	65.62
	Case 5	74.63	72.39	73.51	47.02	56.55	51.79
	Case 6	62.69	61.01	61.85	61.90	44.05	52.98
3	Case 7	60.82	41.79	51.31	35.42	53.57	44.49
	Case 8	56.53	48.32	52.43	71.43	71.73	71.58
	Case 9	52.61	52.61	52.61	65.77	69.94	67.86
Average		64.61	58.77	61.69	58.17	62.14	60.15

In this chapter, the SRS-LS-SVM algorithm is implemented on separated training set and testing set where training set is used to train the method and then testing set is used to evaluate effectiveness of the method. The sizes of training and testing set are not same in Chapter 3 and Chapter 4. That's why, the performance of the SRS-LS-SVM method is different for the same dataset in Chapter 4 compared to Chapter 3.

In order to test further the efficacy of the present method, we also employ the 10-fold cross validation method for the different cases of all subjects of the mental imagery tasks data. The classification accuracy rate of the proposed method by the 10-fold cross validation method is presented in Table 4.5. It is seen from Table 4.5 that most of the experiments produce good performance and the average correct classification rate is 61.14% for all the subjects.

Table 4.5: Classification accuracy of the proposed CT-LS-SVM method by the 10-fold cross validation for the mental imagery tasks EEG data.

Subject	Different cases	10- fold cross validation accuracy of the proposed method (%)
1	Case 1	65.88
	Case 2	75.35
	Case 3	62.68
2	Case 4	58.95
	Case 5	73.04
	Case 6	62.35
3	Case 7	47.84
	Case 8	51.47
	Case 9	52.71
Average for all cases of all subjects		61.14

We also observe from Table 4.4 and Table 4.5 that the accuracies of the proposed method for some cases are not significantly high as noises and artifacts were not removed from this dataset.

To accurately evaluate the computational efficiency of the proposed method, the execution (running) time of the CT-LS-SVM algorithm is compared in Figure 4.3 with the SRS-LS-SVM method. Figure 4.3, which is generated using Matlab (version 7.7, R2008b) reports the execution time behaviours of the both algorithms. The total numbers of sample points of the two-class algorithm are plotted in X-axis and total program running time (second) is put in Y-axis. For the two methods, the running time increases in conjunction with an increase in the sample numbers. For example, it is seen from Figure 4.3 that the program running time is 3.2 seconds for the

proposed method when the algorithm uses 76608 sample points. On the other hand the SRS-LS-SVM method takes 5.8 seconds for the same samples. Again the proposed CT-LS-SVM method takes 4.8 seconds for the total observations of 153216 where the SRS-LS-SVM takes 9.6 seconds for the same observations.

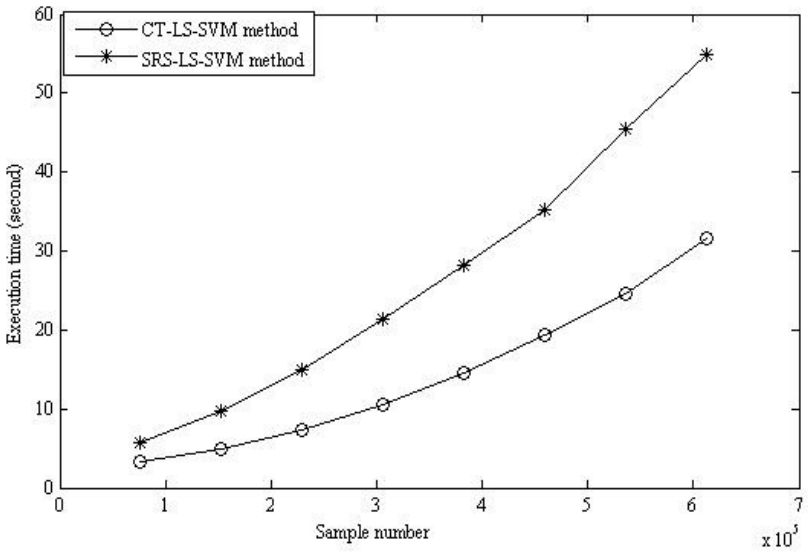


Figure 4.3: Comparison of execution time between the CT-LS-SVM and SRS-LS-SVM methods for the mental imagery tasks EEG data.

Figure 4.3 clearly shows that the proposed technique takes less time than the SRS-LS-SVM method for all sample numbers. The results demonstrate that the computational complexity of the CT method based on the LS-SVM is much less than the SRS method with the LS-SVM to classify any pair of EEG signals.

Table 4.6: Performance comparison of different methods with the proposed CT-LS-SVM algorithm for the mental imagery tasks EEG data.

Method	Overall Classification accuracy (%)			
	S1	S2	S3	Average
CT- LS-SVM (proposed method)	68.19	64.77	52.12	61.69
Neural networks based on improved particle swarm optimization (IPSO) . (Lin and Hsieh, 2009)	78.31	70.27	56.46	68.35
Random electrode selection ensemble (RESE) (Sun et al. 2008)	68.75	56.41	44.82	56.66
Ensemble methods (Sun et al. 2007)	70.59	48.85	40.92	53.45
Adaptive common spatial patterns (ACSP) (Sun and Zhang, 2006)	67.70	68.10	59.55	65.12
Decorrelated least mean square (DLMS) (Sun and Zhang, 2005)	69.23	48.97	45.80	54.67

Note: S1=Subject 1; S2= Subject 2; S3= Subject 3.

The performance comparison of the proposed method with other methods reported in the literature for the mental imagery tasks EEG data is presented in Table 4.6. Table 4.6 gives the classification accuracies of the proposed algorithm and some existing methods. The average classification accuracy of IPSO (Lin and Hsieh, 2009), RESE (Sun et al. 2008), Ensemble methods (Sun et al. 2007), ACSP (Sun and Zhang, 2006) and DLMS (Sun and Zhang, 2005) methods were obtained 68.35%, 56.66%, 53.45%, 65.12% and 54.67%, respectively, for the mental imagery tasks data whereas this value is 61.69% for our present method. From these results, one can see that our proposed approach has the capability to classify two-class EEG signals with the computational efficiency.

4.5.2 Classification results for the EEG epileptic data

The present method is employed in this section to classify different pairs of two-class EEG signals from five datasets (Sets A-E) in the EEG epileptic data. In this study, 1600 vectors of 36 dimensions (the dimensions of the extracted features) from each dataset are obtained using the CT method. We use the first 1100 vectors of 36 dimensions for the training and the remaining 500 vectors of the same dimensions for the testing of each class. The training vectors are used to train the LS-SVM classifier, while the testing vectors are used to verify the accuracy and the effectiveness of the trained LS-SVM for the classification of two-class of EEG signals.

Five experiments are performed for the EEG epileptic dataset and each experiment is considered as a case. In each case, we use a pair of two-class of EEG signals, which have 3200 vectors with 36 dimensions taking 1600 vectors from each class. The cases are defined as follows:

Case I: Set A versus Set E

Case II: Set B versus Set E

Case III: Set C versus Set E

Case IV: Set D versus Set E

Case V: Set A versus Set D

Table 4.7 displays the performance comparison of the proposed CT-LS-SVM method versus the SRS-LS-SVM method for the different cases in the EEG epileptic data. The classification accuracies of Case I, Case II, Case III, Case IV and Case V are 99.90%, 96.30%, 96.20%, 93.60% and 84.90%, respectively. The average

classification accuracy of the proposed approach for all cases is achieved as 94.18% while the SRS-LS-SVM gained 95.58% for the same dataset. An average value of sensitivity is obtained as 94.92% for the proposed approach while the SRS-LSVM method attained sensitivity of 95.96% as the average for this dataset. The CT-LS-SVM method achieved a 93.44% average specificity but the SRS-LS-SVM technique obtained a 95.20% average specificity. From Table 4.7, it is noticeable that the SRS-LS-SVM attained slightly higher classification accuracies (sensitivity, specificity, accuracy) compared to the present method, as in the SRS, random samples were replicated 10 times from each channel.

Table 4.7: Performance comparison of the proposed CT-LS-SVM versus the SRS-LS-SVM method for different pairs of two-class EEG signals from the EEG epileptic data.

Different Cases	CT-LS-SVM (proposed)			SRS-LS-SVM		
	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Case I	100.00	99.80	99.90	100.00	100.00	100.00
Case II	99.20	93.40	96.30	99.80	99.20	99.50
Case III	96.20	96.20	96.20	98.00	94.80	96.40
Case IV	89.40	97.80	93.60	94.00	82.00	94.00
Case V	89.80	80.00	84.90	88.00	100.00	94.00
Mean/Average	94.92	93.44	94.18	95.96	95.20	95.58

From Table 4.7, it is noted that the highest sensitivity, specificity and accuracy are obtained for both methods in Case I and the lowest for both methods in Case V. It is important to note that the EEG data from Case I are more classifiable than the other cases because there are large variations among the recorded EEG values in Case I, which consists of Set A and Set E. Due to the nature of the large differences in the data, it is easier to classify Set A and Set E as demonstrated by the 99.90% of classification accuracy in the proposed method. In contrast, Case V produces the lowest classification accuracy which is 84.90% for Sets A and D. The accuracy is not significant as the EEG data from Set A and Set D are more analogous to each other.

Table 4.8 shows the performance of the proposed method through the 10-fold cross validation method for every case from the epileptic data. Using the 10-fold cross validation method, we can achieve an overall classification performance of 94.12%. It is observed from Table 4.8 that the classification accuracies for almost all cases are quite satisfactory, which indicate the high performance of the proposed method for EEG signal classification.

Table 4.8: Classification accuracy of the proposed CT-LS-SVM method by the 10-fold cross validation for the EEG epileptic data.

Different Cases	10-fold cross validation accuracy of the proposed method (%)
Case I	99.69
Case II	96.78
Case III	97.69
Case IV	93.91
Case V	82.53
Average	94.12

We next compare the execution time of the proposed method to the SRS-LS-SVM method for this dataset. Figure 4.4 depicts the execution time (running time) of the CT-LS-SVM versus the SRS-LS-SVM for the epileptic data. It is generated using MATLAB (version7.7, R2008b). The total numbers of the observations of two-class raw EEG data are indicated in the horizontal axis and the total program running time in second is plotted in the vertical axis. The proposed method takes 8.9 seconds and 9.4 seconds for 81940 and 163880 data samples, respectively, but the SRS-LS-SVM method takes 11.4 seconds and 14.2 seconds for the same samples, respectively. Figure 4.4 illustrates that the execution time of the proposed algorithm with all different samples is much smaller than the SRS-LS-SVM algorithm. The SRS technique with the LS-SVM takes longer time as it spends more time selecting different random samples from the original data set.

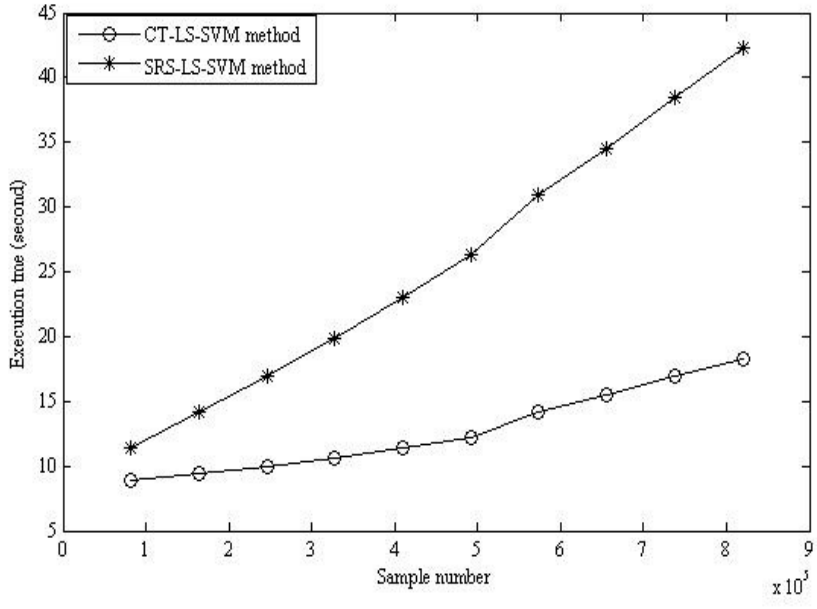


Figure 4.4: Comparison of execution time between the CT-LS-SVM and SRS-LS-SVM methods for the EEG epileptic data.

Thus it is noted that the proposed method is superior to the SRS method with the LS-SVM in terms of the execution time. As Figure 4.3 and Figure 4.4 show, the proposed algorithm results in a less execution time, which demonstrates the flexibility and the usability of the proposed method.

Different methods from the literature and their respective classification accuracies for healthy subjects with eyes open (Set A) and epileptic patients during seizure activity (Set E) from the epileptic dataset are provided in Table 4.9 for the performance comparison. Based on these results shown in Table 4.9, our proposed method produces a good classification accuracy rate (99.90%) while the SRS-LS-SVM method (Siuly et al. 2011d) reported 100% classification accuracy. The classification accuracy of the wavelet-artificial neural networks (Guo et al. 2009), wavelet-neural networks (Jahankhani et al. 2009) and expert model with a double-loop EM algorithm (Subasi, 2007) were reported at 95.00%, 98.00% and 94.50%, respectively.

Table 4.9: The obtained performance with the proposed CT-LS-SVM method and other methods from the literature for healthy subjects with eyes open (Set A) and epileptic patients during seizure activity (Set E) of the EEG epileptic data.

Method	Classification accuracy
CT-LS-SVM (proposed method)	99.90%
SRS based LS-SVM (SRS-LS-SVM) (Siuly et al. 2010)	100.00%
Wavelet-artificial neural networks (Guo et al. 2009)	95.00%
Wavelet-neural networks (Jahankhani et al. 2009)	98.00%
Expert model with a double-loop EM algorithm (Subasi, 2007)	94.50%
Decision tree classifier-FFT (Polat and Gunes, 2007)	98.72%
Cross-correlation aided SVM classifier (Chandaka et al., 2009)	95.96%
Model based methods-LS-SVM (Ubeyli, 2010)	99.56%

On the other hand, the decision tree classifier-FFT (Polat and Gunes, 2007), cross-correlation aided SVM classifier (Chandaka et al., 2009) and model based methods-LS-SVM (Ubeyli, 2010) obtained 98.72%, 95.96% and 99.56% classification rates. Based on the above results, we conclude that the CT method with the LS-SVM obtains more promising results in classifying the two-class EEG signals. We believe that the proposed approach can be very helpful to physicians for their final decisions on the diagnosis of their patients. By using such a reliable tool, they can make more accurate medical diagnosing decisions.

4.5.3 Classification results for the motor imagery EEG data

This section discusses the classification results of the proposed approach for the motor imagery EEG dataset. Applying the CT method to this data set, we obtain 1888 feature vectors of 36 dimensions for one class from each subject where 1000 vectors of 36 dimensions are used as the training vector and 888 vectors of the same dimensions for the testing set. We know that the motor imagery data contain the EEG recorded data of five healthy subjects where every subject performed two tasks, imagination of 'right hand' and 'right foot' movement. Each task indicates a class of EEG data. For this dataset, five experiments are conducted for five subjects and every experiment contains 3776 feature vectors of 36 dimensions for two classes of a subject, with 1888 vectors of the same dimension in each class.

Table 4.10 shows the sensitivity, specificity and classification accuracy of the proposed method compared to the SRS-LS-SVM algorithm for the motor imagery EEG data. As shown in Table 4.10, the classification accuracy of the proposed method for almost all subjects is higher than the previous method, SRS-LS-SVM.

Table 4.10: Performance comparison between the CT-LS-SVM and SRS-LS-SVM for the motor imagery EEG data.

Subject	CT-LS-SVM			SRS-LS-SVM		
	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
1 (aa)	96.17	88.18	92.17	100.00	55.29	77.65
2 (al)	89.86	73.42	81.64	88.85	38.40	63.63
3 (av)	87.16	88.96	88.06	100.00	64.75	82.38
4 (aw)	63.74	88.63	76.18	97.29	58.22	77.76
5 (ay)	82.99	82.66	82.83	100.00	64.30	82.15
Mean/Average	83.98	84.37	84.17	97.23	56.19	76.71

The average sensitivity and specificity for the CT-LS-SVM are 83.98% and 84.37%, while they are 97.23% and 56.19%, respectively, for the SRS-LS-SVM. The proposed approach produces a 84.17% average classification accuracy for all five subjects while the SRS-LS-SVM has 76.71%. The results demonstrate that the proposed algorithm has better potential in the classification environment than the SRS-LS-SVM.

The 10-fold cross validation accuracy rate of the proposed approach for the motor imagery data is depicted in Table 4.11. The classification accuracy is reached at 84.17% for the same dataset when the data are divided into two groups as the

training and testing sets that results are shown in Table 4.10. We achieve an overall performance of 88.32% for the same dataset with the same parameters of the classifier. It is worth mentioning that the proposed method is very effective in identifying the different motor imagery signals from EEG data.

Table 4.11: Classification accuracy of the CT-LS-SVM method by the 10-fold cross validation for the motor imagery EEG data.

Subject	10-fold cross validation accuracy of the proposed method (%)
1 (aa)	92.63
2 (al)	84.99
3 (av)	90.77
4 (aw)	86.50
5 (ay)	86.73
Mean/Average	88.32

Figure 4.5 presents the experimental time (execution time in second) of the CT-LS-SVM and SRS-LS-SVM for different numbers of samples with the motor imagery EEG data. The total number of samples for the two-class EEG data is presented in X axis, and the total program running time (execution time) in second is with Y-axis.

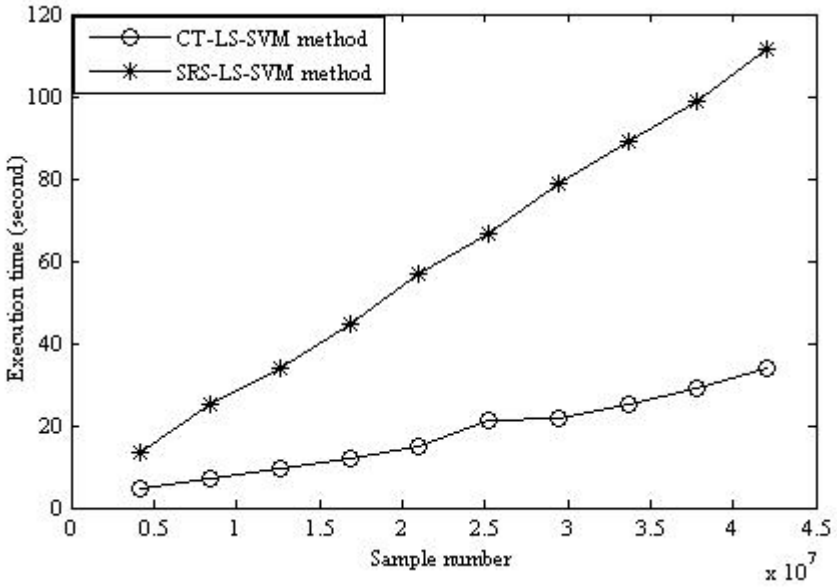


Figure 4.5: Comparison of execution time between the CT-LS-SVM and SRS-LS-SVM methods for the motor imagery EEG data.

Figure 4.5 shows that the proposed method and the SRS technique with LS-SVM take 4.7 seconds and 13.3 seconds, respectively, when the total number of samples

for two-class signals is 4206664. Again the program running time is 7.3 seconds for the proposed CT-LS-SVM algorithm when the algorithm uses 8413328 samples whereas the SRS-LS-SVM method takes 25.1 seconds for the same samples. Thus Figure 4.5 depicts that the SRS-LS-SVM method takes longer time than the proposed technique for all samples. Figures 4.3 and 4.4 show the same patterns for the mental imagery tasks data and the epileptic data, respectively. As Figure 4.5 indicates, the proposed approach is a faster running algorithm compared to the SRS-LS-SVM.

Table 4.12 displays an overall comparison of our method with a few other EEG signal classification methods for the motor imagery EEG data set. The results are presented with respect of the classification accuracy for the five subjects and their averages. The average classification accuracies of sparse spatial filter optimization (Yong et al. 2008), R-CSP (Lu et al. 2009), composite CSP (Kang et al., 2009) and SRCSP methods (Lotte et al. 2010) for the motor imagery data are 73.50%, 74.20%, 76.22% and 78.62%, respectively, whereas it is 84.17% for the proposed method. The classification accuracy is improved when the proposed methodology is employed on the motor imagery EEG data.

Table 4.13 shows a summary of the performance of the proposed CT-LS-SVM method versus the SRS-LS-SVM for the three databases. It is observed from Table 4.13 that the CT approach achieves 94.18 % of the average classification accuracy with the epileptic EEG data while the SRS-LS-SVM technique obtained a 95.58% average classification accuracy.

Table 4.12: Comparison of classification accuracy for the motor imagery EEG data with other EEG signal classification attempts.

Method	Classification accuracy (%)					
	S1	S2	S3	S4	S5	Average
CT-LS-SVM (proposed method)	92.17	81.64	88.06	76.18	82.83	84.17
Spatially regularized common spatial pattern (SRCSP) (Lotte et al. 2010)	72.32	96.43	60.2	77.68	86.51	78.62
Composite common spatial pattern (composite CSP) (method 1; *n=3)* (Kang et al., 2009)	67.66	97.22	65.48	78.18	72.57	76.22
Regulized common spatial pattern with generic learning (R-CSP) (Lu et al. 2009)	69.6	83.9	64.3	70.5	82.5	74.20
Sparse spatial filter optimization (Yong et al. 2008)	57.5	54.4	86.9	84.4	84.3	73.50

Note: S1=Subject 1 (aa); S2= Subject 2 (al); S3= Subject 3 (av); S4= Subject 4 (aw); S5= Subject 5 (ay).

The average classification accuracy of the SRS-LS-SVM method is a little higher than the proposed method because random samples were repeated 10 times from

each EEG channel data in the SRS technique. For the motor imagery data, we obtain an average classification accuracy of 84.17% for the present method whereas it is 76.71% for the SRS-LS-SVM method.

Table 4.13: Summary results of the proposed CT-LS-SVM approach and the SRS-LS-SVM method for the mental imagery tasks EEG data, the EEG epileptic data and the motor imagery EEG data.

Database	Average classification accuracy (%)	
	CT-LS-SVM	SRS-LS-SVM
EEG epileptic data	94.18	95.58
Motor imagery EEG data	84.17	76.71
Mental imagery tasks EEG data	61.69	60.15

On the other hand, the average classification accuracy is improved at 61.69% compared to 60.15% when the proposed method is adapted to the mental imagery tasks data. The study demonstrates that the obtained signal features using the CT approach accurately represent the most important information in the recorded EEG data. The CT-LS-SVM is a powerful and less complex algorithm for the EEG signal classification.

4.6 Conclusions

This chapter proposes the CT-LS-SVM algorithm for the EEG signal classification where the CT approach is employed for the feature extraction and the LS-SVM classifier with RBF kernel function is used for the classification of the extracted features. The major aim of the proposed approach is to develop a system that can distinguish two categories of EEG signals and to investigate whether the CT method is appropriate for feature extraction from EEG data. Experiments are carried out over three publicly available benchmark databases, an EEG epileptic data, a mental imagery tasks EEG data for brain-computer interface (BCI) Competition III (Data set V), and a motor imagery EEG data obtained from BCI Competition III (Data set IVa), respectively. The efficacy and superiority of the proposed CT-LS-SVM method over the SRS-LS-SVM method are validated through different measures. For the EEG epileptic data, we obtain 94.92%, 93.44% and 94.18% as the average sensitivity, specificity and classification accuracy, respectively, using the CT-LS-SVM while the SRS-LS-SVM has 95.96%, 95.20% and 95.58%, respectively. The

average classification accuracy of the proposed method is 1.40% lower compared to the SRS-LS-SVM as in the SRS, random samples are repeated 10 times from each EEG channel data. In the motor imagery data, the average classification accuracy of the proposed approach is 84.17%, while the SRS-LS-SVM has 76.71%. The proposed method attains an average sensitivity, specificity and classification accuracy of 64.61%, 58.77% and 61.69%, respectively, for different pairs of two-class EEG signals from the mental imagery tasks EEG data while the average sensitivity, specificity and classification accuracy of the SRS-LS-SVM were reported at 58.18%, 62.14% and 60.15% for the same dataset. The average classification rate achieved by the CT-LS-SVM presented for EEG signal classification is found higher than that in the SRS- LS-SVM for the mental imagery tasks data. Using the 10-fold cross validation method, we achieve an overall classification accuracy of 94.12% for the epileptic data, 88.32% for the motor imagery data and 61.14% for the mental imagery data by using the CT-LS-SVM algorithm. In the proposed CT-LS-SVM approach, it takes much less time to compute the data compared to the SRS-LS-SVM with the three datasets. One of the advantages of the proposed methodology is the computational efficiency and usability. The experimental results demonstrate that the CT method is efficient for extracting features to represent the EEG signals and the LS-SVM classifier has the inherent ability to solve a pattern recognition task for these features.

CHAPTER 5
CROSS-CORRELATION BASED LOGISTIC REGRESSION METHOD FOR CLASSIFICATION OF MI TASKS

Classification of MI tasks based EEG signals provides an important basis for designing BCI systems. If MI tasks are reliably distinguished by recognizing patterns in EEGs, then motor disabled people could control a device by composing sequences of MI tasks. Thus, it is very crucial to classify EEG signals correlated with various physical or mental activities in BCI applications. This chapter focuses on the identification of MI tasks for the development of BCI technologies combining cross-correlation and logistic regression techniques named as CC-LR. The cross-correlation (CC) method is used for feature extraction and the logistic regression (LR) method is employed as a classifier to classify the extracted features. This study investigates the performance of the CC technique for a randomly selected reference signal and also investigates the LR on how efficient to classify the cross-correlated features. The CC-LR algorithm is tested on two benchmark datasets, IVa and IVb of BCI Competition III, and the performance is evaluated through a 3-fold cross-validation procedure. The experimental outcomes are compared with two recently reported algorithms, R-CSP with aggregation and the CT-LS-SVM and also other four algorithms using dataset IVa. The CC-LR algorithm can help to properly identify MI tasks, which can help to generate control signals for BCI systems.

The content of this chapter has been accepted by the *International Journal of Bioinformatics Research and Applications* (Siuly et al., 2011b).

5.1 Introduction

Motor imagery (MI) is a cognitive task that is one of the major protocols widely used in the field of BCI research. In MI tasks, subjects are instructed to imagine themselves performing a specific motor action (e.g. a hand or foot) without overt motor output (Kayikcioglu and Aydemir, 2010; Thomas et al., 2009) and each task is treated as a class. MI tasks modify bioelectrical brain activities generally measured

through EEGs. A BCI is able to detect such changes in the ongoing EEG signals and translates different brain states into operative control signals (Pfurtscheller et al., 1998). Therefore, BCI technologies can establish a direct communication channel between the human brain and a machine which does not require any motor activity (Vaughan et al., 2003; Wang et al., 2008). When people suffer from severe motor disabilities while being cognitively intact, they need such alternative methods to interact with their environments without any movements.

In the BCI development, MI tasks are key issues as users produce different brain activity patterns from different MI tasks that will be identified by a system and are then translated into commands (Wang et al., 2004). These commands will be used as feedback for motor disabled patients to communicate with the external environments. Thus, a MI-based BCI provides a promising control and communication means to people suffering from motor disabilities (Wolpaw et al. 2002). Therefore, the recognition of MI tasks is very crucial for the BCI development to generate control signals.

If a BCI system is considered as a pattern recognition system, the EEG recognition procedure mainly involves feature extraction and classification. Classification of human EEG signals is a difficult problem as EEG data are naturally non-stationary, highly noisy and contaminated with artifacts (Wu et al., 2008; Long et al., 2010). On the other hand, it is also very important to extract discriminative relevant features from EEG signals for a successful classification of brain activity-related tasks (e.g. MI tasks). The performance and reliability of a recognition system depend greatly on the features and the classification algorithm employed. The current study aims to build up an algorithm for the classification of MI tasks for the development of BCI systems.

Most recently, two algorithms are introduced for classifying MI-based EEG signals in BCI applications for dataset IVa in BCI Competition III. The first one was proposed by Lu et al. (2010) and the second one by Siuly et al. (2011c). Lu et al. (2010) used a regularization and aggregation technique with common spatial pattern (CSP) to separate MI tasks in a small-sample setting (SSS). A number of R-CSPs were aggregated to give an ensemble-based solution. The classification accuracy rates were obtained 76.8%, 98.2%, 74.5%, 92.9% and 77.0% for subject 1, subject 2, subject 3, subject 4, subject 5, respectively, for experiment III. The overall accuracy performance was 83.9%. Siuly et al. (2011c) reported a clustering technique-based

least square support vector machine algorithm (CT-LS-SVM) for the EEG signal classification. They developed a clustering technique (CT) for the feature extraction and the obtained features were used to the LS-SVM as the inputs for classification. The proposed algorithm achieved the classification accuracy of 92.63% for subject 1, 84.99% for subject 2, 90.77% for subject 3, 86.50% for subject 4 and 86.73% for subject 5. The average accuracy performance was 88.32%.

The major weakness of Lu's algorithm is that the method is only applicable for small sample settings. In Siuly's approach, the authors manually selected the parameters for the LS-SVM although the parameters of the LS-SVM play an important role in achieving a classification performance. The classification accuracy rates for each subject in these two algorithms are not good enough. This study proposes an algorithm that combines cross-correlation and logistic regression (CC-LR) techniques for the classification of MI tasks. The cross-correlation (CC) technique is used for the feature extraction from the MI data and the logistic regression (LR) is applied for the classification of the obtained features. The effectiveness of the proposed algorithm is tested on two benchmark datasets, IVa and IVb of BCI Competition III. These two datasets contain two categories MI EEG data.

This study uses six statistical characteristics, *mean, median, mode, standard deviation, maximum* and *minimum* to diminish the dimensionality of a cross-correlation sequence and these features represent the distribution of the MI EEG signals. The reasons of considering these features in this study are provided in Section 5.3.1. These values are employed to the LR model as input variables for the classification. A well known 3-fold cross-validation technique is applied to assess the performance of the proposed method as a means to control over-fitting of the data. The present algorithm is compared with several existing algorithms in the literature. The experimental results demonstrate that the proposed algorithm is superior to the existing algorithms for the MI tasks classification in BCIs.

The main contributions of this work are as follows.

1. The introduction of a CC-LR algorithm for the MI tasks classification, which is first reported in this study.
2. The proposed algorithm is suitable for any size (large or small) of a dataset.
3. The proposed algorithm is reliable because the parameters of the LR classifier are estimated automatically through maximum likelihood estimation (MLE) algorithm rather than by a manual selection.

4. The CC-LR algorithm can enhance the success rate in the MI tasks classification as the CC technique is powerful in signal processing for feature extraction and the LR model is very robust for the identification of the MI tasks.
5. The proposed CC-LR algorithm can be used to properly identify MI tasks, which can help to generate control signals for BCI systems.

5.2 Theoretical Background

5.2.1 Cross-correlation technique

In signal processing, cross-correlation (CC) (Chandaka et al. 2009a) is a statistical tool that measures the degree of similarity of two signals as a function of a time-lag applied to one of them. The similarity of two waveforms may be numerically evaluated by summing the products of identical time samples of each waveform. It is commonly used to search a long duration signal for a shorter, known feature, which has applications in pattern recognition when measuring information between two different time series. The correlation uses two signals to produce a third signal. This third signal is called the cross-correlation of the two input signals. If a signal is correlated with itself, the resulting signal is instead called the autocorrelation. This method basically motivates implementations of the Fourier transformation: signals of varying frequency and phase are correlated with the input signals, and the degree of a correlation in terms of frequency and phase representing the frequency and phase spectrums of the input signals.

The CC of two signals is obtained by multiplying corresponding ordinates and summing for all portions of the signals within a time window. Consider two signals, x and y, with N points, their cross-correlation as a function of lag m is defined as (Chandaka et al. 2009a; Dutta et al. 2010)

$$R_{xy}[m] = \sum_{i=0}^{N-|m|-1} x[i]y[i-m]; \quad m = -(N-1),-(N-2),.....0,1,2,3.....(N-2),(N-1) \tag{5.1}$$

Here, index m represents time-shift parameters known as lag where m=-(N-1),-(N-2),…..-1,0,1,..,(N-2),(N-1), and $R_{xy}(m)$ is the cross-correlated sequence at m lag. If each of the signals, x and y, consists of M finite number of samples, the resultant cross-correlation sequence has 2M-1 samples. If x and y are not the same length, for

example, x and y have N and M number of samples, respectively, and if $N>M$, the resultant cross-correlation sequence has $(2N-1)$ number of samples. The shorter vector, here y, is zero-padded to the length of the longer vector, x.

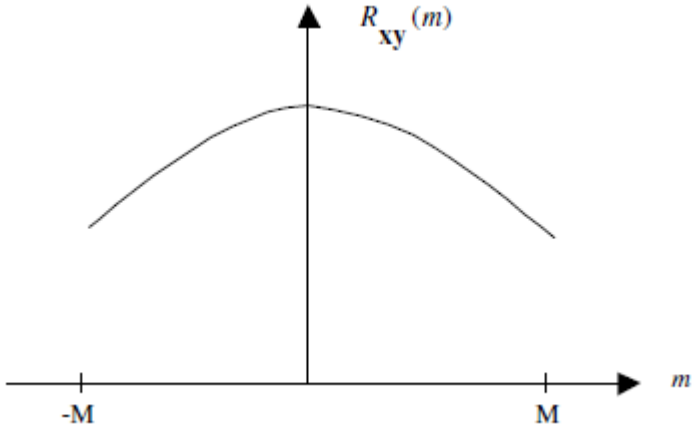


Figure 5.1: Example of a typical cross-correlogram (Chandaka et al. 2009a).

Figure 5.1 shows an example of a typical cross-correlogram. The peak of the cross-correlogram represents offset. If the two signals are identical (same rhythm), the peak of the cross-correlogram curve would appear exactly in the centre. If they are offset from each other, the peak will occur offset from center. In cross-correlation analysis, two signals (e.g. x and y) that alternate are out-of-phase from each other and will have a negative relationship, whereas two signals that are synchronous will be in-phase and have a positive relationship. A high degree of symmetry or stability along the X-axis indicates a stable relationship between the two signals. However, as the relationship between two signals varies, therefore creating decreasing correlation values beyond zero lag, this indicate less stability in the relationship.

5.2.2 Logistic regression model

Logistic regression (LR) model is a workhorse of statistics and is increasingly popular in machine learning, due to its similarity with the support vector machine (SVM). The LR fits a separating hyper plane that is a linear function of input features between two classes. The goal of the LR in Equation (5.2) is to estimate the hyper plane that accurately predicts class label of a new example. To accomplish this goal, a model is created that includes all independent or predictor variables that are useful

in predicting the dependent/response variables. The LR (binary) is used when the dependent variable is a dichotomy (which is usually presented by the occurrence or non-occurrence of some output events, usually coded as 0 and 1) and the independent variables are of any types.

Suppose $x_1, x_2,\ldots\ldots,x_n$ are vectors of input features and y is its class label either 0 or 1. Here $x_1, x_2,\ldots\ldots,x_n$ are treated as independent variables and y is a dependent variable. Under the logistic regression framework, the probability of the dependent variable y, when y belongs to class 1, is defined as (Caesarendra et al. 2010; Hosmer and Lemeshow, 1989; Subasi, 2005)

$$P(y=1|\ x_1, x_2,\ldots\ldots,x_n)=\pi = \frac{e^{\beta_0+\sum_{i=1}^{n}\beta_i x_i}}{1+e^{\beta_0+\sum_{i=1}^{n}\beta_i x_i}} \tag{5.2}$$

Here, π is a conditional probability of the form $P(y=1|\ x_1, x_2,\ldots\ldots,x_n)$. On the other hand, the probability of y, when y belongs to class 0 denoted as $P(y=0|\ x_1, x_2,\ldots\ldots,x_n)$ can be calculated as $1-\pi=1-P(y=1|\ x_1, x_2,\ldots\ldots,x_n)$. In Equation (5.2), β_0 is an intercept and $\beta_1, \beta_2, ..., \beta_n$ are the regression coefficient related to the independent variables $x_1, x_2, \ldots x_n$. These parameters are estimated by maximum likelihood estimation (MLE) (Hosmer and Lemeshow, 1989). The above cost function results in a solution that accurately predicts class label of a new example.

The *logit* model of the LR is given below as (Caesarendra et al. 2010; Hosmer and Lemeshow, 1989; Subasi, 2005)

$$\log it(\pi) = \log_e\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n = \beta_0 + \sum_{i=1}^{n}\beta_i x_i \tag{5.3}$$

In Equation (5.3), *logit* (π) is a linear combination of the independent variables, $x_1, x_2, \ldots x_n$ and regression coefficients, $\beta_0, \beta_1, \beta_2, ..., \beta_n$. The LR applies the MLE after transforming the dependent variable into *logit* variable in order to calculate the parameters $\beta_0, \beta_1, \beta_2, ..., \beta_n$. A linear relationship is not assumed in general between the independent and dependent variables nor requires normally distributed independent variables (Subasi, 2005).

5.3 Proposed Methodology

This section presents the CC-LR algorithm to classify the MI tasks in BCI applications. The CC-LR algorithm combines two techniques, cross-correlation (CC) and logistic regression (LR), where the CC technique is used for the feature extraction and the LR model is employed for the classification of the MI tasks described in the following two sections.

5.3.1 Feature extraction using cross-correlation (CC) technique

In the present study, we develop a CC technique for the feature extraction from MI tasks data. There are strong reasons for choosing the CC method for the feature extraction in this research. EEGs record brain activities as multichannel time series from multiple electrodes placed on the scalp of a subject. It is found that all the channels on the head do not provide independent information and there are highly correlations between channels in EEGs. The anatomical difference of the brain of the subjects could affect the correlation of signals. The recorded multi-channel EEG signals are also typically very noisy. They are not directly usable in BCI applications. The CC method can measure the degree of similarities between two signals with respect to time (Chandaka et al. 2009a). In addition, the cross correlation can diminish noise from the EEG signals by means of correlation calculation because of the characteristics of signal periodicity. Hence the cross-correlogram is a nearly noise-free signal that can provide more signal information compared to the original signal (Hieftje et al., 1973). The process also takes into consideration any potential phase differences between the two signals via the inclusion of a lead or lag term (Dutta et al., 2009). Thus, a cross-correlation technique works better for the feature extraction from the MI EEG data.

The CC method has been successfully used in many applications like ECG beat detection (Dutta et al. 2010; Last et al. 2004), gait signal processing (Dutta et al. 2009; Joshi and Anand 2010), emotional speech recognition (Chandaka et al. 2009b), heart rate variability classification (Abdullah et al. 2010), signal to noise enhancement (Hieftje et al. 1973), seizure prediction (Subasi and Ercelebi, 2005) and their promising results have been reported. According to our best knowledge, this technique is not yet used for the MI feature extraction in BCI development. These

facts give a possibility to use the CC technique for representative feature extraction from MI data in this work.

In this research, the CC technique follows three steps to extract features from MI tasks data. At first, one of the EEG channels is selected randomly as a reference channel (reference signal) from a class of a subject as there are no specific requirements for selecting a reference signal in the cross-correlation analysis. In equation (5.1), $x[i]$ is considered as the reference signal and $y[i]$ is one of any other signals named as a non-reference signal. Secondly, the reference channel of a class is cross-correlated with the data of the remaining channels in this class and the data of all channels of another class. Thirdly, six statistical features, *mean, median, mode, standard deviation, maximum* and *minimum* are calculated from each cross-correlation sequence to reduce the dimensions, which ideally represents the distribution of the signal containing important information.

It is necessary to describe why the above mentioned characteristics are used in this chapter for the representations of the MI data. When we are interested in describing an entire distribution of some observations or characteristics of individuals, there are two types of indices that are especially useful. These are the measure of central tendency and the measure of variability (Islam 2004; De Veaux et al., 2008). Measures of central tendency are numerical indices that attempt to represent the most typical value (centre value/representative value) of the observations. The purpose of a typical value is to represent the distribution and also to afford a basis of comparison with other distributions of a similar nature. The three measures of the central tendency, *mean*, *median* and *mode* are the most used typical values which can describe almost all distributions with a reasonable degree of accuracy. *Mean* corresponds to the centre of a set of values while *median* is the middle most observation. *Mode* is the value in the data set that occurs most often. These three features give a fairly good idea about the nature of the data (shows the "middle value"), especially when combined with measurements on how the data is distributed. Measures of variability describe how the observations in the distribution are different from each other or how the observations in the distribution spreading out around the typical values. *Standard deviation* is the most popular measure for the variability, which is the average distance between the actual data and the mean. This feature gives information about the spread of data how close the entire set of data is to the average value in the distribution. Data sets with a small *standard deviation*

have tightly grouped, precise data. Data sets with large *standard deviations* have data spread out over a wide range of values. Measures of central tendency and measures of variability are both used to describe the distribution of observations or characteristics of individuals under study. *Maximum and minimum* values are used to describe the range of observations in the distribution. Hence *mean, median, mode, standard deviation, maximum* and *minimum* values are considered as the most valuable parameters for representing the distribution of the MI EEG signals and also for representing the brain activities as a whole.

5.3.2 MI tasks signal classification by logistic regression (LR)

This study develops a LR to predict the probability of two categories of MI tasks from the EEG datasets in a BCI system as the utilization of the LR in pattern recognition is still in its infancy although it is frequently used in statistic science (Liao et al., 2007; Xie et al., 2008), engineering (Caesarendra et al., 2010) and biomedical research (Mrowski et al., 2009; Ryali et al., 2010). The LR is a standard method for the identification of binary outcomes. So it has been extended to the MI tasks classification in this study.

In this study, a MI task in two categories is used as a dependent variable y and the six statistical features are considered as the independent variables, which are x_1= *mean*, x_2= *maximum*, x_3= *minimum* values, x_4= *standard deviation*, x_5= *median*, x_6= *mode* in Equation (5.2).

5.3.3 Performance evaluation methods

5.3.3.1 *k*-fold cross validation method

The *k*-fold cross-validation (Abdulkadir, 2009; Siuly et al. 2011c) used in this research is very popular for assessing the performance of a classification method in pattern recognition. In *k*-fold cross-validation procedure, a data set is partitioned into *k* mutually exclusive subsets of approximately equal size and the method is repeated *k* times (folds). Each time, one of the subsets is used as a test set and the other *k-1* subsets are put together to form a training set. Then the average accuracy across all *k* trials is computed.

A common choice for the *k*-fold cross validation is *k*=10 but in practice, the choice of the number of folds depends on the size of the dataset. For our experimental datasets, the 3-fold cross-validation is found adequate. We attempt to reduce the computation time and the number of experiments. For this reason, this study considers *k=3* in the *k*-fold cross-validation method for all experiments.

5.3.3.2 Classification accuracy

As accuracy is a major concern in BCI systems, this study uses classification accuracy as criterion to evaluate the performance of the proposed method. The accuracy of a measurement system is the degree of closeness of measurements of a quantity to its actual (true) value (Siuly et al., 2011c; Siuly et al., 2011d). The classification accuracy of the proposed classifier for each fold in the 3-fold cross-validation method is measured using the following formula (Dutta et al., 2010):

$$\text{CA}= (\frac{N_{cn}}{N_{tn}})*100 \tag{5.4}$$

where CA is the percentage classification accuracy, N_{cn} is the number of correctly classified samples and N_{tn} is the total number of samples.

5.3.3.3 Confusion matrix

In this work, a confusion matrix is used to present the actual and the predicted classification outcomes by the proposed method. It identifies the common misclassifications of the proposed classification system (Ubeyli, 2010).

5.3.4 Experimental Data

This study uses two benchmark datasets, IVa and IVb from BCI Competition III to evaluate the efficacy of the proposed approach. All EEG recorded data for these two sets were collected during MI tasks. The description of dataset IVa is provided in Section 4.3 of Chapter 4 and dataset IVb is described as below.

Dataset IVb (BCI competition III, 2005; Blankertz et al., 2006) was collected from one healthy male subject. He sat in a comfortable chair with arms resting on

armrests. This data set has the data from seven initial sessions without feedback. The EEG data consisted of two classes: left hand and right foot MI tasks signals were recorded from 118 channels in 210 trials. 118 EEG channels were measured at the positions of the extended international 10-20-system. Signals were band-pass filtered between 0.05 and 200 Hz and digitized at 1000 Hz with 16 bit (0.1 μV) accuracy. The data was down-sampled at 100 Hz, which is used in this research.

5.4 Results and Discussions

This section presents the experimental results of the proposed algorithm on two benchmark EEG datasets, Sets IVa and IVb, used in the BCI Competition III, and also provides a comparison of the present method with two recent reported methods for dataset IVa. As we did not find any research reports for the dataset IVb in the literature, we could not compare the experimental results with other methods. Each subject of both datasets is considered separately for an experiment as the MI EEG signals are naturally highly subject-specific depending on physical and mental tasks. In this study, all experimental results for both datasets are presented based on testing sets. In this chapter, we used MATLAB software package (version7.7, (R2008b)) for all mathematical calculations and PASW (Predictive Analytics SoftWare) Statistics 18 for the LR model.

5.4.1 Classification results for dataset IVa

In our proposed algorithm, we develop the CC technique to extract the representative features from the MI based EEG data and employ the LR model for the classification of the extracted MI features. Dataset IVa contains MI EEG records from five healthy subjects labelled ‘aa’, ‘al’, ‘av’, ‘aw’, ‘ay’ which are denoted as subject 1, subject 2, subject 3, subject 4 and subject 5, respectively, in this chapter. Each of the five subjects performed two MI tasks denoted as two classes: right hand denoted by ‘RH’ and right foot denoted by ‘RF’. Table 5.1 presents the information about the structure of the dataset. As shown in Table 5.1, every sample of the training trials contains class labels but the testing trials do not have class labels with the samples. In this research, we used the training trials in our experiments as the proposed algorithm requires a class label at each data sequence.

Table 5. 1: The information of original data for BCI Competition III, dataset IVa.

Subject	Size of data with two classes (RH and RF)	Among 280 trials	
		Number of trials considered as a training trial with class label	Number of trials considered as a testing trial without class label
1 (aa)	298458×118	168	112
2 (al)	283574×118	224	56
3 (av)	283042×118	84	196
4 (aw)	282838×118	56	224
5 (ay)	283562×118	28	252

Figure 5.2 shows the typical RH and RF MI signals for each of the five subjects for dataset IVa. In each of the five subjects, the Fp1 channel of the RH MI class is considered as a reference channel (reference signal) as there is no specific selection criterion in the CC system. As mentioned before, there are 118 channels in each of the two classes of a subject. In this study, the reference channel data is cross-correlated with the data from the remaining 117 channels of the RH class and 117 cross-correlation sequences are obtained for this class. Again, in the RF class of the same subject, the reference channel data is cross-correlated with each of 118 channels data and produces 118 cross-correlation sequences. Thus, a total of 235 cross-correlation sequences are obtained for the two-class MI data for each subject.

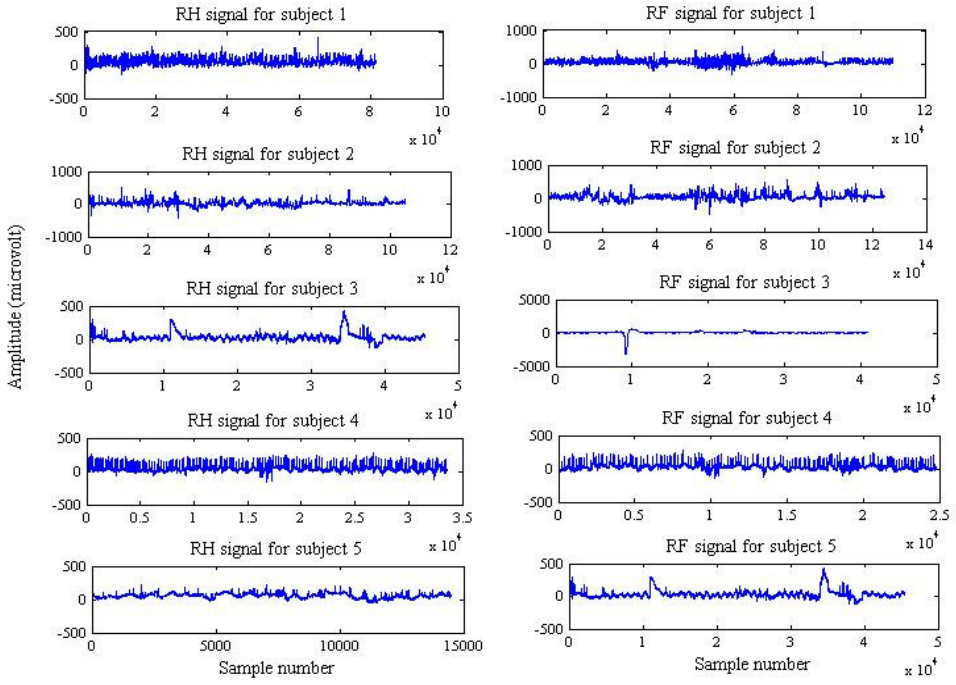


Figure 5.2: The typical signals of the RH and the RF MI tasks for each subject of dataset IVa.

Figure 5.3 presents the results of cross-correlation sequences called cross-correlograms for the RH and the RF MI data of subject 1, subject 2, subject 3, subject 4 and subject 5, respectively. It is important to note that the cross-correlogram or cross-correlation sequences (R_{xy}) are calculated using equation (5.1) for each lag. From the figure, one can see that in most of the cases, the shapes of the two curves for a subject are not exactly same, which indicate the statistical independency. That means, there is more of a chance to achieve better separation. From each cross-correlogram of a subject of dataset IVa, the six statistical features as described in Section 5.3.1 are calculated. Thus, in the case of each subject, we obtain 117 feature vectors of six dimensions for the RH class and 118 feature vectors of six dimensions for the RF class. Thus a total 235 feature vectors of six dimensions are obtained for the two classes of each subject.

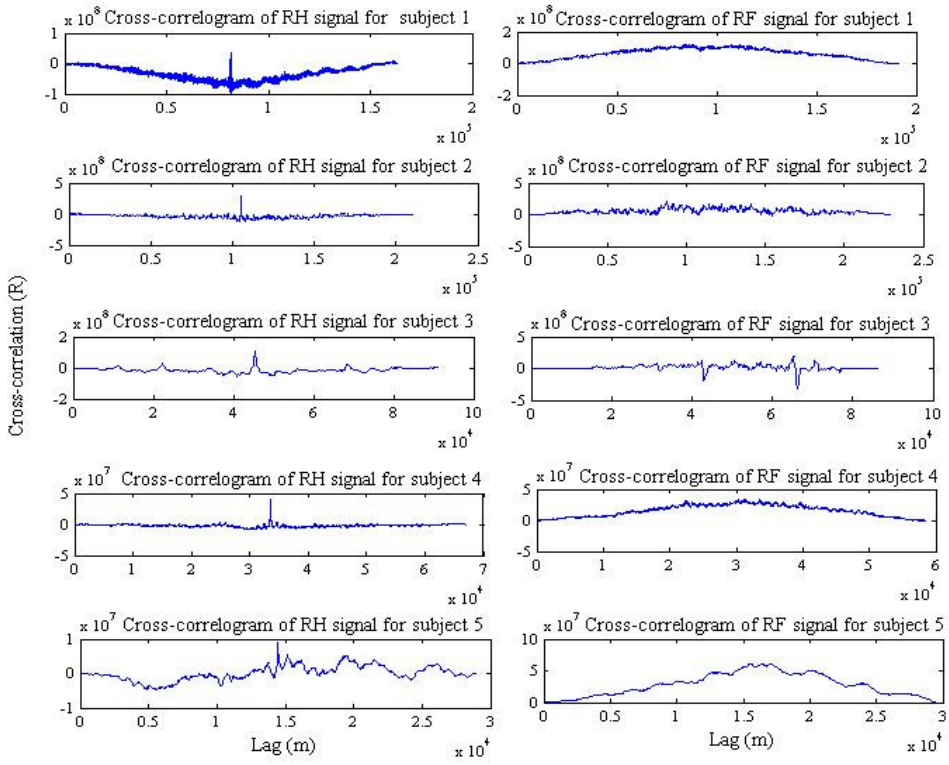


Figure 5.3: The typical cross-correlograms for the RH and the RF MI signals of each subject in dataset IVa.

According to the 3-fold cross validation procedure, the 235 feature vectors of six dimensions are divided into three subsets containing the equal number of observations. In this study, each of the three subsets consists of 78 feature vectors (39 vectors from each class). Each time, a subset is used as a testing set and the remaining two subsets comprise a training set. The procedure is repeated three times

(the folds) with each of the subsets as the test set. Finally the average classification accuracy is then evaluated across all three folds, that is called the cross validation accuracy.

We utilize the training set and the testing set to the LR model in equation (5.2) for estimating the probability of the dependent variable. In equation (5.2), we consider the MI tasks as the dependent variable *y,* which has two values, 0 or 1. Here the RH class is treated as 0 and the RF is as 1 for dataset IVa. The six statistical features are considered as six independent variables in equation (5.2) where x_1=*mean* values, x_2=*maximum* values, x_3=*minimum* values, x_4=*standard deviation* values, x_5=*median* values and x_6=*mode* values. Parameters $\beta_0, \beta_1, \beta_2, ..., \beta_6$ are calculated using the MLE.

Table 5.2 shows the classification accuracies and standard deviations of each of the three folds. It also presents the average classification accuracy for all five subjects of dataset IVa. Using the 3-fold cross validation procedure, the proposed CC-LR algorithm produces the cross validation accuracy of 96.57%, 82.9%, 100%, 96.6%, and 82.9% for subject 1, subject 2, subject 3, subject 4 and subject 5, respectively. The standard deviations among three folds of subject 1, subject 2, subject 3, subject 4 and subject 5, are 2.65, 5.35, 0.0, 3.21 and 0.69, respectively. One can see that there is no significant differences among the three fold accuracies of a subject, which indicate a consistency of the proposed method.

Table 5.2: The 3-fold cross validation results by the CC-LR method on testing set for dataset IVa of BCI Competition III.

Subject	3- fold cross validation accuracy and their standard deviation (%)				
	1-fold	2-fold	3-fold	Cross validation accuracy (average of three folds)	Standard deviation among three folds accuracy
1	93.6	98.7	97.4	96.57	2.65
2	76.9	84.6	87.2	82.9	5.35
3	100.0	100.0	100.0	100.0	0.0
4	93.6	100.0	96.2	96.6	3.21
5	82.1	83.3	83.3	82.9	0.69
Average for all five subjects				91.79	2.38

As shown in Table 5.2, the CC-LR method provides the highest cross validation accuracy at 100% and zeros the standard deviation for subject 3.The average cross validation accuracy and the standard deviation for all five subjects obtained was 91.79% and 2.38, respectively.

To provide the information on how the MI tasks are predicted and classified in each fold, we describe the output for the 1-fold of subject 1in dataset IVa, as an example. Table 5.3 displays a confusion matrix to show the classification results of the LR classifier for the 1-fold of subject 1 to dataset IVa. From the confusion matrix, it is observed that in the 1-fold of subject 1, two values of the RH class are misclassified to the RF class and three values of the RF class are misclassified as the RH class. The correct classification rate is 94.9% for the RH class and 92.3% for the RF class. The overall accuracy is 93.6% in the 1-fold for subject 1.

Table 5.3: Confusion matrix for the 1-fold of subject 1 from dataset IVa.

	Predicted outcome			Correct classification rate (%)
Observed values		RH	RF	
	RH	37	2	94.9
	RF	3	36	92.3
Overall				93.6

Figure 5.4 depicts typical scenery of a classification plot on how observed values are predicted and classified by the LR model in the 1-fold of subject 1for the dataset. In this figure, the observed group values of the two MI tasks and the predicted probability values obtained from Equation (5.2) are plotted in the X-axis. The frequency values of the observed and predicted probability are plotted in the Y-axis. As mentioned before, in dataset IVa, 0 is indicated as the RH class and 1 as the RF class.

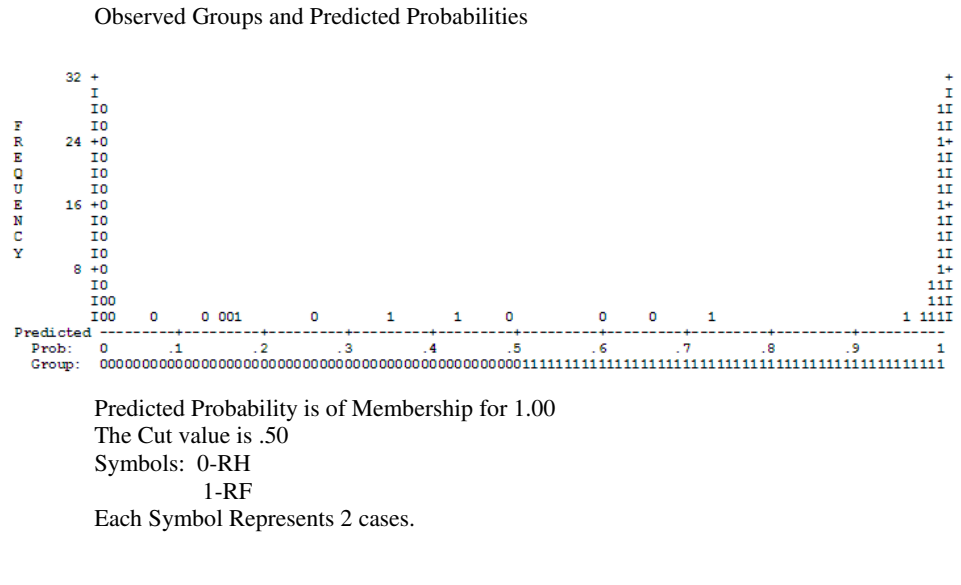


Figure 5.4: Classification plots for the 1-fold of subject 1 in dataset IVa.

From Figure 5.4, it is seen that three values of the RF class (denoted by 1) are misclassified as the RH class (denoted by 0) whereas two values of the RH class are misclassified with the RF class. Table 5.3 also shows similar results. The reflection of the confusion matrix is shown in Figure 5.4.

A comparison of the proposed algorithm with two recent reported algorithms is shown in Table 5.4. The highest classification accuracy from three algorithms is highlighted for each subject and their average. The proposed CC-LR algorithm provides better classification accuracies than the other two recent reported algorithms in three out of the five subjects. From Table 5.4, it is seen that the proposed algorithm produces the highest classification accuracy of 96.57% for subject 1, 100.0% for subject 3 and 96.6% for subject 4 among the R-CSP with aggregation (Lu et al., 2010) and CT-LS-SVM (Siuly et al. 2011c) algorithms.

Table 5.4: Performance comparison of the CC-LR algorithm with the R-CSP with aggregation and the CT-LS-SVM algorithms for dataset IVa, BCI III.

subject	Classification accuracy rate (%)		
	CC-LR method	R-CSP with aggregation (Lu et al., 2010)	CT-LS-SVM (Siuly et al., 2011c)
1	96.57	76.8	92.63
2	82.9	98.2	84.99
3	100.0	74.5	90.77
4	96.6	92.9	86.50
5	82.9	77.0	86.73
Average	91.79	83.9	88.32

We obtain 82.9% accuracy with the proposed method for subject 2, which is a bit less than the highest rate (98.2%) of the R-CSP with aggregation algorithm. The classification accuracy of the present method for subject 5 is 82.9%, which is better than the R-CSP with aggregation algorithm (77.0%) but the CT-LS-SVM algorithm obtained 86.73%. The results demonstrate that the average classification accuracy of the proposed method increases by 3.47% in comparison to the CT-LS-SVM algorithm and 7.89% compared to R-CSP with aggregation. Based on these results, it can be concluded that the CC-LR method do better than the recent reported two algorithms in the MI tasks signal classification.

For further justification, we also compare the proposed method with four other existing methods in the literature for the dataset shown in Figure 5.5. It is seen from the figure that the present algorithm is compared with composite CSP (method 1, n=3) (Kang et al. 2009), composite CSP (method 2, n=3) (Kang et al. 2009), R-

CSP (Lu et al. 2009) and SSFO (Yong et al. 2008). As shown in Figure 5.5, the proposed algorithm yields the best accuracy for subject 1, subject 3 and subject 4 compared to the other existing methods.

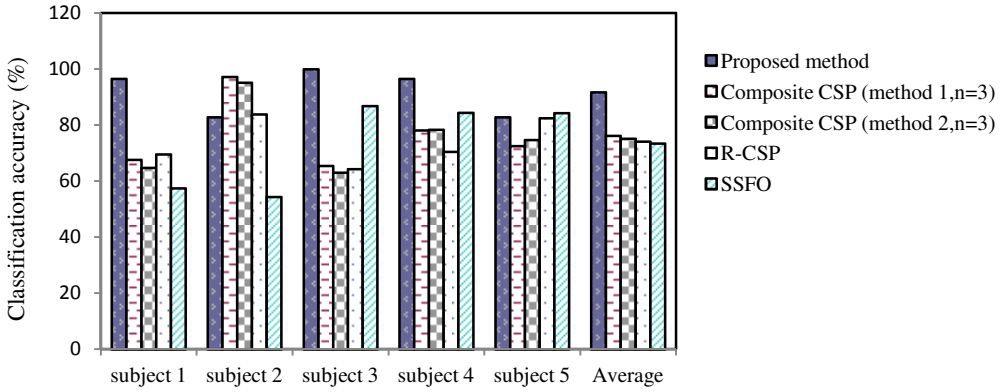


Figure 5.5: Performance comparisons of four other existing methods of the literature with the proposed CC-LR method.

The composite CSP (method 1, n=3) achieved a better classification accuracy than our method for subject 2. The classification accuracy of the proposed algorithm is similar to the other methods for subject 5. It is observed from Figure 5.5 that the highest average classification accuracy is obtained by our proposed algorithm among the other four existing methods in the literature.

5.4.2 Classification results for dataset IVb

As mentioned before, dataset IVb of BCI Competition III was formed from one healthy subject who performed left hand denoted by 'LH' and right foot denoted by 'RF' MI tasks. Here each task is considered as a class. The dataset has two portions, training data and testing data. We use the training data in our experiment as the training data includes with class labels of each observation. The original data size of training set is 210259×118.

For this dataset, the Fp1 channel has been chosen as the reference channel from the RF MI class. This reference channel is cross-correlated with the data from the rest of 117 channels of the RF MI class to create 117 cross-correlation sequences. The reference channel is again cross-correlated with each of 118 channels of LH MI class to produce 118 cross-correlation sequences. Therefore, a total of 235 cross-correlation sequences are obtained from the dataset.

Figure 5.6 depicts the typical signals of the RF and LH MI tasks and their cross-correlograms for dataset IVb. From this figure, it is observed that the shapes of two waveforms are not the same, so there is greater chance to get a better separation. The six statistical features mentioned before (see Section 5.3.1) are calculated from each cross-correlogram. We obtain 117 feature vectors of six dimensions for the RF MI class, and 118 feature vectors of the same dimensions for the LH MI class. Finally we obtain a total of 235 feature vectors with six dimensions from the dataset. These features are segregated as the training and testing sets through the 3-fold cross validation process. The obtained feature vector sets are used as input variables to the LR classifier for the prediction and classification of the EEG-based MI tasks.

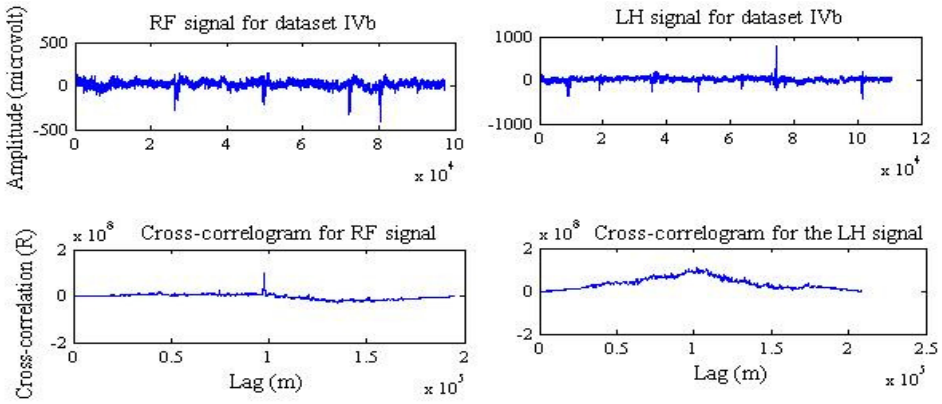


Figure 5.6: The typical signals and cross-correlograms for the RF and the LH MI signals of dataset IVb.

Table 5.5 gives the classification accuracy for each of the three folds and also the average classification accuracies and standard deviations for all three folds from dataset IVb.

Table 5.5: The 3-fold cross validation results by the proposed method on testing set for dataset IVb of BCI Competition III.

Fold	Cross validation accuracy (%)
1	85.9
2	94.9
3	100
Average for three folds	93.6
Standard deviation	7.14

It can be observed from Table 5.5 that the LR model classifies the RF and LH MI data with the accuracy of 85.9%, 94.9% and 100% for the 1-fold, the 2-fold and the

3-fold, respectively. The average cross validation accuracy and standard deviation for the three folds is 93.6% and 7.14, respectively.

Table 5.6 provides the prediction information about the classification for the 1-fold of dataset IVb. From Table 5.6, we see that seven values of the RF class are misclassified as the LH class. On the other hand, four values of the LH class are misclassified as the RF class. The correct classification rate is 82.1% for the RF class and 89.7% for LH class. The overall correct classification rate reaches at 85.9% for the 1-fold of this dataset. We can also see the mirror image of this confusion matrix in Figure 5.5.

Table 5.6: Confusion matrix for the 1-fold of dataset IVb.

	Predicted outcomes			Correct classification rate (%)
Observed values		RF	LH	
	RF	32	7	82.1
	LH	4	35	89.7
Overall				85.9

In order to describe visibly the classification seceneryof the LR classifier, Figure 5.6 exhibits the classification plot for the 1-fold of dataset IVb, as an example. This figure clearly explains the confusion matrix reports presented in Table 5.6. In this figure, 0 denotes the RF class and 1 denotes the LH class.

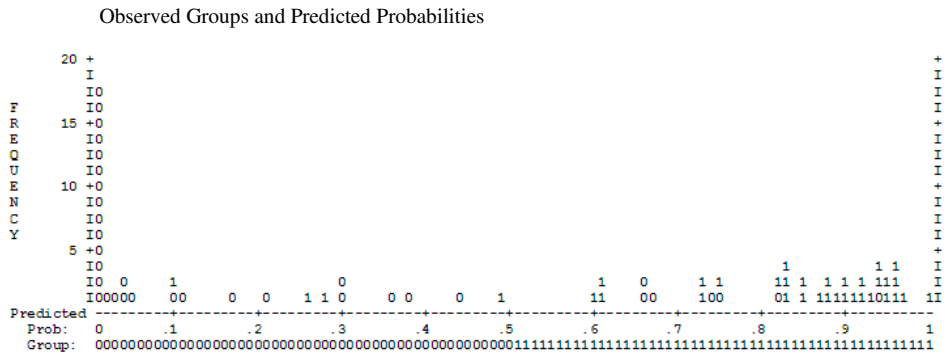


Figure 5.7: Classification plot for the 1-fold of dataset IVb.

As shown in Figure 5.7, four values of the LH class are misclassified as the RF class and seven values of the RF class are misclassified to the LH class in the 1-fold of

dataset IVb. From the experimental results, it is obvious that the CC technique is efficient for extracting features from the MI data and the LR classifier has the inherent ability to identify the MI tasks in BCIs. The experimental results of the two datasets prove that the proposed CC-LR algorithm is promising for the classification of MI tasks and it offers a great potential to improve the performance.

5.5 Conclusions and Recommendations

This study develops a novel algorithm for the classification of the MI tasks in BCI applications combining cross-correlation and logistic regression methods. The performance of the proposed algorithm is measured in terms of classification accuracy using a 3-fold cross-validation method. The experiments are performed on datasets, IVa and IVb from BCI Completion III. The current approach is compared with two recent reported algorithms, R-CSP with aggregation (Lu et al., 2010) and CT-LS-SVM (Siuly et al., 2011c) for dataset IVa. To further validate, the efficacy of the proposed algorithm is also compared with four other algorithms from the literature. The experimental results have demonstrated the effectiveness of the proposed CC-LR algorithm, especially its superiority over the reported algorithms. Moreover, the CC-LR method is efficient for the identification of MI tasks that can provide positive impacts on developing BCI systems.

In this chapter, our proposed approach provides a structure for the recognition of the MI EEG signals randomly considering the reference signal of Fp1 electrode. But the Fp1 electrode does not transmit the motor imagery related information well from the brain according to the international 10-20 electrode placement systems. One feature set is extracted from each cross-correlation sequence for using as input to the LR classifier but it could not provide enough performance for competition. That's why, in the next chapter, we modify the CC-LR algorithm with the C3 reference channel employing the diverse features to this algorithm.

CHAPTER 6
MODIFIED CC-LR ALGORITHM WITH DIVERSE FEATURE SETS

In Chapter 5, we introduced the CC-LR algorithm for the classification of the MI signals where the reference signal, Fp1 was randomly considered for the CC technique. We extracted one feature set from each CC sequence to represent the MI tasks EEG data for that method. However, the algorithm could not provide a clear idea about whether the reference signal and the feature set are the optimal choices or not, as all EEG signals are not equal in providing informative measurements about motor tasks.

In order to alleviate these concerns, this chapter develops a modified version of the CC-LR algorithm. The modified CC-LR algorithm investigates which feature set is superior to characterize the distribution of MI tasks EEG data. After investigation of three diverse feature sets, finally this study reaches a conclusion on which features set is best to characterize the EEG signals. This chapter also provides an insight into how to select a reference channel for the CC technique with EEG signals considering the anatomical structure of the human brain. The C3 electrode channel (International 10-20 system) is chosen as a reference channel for this modified algorithm since this channel can provide more relevant information about the MI task. The proposed algorithm is compared with the most recently reported eight well-known methods including the BCI III Winner algorithm. The experimental results demonstrate that the modified CC-LR algorithm provides a classification improvement over the existing methods tested.

The contents of this chapter have been submitted to the journal of *Applied Mathematics and Computation* for publication.

6.1 Background

A communication or control based BCI technology requires patterns of a brain activity that can be consciously generated or controlled by a subject and ultimately clearly distinguishable by a computer system. This system involves a BCI user to

concentrate on a motor imagery (MI) task in order to produce a characteristic brain pattern that identifies with the desired control, for example, the imagination of a hand movement (Thomas et al., 2009) and each MI task is usually treated as a class. In order to improve BCI systems, it is essential to correctly identify different MI classes that are related to a performed task using classification techniques.

Classification techniques assist to predict and identify a qualitative variable, e.g. the class label of a subject's mental state by extracting useful information from the highly multivariate recordings of the brain activities. Feature extraction in BCI research is a major task that would significantly affect the accuracy of classifying MI tasks. We can obtain a good classification rate if the extracted features are efficient for differentiating MI tasks. An efficient feature extraction method can achieve good classification results with a simple classifier. The role of a classification method in BCI systems is to identify a subject's intentions from a number of predefined choices. Following a feature extraction procedure, a suitable classifier needs to be designed. Even though many methods have been reported in the last decade yielding impressive results related BCI data for the MI tasks classification, still most BCI technologies are not satisfactory.

For the MI tasks classification, we developed the CC-LR algorithm (Siuly et al., 2011b) in Chapter 5. In that algorithm, we randomly considered the EEG signal from the electrode position Fp1 as a reference signal for the CC technique. We extracted from each cross-correlation sequence to represent the MI tasks EEG data. To reduce experimental time and the number of experiments, we used the *3-fold cross validation method* for the evaluation of the performance of the algorithm. The experimental results show that the algorithm works well to recognize the MI tasks. However, the algorithm (Siuly et al., 2011b) could not provide a clear idea about whether the feature set and the reference signal are the optimal choices or not, as all EEG signals are not equal in providing informative measurements about motor tasks.

To overcome the problem of the CC-LR algorithm in (Siuly et al., 2011b), this study proposes a modified version of the CC-LR method to classify two-class MI tasks EEG signals. To investigate which features are suitable for the representation of the MI signals, three statistical feature sets (described in Section 6.2) are extracted from each cross-correlation sequence of a subject and then evaluated. Finally this paper reaches a conclusion on which features set is best to characterize the EEG

signals. This study also reports on how a reference channel is selected for the CC method considering the structure of the brain associated with MI tasks.

In the present work, the experimental evaluation is performed on two benchmark datasets, IVa and IVb of BCI competition III like as Chapter 5. A popular *k*-fold cross validation method (*k*=3) is used to assess the performance of the proposed method for reducing the experimental time and the number of experiments in the MI tasks EEG signal classification. This cross-validation procedure is applied as a way to control over-fitting of the data. The performance of the proposed approach is also compared with eight most recent reported methods including BCI III winner in dataset IVa. The study results from the both datasets demonstrate that our proposed algorithm produces a promising performance for the MI tasks EEG signal classification.

6.2 Methodology

This study proposes a potent method for the MI tasks classification in the BCI development, which is shown in Figure 6.1. This method provides an important framework to classify the two-class MI tasks based-EEG signals for BCI data. As we consider a pattern recognition system for the MI data, the EEG recognition procedure mainly involves feature extraction and classification process.

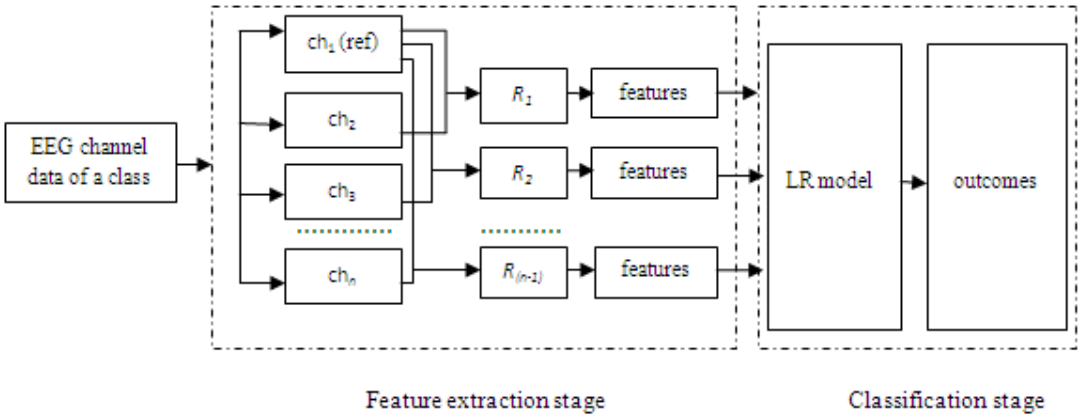
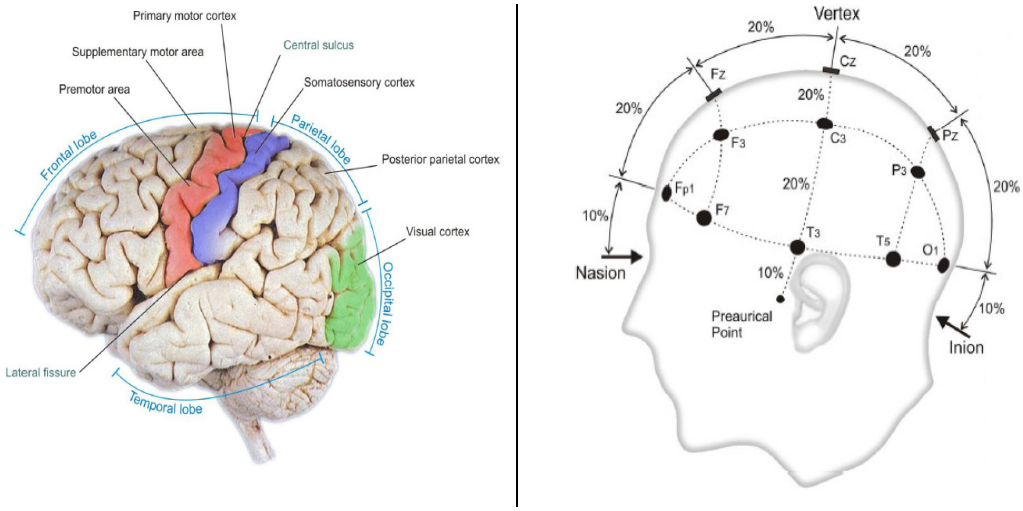


Figure 6.1: Schematic diagram for the classification of the MI tasks based EEG signal in BCIs. Here ch=channel, *R*=cross-correlation sequence, LR=logistic regression, ref=reference channel.

As shown in Figure 6.1, the building components of the proposed system are divided into two major parts where the feature extraction procedure is described in the first part and the classification technique in the second part. In the feature extraction stage, features are extracted using the CC technique (described in Chapter 5) from different channels of the EEG data of each and every MI class of a subject by means of following three steps:

Step 1: In the first phase, one EEG channel is chosen as a reference channel corresponding an electrode, which is likely to provide informative measurements about motor tasks. It is believed that only a particular part of the brain is activated in response to an MI task and the channels that are close to the active brain regions have more relevant information about the MI task compared to all other channels. As the motor cortex area of the brain is typically associated to the MI movements with the EEG position C3 in the international 10-20 system for electrode placement (Sander et al., 2010), so C3 channel can provide more information about brain activities during the MI tasks.



(a) (b)

Figure 6.2: (a) Structure of the human brain (b) the international 10-20 electrode placement system

Figure 6.2 gives the information of the human brain structure and also shows the locations of the 10-20 electrode placements. From Figure 6.2(b), it is obvious that C3 is in an important position of the motor cortex area, which is very responsive for supplying the MI information. This study considers C3 channel as a reference

channel for the CC method. As shown in Figure 6.1, ch_1 (channel 1) is considered as the reference channel, as an example.

Step 2: In this step, the CC technique is used to calculate a cross-correlation sequence denoted by 'R_{xy}' between the reference channel data and any other channel data. The reference channel of a class is cross-correlated with the data of the remaining channels of the current class and the data of all channels of any other classes. The graphical presentation of a cross-correlation sequence is called a cross-correlogram. From Figure 6.1, it is seen that a cross-correlation sequence R_1 is created for the reference channel (ch_1) and the ch_2 channel; R_2 for the reference channel and the ch_3 channel; and $R_{(n-1)}$ for the reference channel and the ch_n channel in a class. So the total $(n-1)$ cross-correlation sequences are obtained for n channels for this class when one channel is treated as the reference channel. If the reference channel is not chosen from this class (if considered from another class), the total n cross-correlation sequences are obtained from this class.

Step 3: Statistical features are extracted from each cross-correlation sequence to characterize the distribution of EEG signals, which reduce the dimension of the cross-correlation sequence. To investigate which features can produce the best performance with the proposed classifier, the following three sets of features are extracted from each cross-correlation sequence.

- **Two-feature set:** This set consists of two features, *mean* and *standard deviation that* are calculated from each cross-correlation sequence. When one is interested in describing an entire distribution of some observations, these two features are especially useful to represent a distribution (Islam, 2004; De Veaux, 2008). *Mean* represents the distribution of a signal and *standard deviation* describes the amount of variability in a distribution.
- **Four-feature set:** In this set, another two features, *skewness* and *kurtosis* are added with the features, *mean* and *standard deviation* in the two-feature set. *Skewness* describes the shape of a distribution that characterizes the degree of asymmetry of a distribution around its mean (Islam, 2004; De Veaux, 2008). K*urtosis* measures of whether the data are peaked or flat relative to a normal distribution.
- **Six-feature set:** *maximum* and *minimum* features are added in the four-feature set. Thus the six-feature set consists of six features, which are *mean*, *standard deviation*, *skewness*, *kurtosis*, *maximum* and *minimum*. *Maximum* and *minimum*

provide a range of the observations and a non-parametric prediction interval in a data set from a population (Islam, 2004; De Veaux, 2008).

As shown in Figure 6.1, the classification stage is carried out with two phases described below:

Phase 1: In this phase, the LR model (described in Chapter 5) is employed as a classifier to classify extracted features. Each of the above mentioned three features sets are used in the LR model, individually, as the input. Then the classification performances are evaluated for each of the three sets.

Phase 2: Classification outcomes from each feature set are obtained in this stage. Based on the outcomes, we can decide which feature set is the best for the LR classifier to classify the MI signals.

6.3 Experimental Evaluation and Discussion

This section presents an implementing procedure and experimental results of the proposed algorithm for the two benchmark EEG datasets, IVa (described in Chapter 4) and IVb (described in Chapter 5), used in the BCI Competition III. These datasets are used for the evaluation of the proposed algorithm to classify different EEG signals during the MI tasks. In this study, MATLAB software package (version7.7, (R2008b)) is used for the computation of the CC technique and PASW (Predictive Analytics SoftWare) Statistics 18 is used for the LR model.

6.3.1 Implementation of the CC technique for the feature extraction

One of the challenging tasks in BCIs is to detect features reliably. Those features represent very weak brain activities often corrupted by noise and various interfering artifacts of physiological and non-physiological origins. In this study, we develop a CC technique to extract the representative features from the MI tasks EEG data. To reduce the dimensionality of a cross-correlation sequence, three different statistical features sets are extracted over the resultant data (the reasons of choosing these features for this study are described in Section 6.2).

As mentioned in Chapter 4, dataset IVa of BCI Competition III contains MI tasks EEG records from five healthy subjects labelled ‘aa’, ‘al’, ‘av’, ‘aw’, ‘ay’

which are denoted as S1, S2, S3, S4 and S5, respectively, in this chapter. Each of the five subjects performed two MI tasks categorized as two classes: right hand (denoted by 'RH') and right foot (denoted by 'RF'). As discussed previous chapter, every sample of the training trials contains class labels but the testing trials do not have class labels attached to the samples. In this research, we used the training trials as our proposed algorithm requires a class label at each data point.

In this study, each subject in both datasets is considered separately for experiments as the MI tasks EEG signals are naturally highly subject-specific depending on physical and mental tasks. In each subject of dataset IVa, C3 channel of the RH MI class is considered as a reference channel (reference signal) for the CC approach. As there are 118 channels in each of the two classes for a subject, the reference channel data is cross-correlated with the data from the remaining 117 channels of the RH class and 117 cross-correlation sequences are generated for this class. Again, in the RF class of the same subject, the reference channel data is cross-correlated with each of 118 channel data and produces 118 cross-correlation sequences. Thus, a total of 235 cross-correlation sequences are obtained for a two-class MI data of each subject.

Figures 6.3-6.7 show typical results of cross-correlation sequences called cross-correlograms for the RH and the RF MI data of S1, S2, S3, S4 and S5, respectively. It is important to note that the cross-correlogram or cross-correlation sequences (R_{xy}) are calculated using Equation (5.1) for each time *lag*. From these figures, one can see that in most of the cases, the shapes of the two curves are not exactly the same, which indicate the statistical independency. It indicates, there is more of a chance to achieve a better separation.

From each cross-correlogram of a subject in dataset IVa, the three sets of statistical features as described in Section 6.2 are calculated. Thus, for the RH class of each subject, we obtain 117 feature vectors of two dimensions for the two-feature set, 117 feature vectors of four dimensions for the four-feature set and 117 feature vectors of six dimensions for the six-feature set. For the RF class, 118 feature vectors with two dimensions are obtained for the two-feature set, 118 feature vectors with four dimensions for the four-feature set and 118 feature vectors with six dimensions for the six-feature set. Thus we obtain a total of 235 feature vectors of two dimensions for the two-feature set, 235 feature vectors of four dimensions for the

four-feature set and 235 feature vectors of six dimensions for the six-feature set from the two-class MI tasks EEG data for each subject in the dataset.

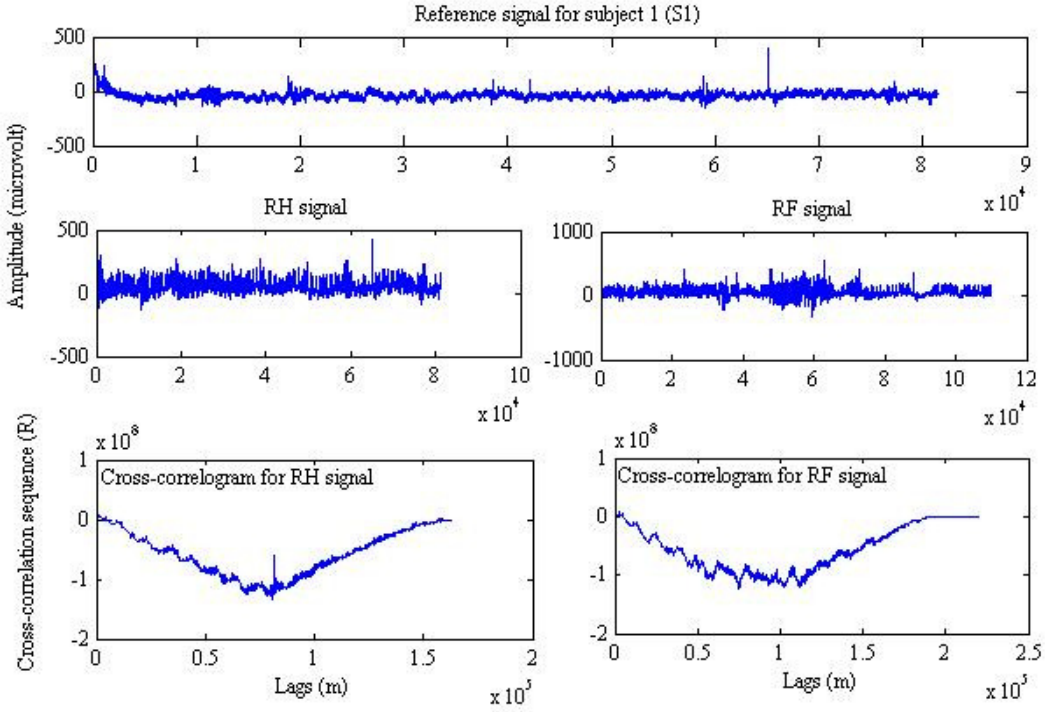


Figure 6.3: The typical cross-correlograms for the RH and the RF MI tasks signals of S1 of dataset IVa.

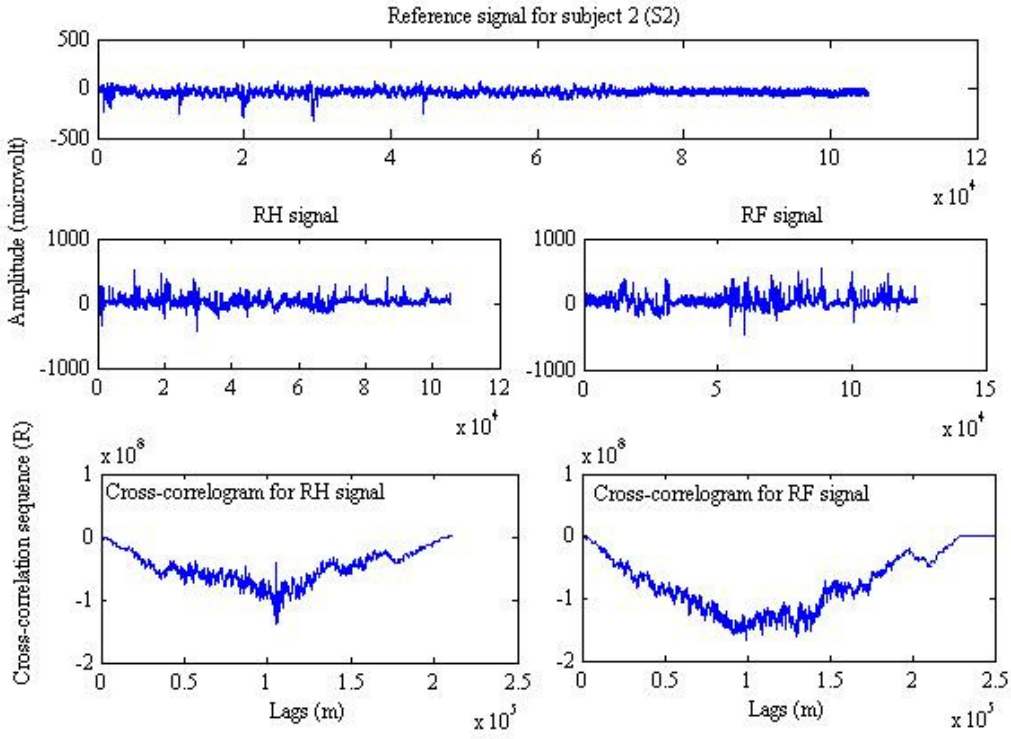


Figure 6.4: The typical cross-correlograms for the RH and the RF MI tasks signals of S2 of dataset IVa.

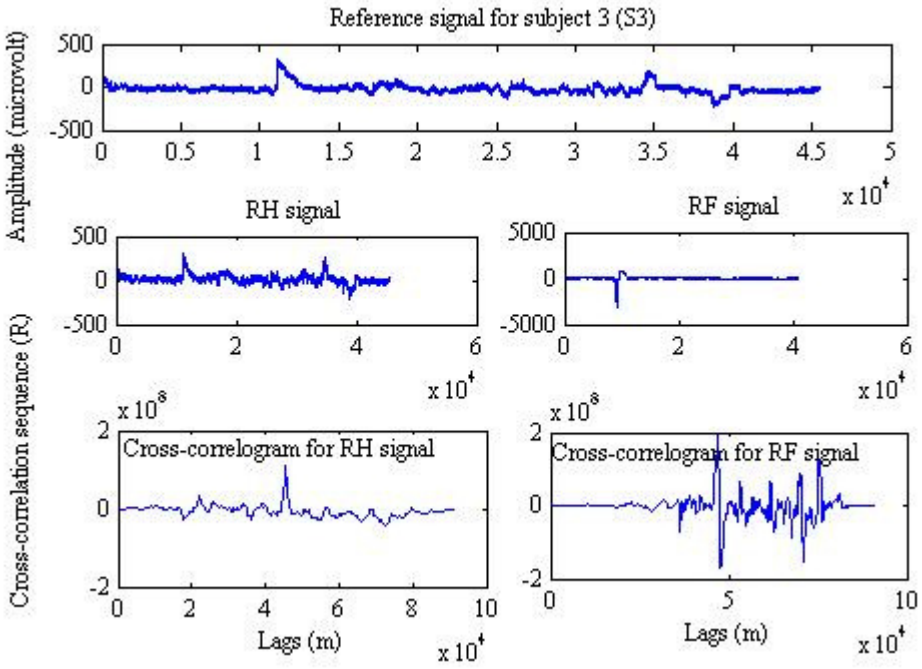


Figure 6.5: The typical cross-correlograms for the RH and the RF MI tasks signals of S3 of dataset IVa.

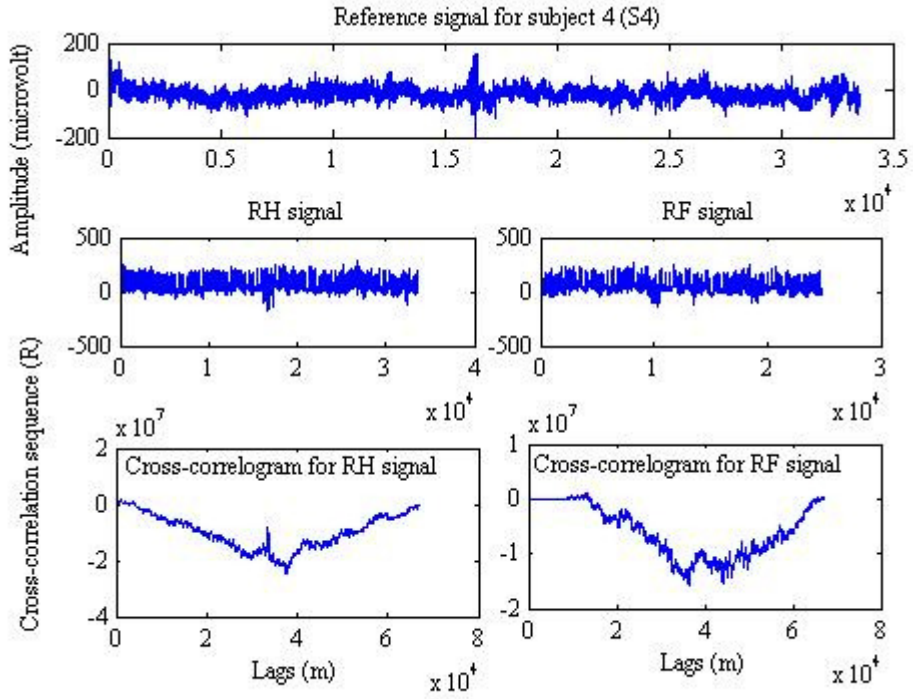


Figure 6.6: The typical cross-correlograms for the RH and the RF MI tasks signals of S4 of dataset IVa.

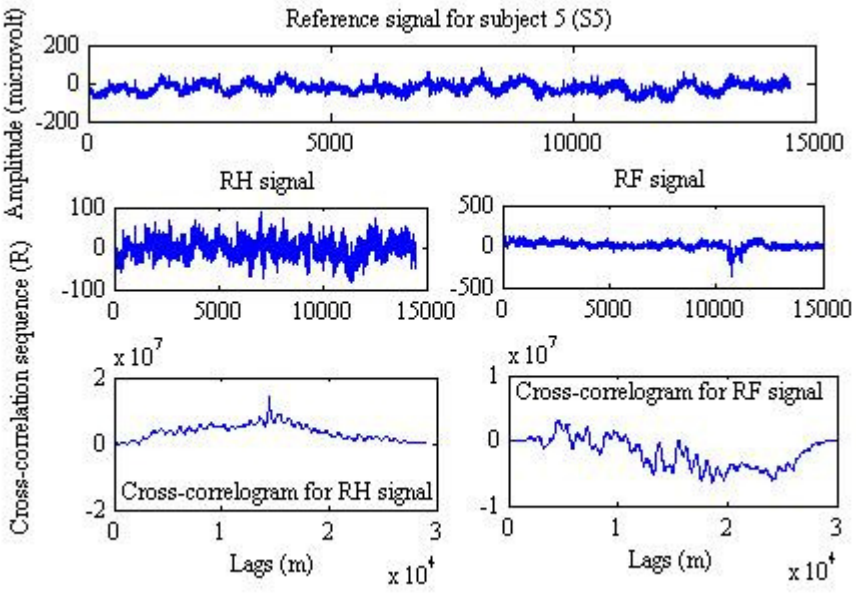


Figure 6.7: The typical cross-correlograms for the RH and the RF MI tasks signals of S5 of dataset IVa.

As introduced in Chapter 5, dataset IVb was generated from one healthy subject who performed left hand (denoted by 'LH') and right foot (denoted by 'RF') MI tasks. Here each task is considered as a class. Dataset IVb has two portions, training data and testing data. We use the training data for our experiment as the training data involves with class labels of each observation but not the testing data. The original data size of training set is 210259×118.

For dataset IVb, C3 channel of the RF class is considered as a reference channel (reference signal). This reference channel is cross-correlated with the data of the rest of 117 channels of the RF MI class and results in 117 cross-correlation sequences for this class. The reference channel is again cross-correlated with each of 118 channels of LH MI class and produces 118 cross-correlation sequences. Therefore, a total of 235 cross-correlation sequences are obtained from this dataset.

Figure 6.8 depicts the typical cross-correlograms for the RF and LH MI signals for dataset IVb. From this figure, it is observed that the shapes of two waveforms are not the same, so there is a greater chance to get a better separation. The three statistical feature sets mentioned before (see Section 6.2) are calculated from each cross-correlogram.

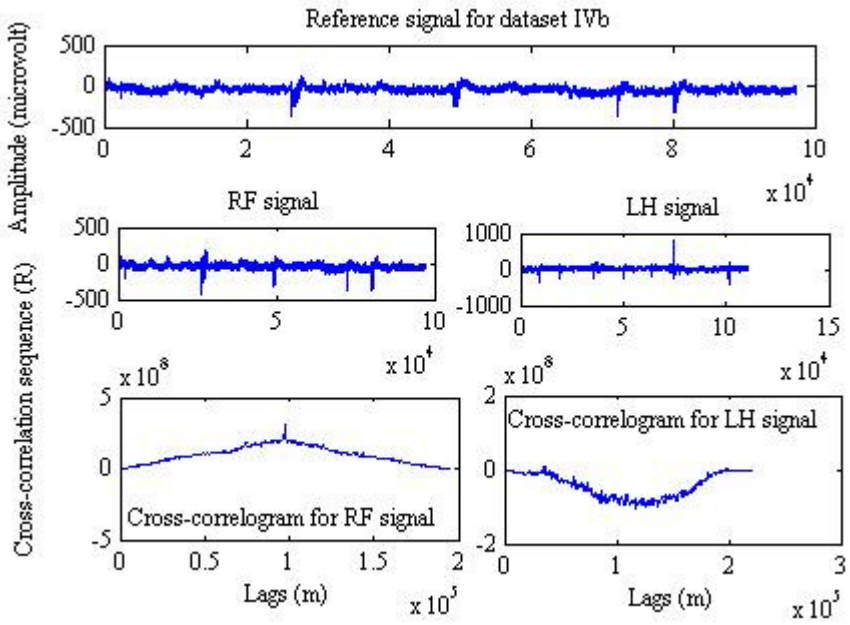


Figure 6.8: The typical cross-correlograms for the RF and the LH MI tasks signals of dataset IVb.

For the RF MI class of dataset IVb, we obtain 117 feature vectors of two dimensions for the two-feature set, 117 feature vectors of four dimensions for the four-feature set and 117 feature vectors of six dimensions for the six-feature set. For the LH MI class, we acquire 118 feature vectors of two dimensions for the two-feature set, 118 feature vectors of four dimensions for the four-feature set and 118 feature vectors of six dimensions for the six-feature set. Finally we obtain a total of 235 feature vectors of two dimensions for the two-feature set, 235 feature vectors of four dimensions for the four-feature set and 235 feature vectors of six dimensions for the six-feature set from the two-class MI EEG signals of the dataset.

In the both datasets, the feature vectors of each of the three sets are segregated randomly as the training and testing sets through the 3-fold cross validation process. Those features are used as the input variables to the LR classifier for classifying the EEG-based MI tasks. In the next section, we are going to discuss how the features of the three sets are used in the proposed LR model to classify the two-class MI tasks EEG data and what results are obtained.

6.3.2 MI classification results testing different features

In this study, the LR model is employed to classify two-class MI EEG signals where the three feature sets (discussed in Section 6.2) are used separately as the input to the

LR model. The expectation of this study is to find an appropriate feature set that are accurate for the two-class MI classification in the LR method. Feature vectors are extracted in a way that they hold the most discrimination information and represent the distribution of the MI EEG data. To evaluate the general efficacy of the proposed classifier, the 3-fold cross validation method is utilized to calculate the classification accuracy on the testing dataset for each of the three feature sets.

Based on the 3-fold cross validation procedure, a total of 235 feature vectors of each set of a subject are randomly divided into three subsets. The first two subsets consist of 78 feature vectors (39 vectors from each class) and the last subset consists of 79 feature vectors of the same dimensions (39 vectors from the reference class and 40 vectors from another class). Each time, a subset is used as a testing set and the remaining two subsets comprise a training set. The procedure is repeated three times (the folds) with each of the subsets as the test set. Finally the average classification accuracy is evaluated across all three folds on the testing set, which is named as 3-fold cross-validation accuracy. In this study, the training set is applied to train the classifier and the testing set is used to verify the classification accuracy and the effectiveness of the classifier. Note that all experimental results for the datasets, IVa and IVb, are presented based on the testing set.

We utilize the training set and the testing set to the LR model in Equation (5.2) for estimating the probability of the dependent variable y. We consider the MI tasks with two classes as the dependent variable y where the RH class is treated as 0 and the RF is as 1 for dataset IVa. For dataset IVb, the RF class is denoted as 0 and the LH is marked as 1. The statistical features mentioned in this study are considered as independent variables. In the case of both datasets, independent variables are considered for the three feature sets in Equation (5.2) as follows.

For the two-feature set:

x_1=*mean* values; x_2=*standard deviation* values

For the four-feature set:

x_1=*mean* values; x_2=*standard deviation* values; x_3=*skewness* values and x_4=*kurtosis* values

For the six-feature set:

x_1=*mean* values; x_2=*standard deviation* values; x_3= *skewness* values; x_4= *kurtosis* values; x_5=*maximum* values and x_6=*minimum* values

The obtained classification results for the two, four and six features sets are presented in Table 6.1 for each subject in dataset IVa. Table 6.1 provides the classification accuracy of the proposed algorithm through the 3-fold cross-validation accuracy. The results of each subject are reported in terms of mean ± standard deviation of the accuracy over a 3-fold cross-validation method on the testing set. The proposed approach for the two-feature set produces the classification accuracies of 55.33% for S1, 55.73% for S2, 95.37% for S3, 56.17% for S4 and 66.4% for S5. The classification accuracies for S1, S2, S3, S4 and S5, reach at 80.8%, 68.76%, 97.47%, 83.87% and 75.73% for the four-feature set, and 100%, 94.23%, 100%, 100% and 75.33% for the six-feature, respectively. The results show that the classification performance increases gradually for all subjects with additional features. Among the three feature sets, the six-feature set results in the highest classification performance for each subject where the six features are used as the inputs to the LR. The four-feature set generates better performance compared to the two-feature set.

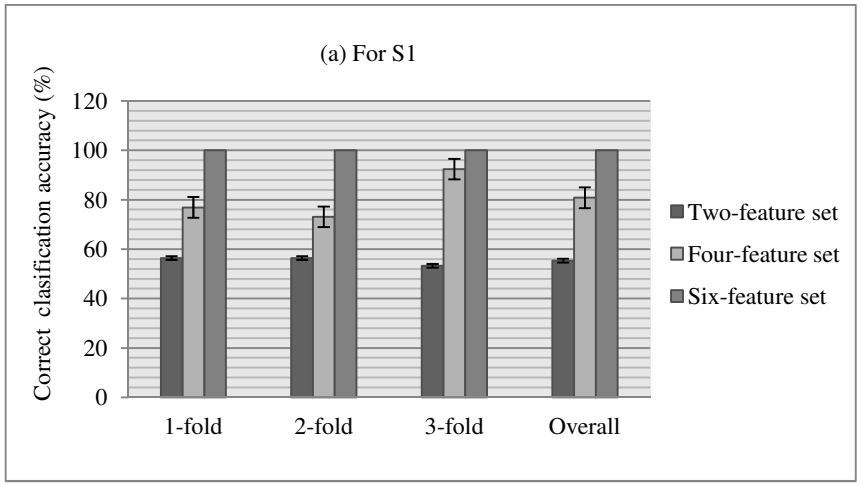
From Table 6.1, one can see that the average classification accuracy of five subjects is 65.80% for the two-feature set, 82.33% for the four-feature set and 93.91% for the six-feature set. The experimental results demonstrate that the performance of the proposed method for the four-features set has been improved by 16.53% compared to the two-feature set for adding two more features, *skewness* and *kurtosis*. The performance of the six-feature set is increased by 11.58% compared to the four-feature set for adding another two features, *maximum* and *minimum*. It can be concluded that more features can substantially improve the performance for all subjects.

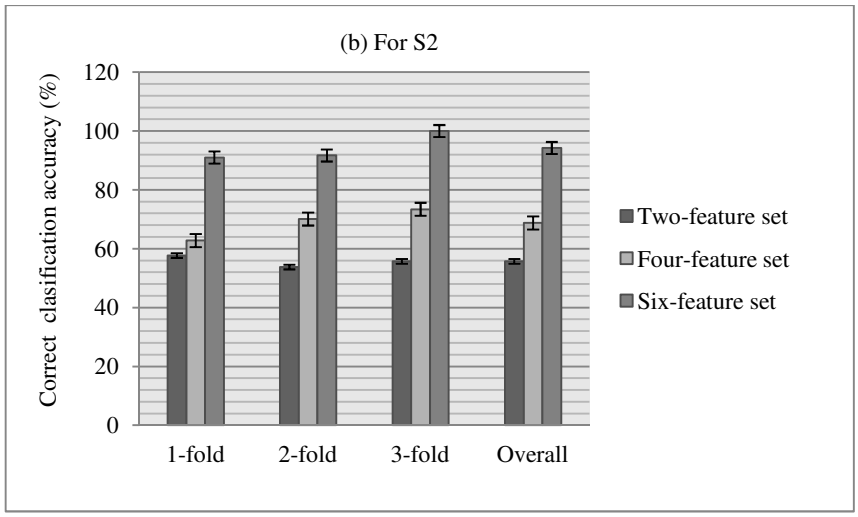
Table 6.1: Cross-validation results with the proposed method on testing set for dataset IVa.

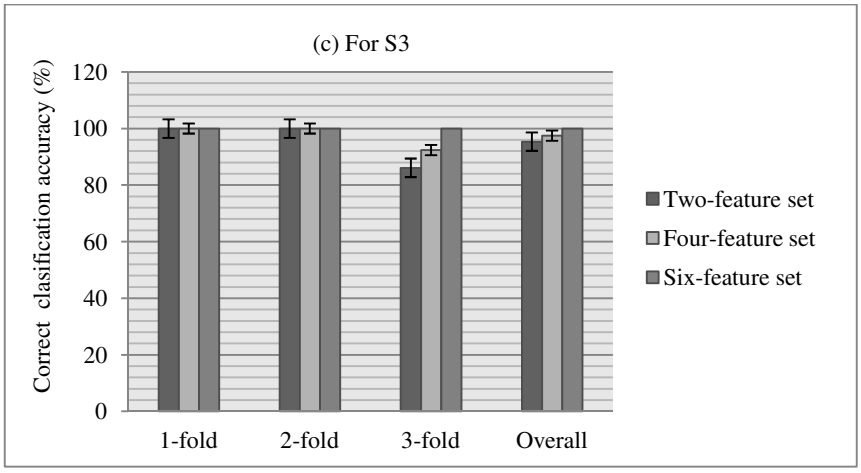
Subject	3-fold cross-validation accuracy (mean± standard deviation) (%)		
	Two-feature set	Four-feature set	Six-feature set
S1	55.33±1.85	80.80±10.22	100.0±0.0
S2	55.73±1.95	68.76±5.42	94.23±5.01
S3	95.37±8.03	97.47±4.39	100.0±0.0
S4	56.17±2.63	83.87±4.05	100.0±0.0
S5	66.40±0.89	75.73±3.49	75.33±7.87
Overall	65.80±2.84	82.33±2.72	93.91±3.68

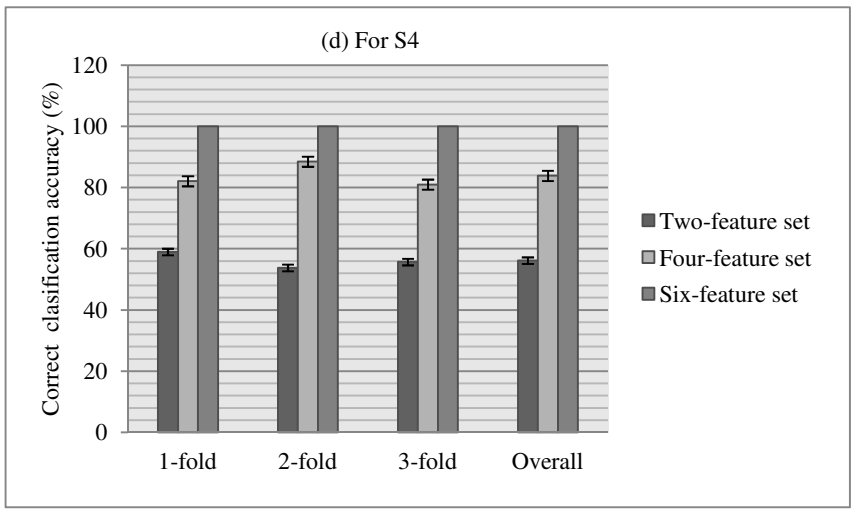
The results confirm us that the six features; *mean*, *standard deviation*, *skewness*, *kurtosis*, *maximum* and *minimum* are potential characteristics to represent the original data for the MI tasks signal classification in BCI applications. Table 6.1 also shows that there are no significant differences of the standard deviation in the three fold accuracies in a subject demonstrating the consistency of the performance.

To provide more detailed information about the classification performance for each of the three feature sets, we show the classification performance for each of the three folds and also overall (average of the three folds) performance for all subjects. Figures 6.9(a)-(e) present the correct classification rate for the two-feature set, the four-feature set and the six-feature set for each of the three folds and also overall performance for all subjects in dataset IVa. From Figures 6.9(a)-9(d), it is seen that the six-feature set produces the highest accuracy among the three feature sets in each of the three folds for S1, S2, S3and S4, respectively. The overall performances are also showing the same results for those subjects. Figure 6.9(e) shows a bit lower performance for the six-feature set in the 2-fold, the 3-fold and overall of S5 than the other two feature sets. Vertical lines on the top of the bar charts show standard error of the three folds and overall. Figures 6.9(a)-6.9(d) also report lower standard errors for the six-feature set in S1, S2, and S3 and S4 which indicate a reliable performance of the method compared to other two features sets. Figure 6.9(e) illustrates a bit higher standard error for the six-feature set in S5 than the other two sets.




(b) For S2
Correct clasification accuracy (%)
120
100
80
60
40
20
0
1-fold
2-fold
3-fold
Overall
Two-feature set
Four-feature set
Six-feature set


(c) For S3
Correct clasification accuracy (%)
120
100
80
60
40
20
0
1-fold
2-fold
3-fold
Overall
Two-feature set
Four-feature set
Six-feature set


(d) For S4
Correct clasification accuracy (%)
120
100
80
60
40
20
0
1-fold
2-fold
3-fold
Overall
Two-feature set
Four-feature set
Six-feature set

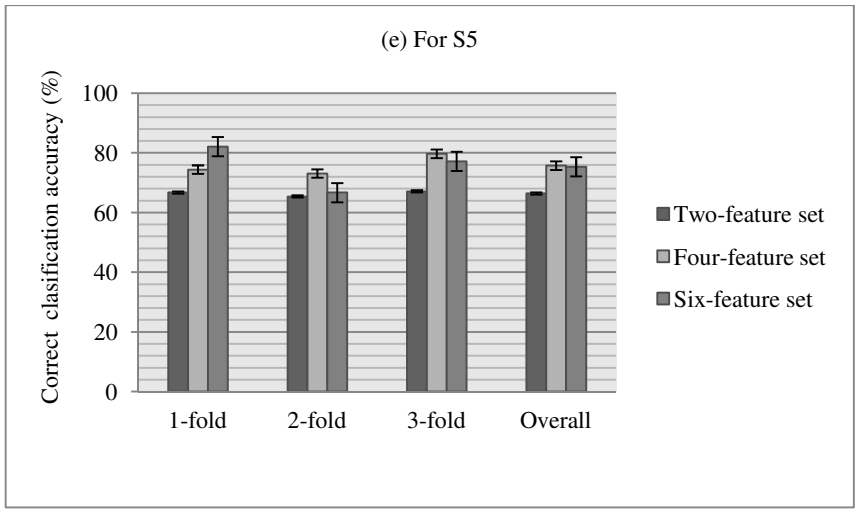


Figure 6.9: Correct classification rate for the two-feature set, the four-feature set and the six-feature set in each of the three folds for: (a) S1 (b) S2 (c) S3 (d) S4 (e) S5 in dataset IVa. Error bars indicate the standard error.

Table 6.2 gives the 3-fold cross-validation accuracy with their standard deviation for each of the three sets of features for dataset IVb. It can be observed from this table that the LR model classifies the LH and RF MI tasks EEG data with the accuracy of 56.57%, 78.33% and 100.0% for the two-feature set, the four-feature set and the six-feature set, respectively. These results demonstrate that the accuracy rate of the four features set increases by 21.76% for including two more features, *skewness* and *kurtosis,* into the two-feature set and the accuracy rate of the six-feature set is improved by 21.67% for adding another two features, *maximum* and *minimum,* into the four-feature set. This table also reports that there is no significant difference of the standard deviation values, which is a good indication of a reliable method. Finally, the experimental outcomes indicate that the proposed algorithm is capable of classifying the MI signals for the six-feature set in BCIs.

Table 6.2: Cross validation results by the proposed method on testing set for dataset IVb.

Feature	3-fold cross-validation accuracy (mean± standard deviation) (%)
Two-feature set	56.57±2.41
Four-feature set	78.33±8.24
Six-feature set	100.0±0.0

Figure 6.10 illustrates the detailed information about the classification rate for the two- feature set, the four-feature set and the six-feature set in each of the three

folds for a subject and also overall rate in dataset IVb. From this figure, it is seen that among the three sets, the six-feature set yields the highest accuracy in each of the three folds for the proposed LR classifier. The figure also shows that standard error is significantly lower in the six-feature set than the two-feature set and the four-feature set; it indicates the consistency of the method.

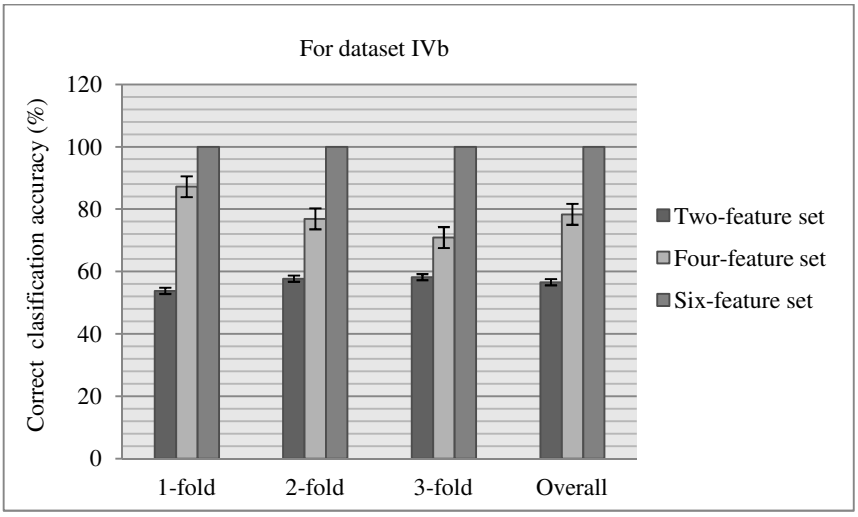


Figure 6.10: Correct classification rate for the two-feature set, the four-feature set and the six-feature set for each of the three folds in dataset IVb. Error bars indicate the standard error.

From the experimental results for both test datasets, it is obvious that the CC technique is capable of feature extraction by using the six mentioned characteristics for the MI tasks data and the LR classifier has the ability to solve a pattern recognition task in BCI applications.

6.3.3 A comparative study

As accuracy is the key criterion for the comparison of different methods in the BCI technology, the classification accuracy of the proposed cross-correlation based logistic regression algorithm is considered as an indicator for the performance evaluation. This section presents a comparative study to compare the performance of our modified CC-LR with BCI III Winner (Blankertz et al., 2006), ISSPL (Wu et al., 2008), CT-LS-SVM (Siuly et al., 2011c), R-CSP with aggregation (Lu et al., 2010), SSRCSP (Lotte and Guan, 2011), TRCSP (Lotte and Guan, 2011), WTRCSP (Lotte

and Guan, 2011) and SRCSP (Lotte and Guan, 2011) for dataset IVa. We could not present the comparison results for dataset IVb as there are no reported research results available. The highest classification accuracy rate among the nine algorithms is highlighted in bold font for each subject and their averages.

Table 6.3 provides a comparative study of the quantitative performance achieved by employing this proposed algorithm, *vs* other recently reported eight well-known algorithms including BCI Competition III Winner for dataset IVa.

Table 6.3: Comparison of the classification performance between our proposed algorithm and the most recent reported eight algorithms for dataset IVa in BCI Competition III.

Method	Classification accuracy rate (%) for dataset IVa					
	S1	S2	S3	S4	S5	Average
Modified CC-LR (Proposed method)	**100**	94.23	**100**	**100**	75.33	93.91
BCI III Winner (Blankertz et al., 2006)	95.5	**100.0**	80.6	**100**	97.6	94.20
ISSPL (Wu et al., 2008)	93.57	**100.0**	79.29	99.64	**98.57**	**94.21**
CT-LS-SVM (Siuly et al., 2011c)	92.63	84.99	90.77	86.50	86.73	88.32
R-CSP with aggregation (Lu et al., 2010)	76.8	98.2	74.5	92.9	77.0	83.90
SSRCSP (Lotte and Guan, 2011)	70.54	96.43	53.57	71.88	75.39	73.56
TRCSP (Lotte and Guan, 2011)	71.43	96.43	63.27	71.88	86.9	77.98
WTRCSP (Lotte and Guan, 2011)	69.64	98.21	54.59	71.88	85.32	75.93
SRCSP (Lotte and Guan, 2011)	72.32	96.43	60.2	77.68	86.51	78.63

In Table 6.3 it is noted that the proposed CC-based LR algorithm provides better classification accuracies than the other eight algorithms in three out of the five subjects. The highest classification accuracy produced by our proposed approach is 100% for S1, S3 and S4**.** With the proposed method, the accuracy rates of 94.23% and 75.33% for S2 and S5 are obtained which are a bit less than the BCI Competition III Winner (Blankertz et al., 2006) and the ISSPL (Wu et al., 2008), while these values are 100% and 98.57%. From the literature review shown in Table 6.3, it is seen that the proposed method has produced the best performance for most of the subjects among the other eight algorithms.

As shown in Table 6.3, the average classification accuracy of the proposed algorithm is 93.91% for the dataset while this value is 94.20% for BCI Competition III Winner, 94.21% for the ISSL, 88.32% for the CT-LS-SVM algorithm, 83.90% for the R-CSP with aggregation, 73.56% for the SSRCSP, 77.98% for the TRCSP, 75.93% for the WTRCSP and 78.63% for the SRCSP. The results demonstrate that the performance of the proposed method is very close to the best results of the BCI Competition III Winner and the ISSL algorithm. Based on these results, it can be

concluded that the modified CC-LR method is better than the recently reported eight methods for the MI tasks EEG signal classification. It is worthy to mention that the training set and testing set of the all methods presented in Table 6.3 are not same due to different methodological process.

6.4 Conclusions and Recommendations

This chapter presents a modified version of the CC-LR algorithm where a cross-correlation (CC) technique is used for feature extraction and a logistic regression (LR) model is applied for the classification of the obtained features. This study investigates what types of features are the best suitable for the representing the distribution of MI EEG signals. The three sets of two-feature, four-feature and six-feature are tested as the input individually to the LR model. The overall classification accuracy for the six-feature set is increased by 28.11% from the two-feature set and 11.58% from the four-feature set for dataset IVa. In dataset IVb, the performance of the proposed algorithm for the six-feature set is improved by 43.43% from the two-feature set and 21.67% from the four-feature set. It is seen from the experimental results that the six-feature set yields the best classification performance for both datasets, IVa and IVb. The performance of the proposed methodology is compared with eight recently reported methods including the BCI Competition III Winner algorithm. The experimental results demonstrate that our present method has improved compared to the existing methods in the literature. The results also report that the CC technique is suitable for the six statistical features, *mean*, *standard deviation*, *skewness*, *kurtosis*, *maximum* and *minimum* representing the distribution of MI tasks EEG data and C3 channel provides better classification results as a reference signal. The LR is an efficient classifier to distinguish the features of the MI data.

CHAPTER 7
LS-SVM WITH TUNING HYPER PARAMETERS: IMPROVING PROSPECTIVE PERFORMANCE IN THE MI TASK RECOGNITION

This study investigates the applications of the least square support vector machine (LS-SVM) classifier with tuning hyper parameters comparing the logistic regression (LR) and kernel logistic regression (KLR) classifiers in the MI task recognition. Chapter 5 and Chapter 6 present the applications of the LR classifier on the cross-correlated features for the classification of a MI tasks based EEG signals in BCI applications. As the LS-SVM is more advance classifier in biomedical research, this chapter proposes a novel algorithm based on the LS-SVM with tuning hyper parameters to distinguish the MI signals for the improvement of the classification performance. The proposed scheme denoted as CC-LS-SVM implements the LS-SVM on cross-correlation features for the two-class MI signal recognition. To compare the effectiveness of the proposed classifier, we replace the LS-SVM classifier by the LR and the KLR classifiers, separately, for the same features extracted from the CC technique. In Chapter 3 and Chapter 4, a LS-SVM classifier is used for classification of EEG signals but the hyper parameters of this classifier are selected manually rather than using an appropriate method although hyper parameters of the LS-SVM play an important role in the classification. This chapter applies a two-step grid search process to select optimal parameters of the LS-SVM which can provide more reliable and consistent results.

In this chapter, we also investigate the effects of two different reference channels, the electrode position Fp1 and C3 channels. The present approach is tested on the same datasets as Chapters 5 and 6, and the performances are evaluated with classification accuracy through a 10-fold cross-validation procedure. After the experimental investigation, finally we conclude on which classifier is better for cross-correlation features to classify the MI signals in BCI applications. The method

reported in this chapter can be useful to assist clinical diagnoses and rehabilitation tasks.

The content of this chapter has been published in the *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* (Siuly and Yan Li, 2012) and also in *the 2011 IEEE International Conference on Complex Medical Engineering (CME 2011)* (Siuly, et al., 2011a).

7.1 Introduction

The translation of brain activities into control signals in BCI systems requires a robust and accurate classification of the various types of information. BCIs convert *human intentions* or *thoughts* into control signals to establish a direct communication channel between the human brain and output devices, as presented in Figure 7.1. This figure depicts the basic structure of BCI technologies, on how a brain signal of a thought is passing on to a BCI system, and how the BCI system processes those signals into a control signal for a user application.

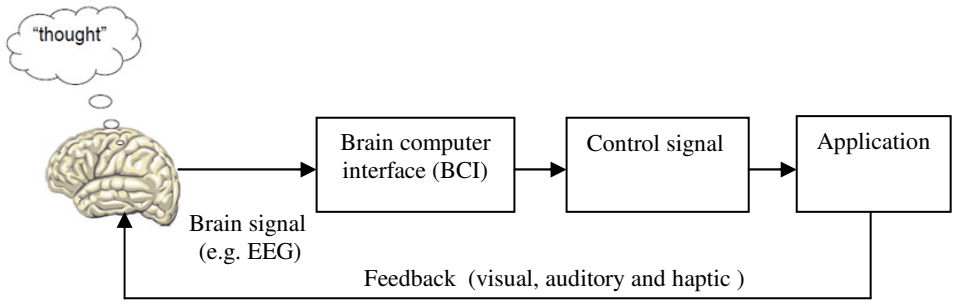


Figure 7.1: Fundamental structure of brain computer interface (BCI).

As shown in Figure 7.1, a BCI system works through EEG brain signals. A big challenge, therefore, is for BCI systems to correctly and efficiently identify different EEG signals of different MI tasks using appropriate classification algorithms. In most current MI based BCIs, machine learning algorithms are carried out in two stages: feature extraction and feature classification (Wu et al., 2008). Because of many factors, such as low topographical resolutions and high noise levels, it has been a key issue to extract effective features from the EEG signals and perform the classification (Long et al., 2010). Although many methods have been

reported for feature extraction as well as classification to produce impressive results in BCI applications as discussed in Section 7.2, nevertheless a MI-based BCI system is not satisfactory due to the lack of classification accuracy.

The goal of the study is to improve the classification accuracy of MI data for BCI systems and also to investigate whether the LS-SVM with tunning hyper parameters is better as a classifier than a logistic regression or kernel logistic regression classifier with cross-correlated features in an EEG-based MI classification. For this purpose, the present study proposes a novel algorithm, the CC-LS-SVM, where a CC technique is developed for feature extraction and a LS-SVM is employed for classifying the obtained features. The performance of the LS-SVM classifier is compared with a logistic regression classifier and a kernel logistic regression classifier for the same feature vector set. In order to further verify the effectiveness of the proposed CC-LS-SVM algorithm, we also compare it with the eight most recently reported methods in the literature.

In this research, there are strong grounds of using a CC technique for feature extraction from MI EEG data explained in Section 5.3.1 of Chapter 5. In order to reduce the dimensionality of the cross-correlation sequences, six statistical features, *mean, median, mode, standard deviation, maximum* and *minimum* values, are extracted from each cross-correlation sequence as discussed in detail in Section 5.3.1 of Chapter 5. These features represent the characteristics of the original MI EEG data without redundancy. The extracted features are then used as the inputs to the LS-SVM and also to the logistic regression and the kernel logistic regression classifiers.

The LS-SVM is a robust intelligent technique for classification in BCI applications. It has the advantage over other techniques of converging to a global optimum, not to a local optimum that depends on the initialization or parameters affecting the rate of convergence. The computation of the LS-SVM is faster compared with other machine learning techniques because there are fewer random parameters and only the support vectors are used in the generalization process (Esen et al. 2009). In spite of its advantages, this method has not been applied for the MI tasks classification in the BCI development except the clustering technique based algorithm (Siuly et al., 2011c). But, in the algorithm (Siuly et al., 2011c), the hyper parameters of the LS-SVM were not selected optimally through a technique. It is well known that the parameters of the LS-SVM play an important role in affecting

the classification performance. To obtain more reliable results, this study, instead of manual selection, employs a LS-SVM for the MI EEG signal classification where the hyper parameters of the LS-SVM are chosen optimally using a two-step grid search algorithm.

The proposed approach in this chapter is evaluated on datasets, IVa and IVb of BCI Competition III, where both sets contain MI EEG recorded data. In this paper, a 10-fold cross-validation method is used for assessing the performance of the proposed method. This procedure divides the feature vector sets into ten approximately equal-sized distinct partitions. One partition is then used for testing, whilst other partitions are used for training the model. To further improve the estimate, the procedure is repeated ten times and all accuracy rates over these ten runs are averaged. The average accuracy over the ten runs obtained from the test data is taken as the performance evaluation criteria in this study. Experimental results show that the proposed LS-SVM classifier achieves a better performance compared to the logistic regression and kernel logistic regression for the same cross-correlated features. The results also demonstrate that the proposed approach do better than the other eight most recently reported methods with respect to the classification performance for dataset IVa.

7.2 Review of the existing classification techniques

A variety of methods have been reported by different researchers using dataset IVa of BCI Competition III. A brief discussion of some recent work is provided below. So far there is no classification results reported for dataset IVb.

Yong et al. (2008) reported a sparse spatial filter optimization for EEG channel reduction in the brain computer interface, where the spatial filter was used to project the signals and the variance of the projected signals was the only feature used in the linear discriminant analysis (LDA) as the input for the classification. They reported a classification accuracy of 57.5% for subject **aa,** 86.9% for subject **al,** 54.4% for subject **av,** 84.4% for subject **aw** and 84.3% for subject **ay,** and the average accuracy was 73.5% using a 10-fold cross validation. But the major limitation of that study was that they manually selected their regularization parameter.

Lu et al. (2009) introduced a regularized common spatial patterns (R-CSP) algorithm by incorporating the principle of generating learning for EEG signal classification. That study used two regularization parameters in regularizing the covariance estimates and these parameters were not selected optimally through a technique. The reported classification accuracy rates were 69.6%, 83.9%, 64.3%, 70.5%, 82.5% for subject **aa, al, av, aw, ay,** respectively. They obtained an average accuracy rate of 74.2% for all subjects. It was reported that the algorithm was particularly effective in small sample settings.

Lotte et al. (2011) proposed four methods representing a family of a theoretical framework based on regularized common spatial patterns (RCSP). Their proposed methods are regularized CSP with selected subjects (SSRCSP) (Lotte et al., 2011), CSP with Tikhonov regularization (TRCSP) (Lotte et al., 2011), CSP with weighted Tikhonov regularization (WTRCSP) (Lotte et al., 2011) and spatially regularization (SRCSP) (Lotte et al., 2011). It was reported that their methods can perform efficiently subject-to-subject transfer for classifying MI data in BCIs. In particular, the TRCSP and WTRCSP algorithms are better than the other two algorithms. The average classification success rate reached at 73.56% for the SSRCSP, 77.98% for the TRCSP, 75.93% for the WTRCSP and 78.63% for the SRCSP. All their four algorithms are based on a common spatial patterns (CSP) method. Although the CSP is a popular method in BCI applications, it is very sensitive to noise, and often over-fits with small training sets.

Lu et al. (2010) introduced a regularization and aggregation technique with CSP for EEG signal classification in a small sample setting (SSS). To tackle the problem of regularization parameter determination, a number of R-CSPs were aggregated to give an ensemble-based solution. The parameter determination problem was solved through a cross-validation procedure. The cross-validation method was employed to determine the regularization parameters of the R-CSP for the EEG signal classification in SSS. The obtained classification accuracy rates were 76.8%, 98.2%, 74.5%, 92.9% and 77.0% for subject **aa, al, av, aw, ay,** respectively, for experiment III. The overall accuracy performance was 83.9%.

Most recently, Siuly et al. (2011c) reported a clustering technique-based least square support vector machine (CT-LS-SVM) algorithm for EEG signal classification (see Chapter 4). They developed a CT approach for feature extraction

and the obtained features were used to the LS-SVM as the inputs for classification. It employed the 10-fold cross-validation method to evaluate the performance and achieved the classification accuracy of 92.63% for subject **aa,** 84.99% for subject **al,** 90.77% for subject **av,** 86.50% for subject **aw** and 86.73% for subject **ay.** The average accuracy performance was 88.32%. The weakness of that method was that they did not select the parameters optimally through any technique. They manually selected the parameters for the LS-SVM method.

From the discussion of the literature, it is observed that most of the reported methods are limited in their success and effective only in a small sample setting. In most of the cases, the methods did not select their parameters using a suitable technique while the parameters significantly affect the classification performance. To overcome these problems, this chapter presents a new approach, CC-LS-SVM which can discriminate two-class MI tasks for the development of BCI systems. In the proposed algorithm, a CC technique is developed for feature extraction and a LS-SVM is applied to classify the obtained features. To the best of our knowledge, the cross-correlation technique and the LS-SVM have not been used together before for the MI task recognition in BCI applications. This study employs a two-step grid search algorithm for selecting the optimal combinations of the parameters for the LS-SVM.

7.3 Proposed Method

The present study develops an algorithm that can automatically classify two categories of MI EEG signals in BCI systems. The proposed cross-correlation-based LS-SVM scheme for the MI signals classification is illustrated in Figure 7.2. The approach employs a CC technique to extract representative features from the original signals, and then the extracted features are used as the inputs to the LS-SVM classifier.

In order to evaluate the performance of the LS-SVM classifier, we test a logistic regression classifier and a kernel logistic regression classifier, separately. They also employ the same features extracted from the cross-correlation method as the inputs.

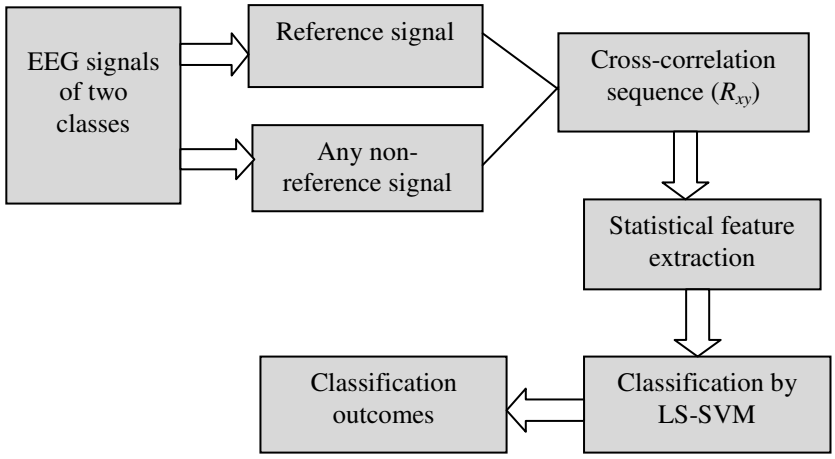


Figure 7.2: Block diagram of the proposed CC-LS-SVM technique for the MI EEG signal classification in BCIs development.

The block diagram of the proposed method in Figure 7.2 depicts the procedure for the MI EEG signal classification as described in the following steps:

7.3.1 Reference signal selection

One signal is selected as a reference signal among all channel signals in one subject of the two-class MI tasks. A reference signal should be noiseless as the signal with the noise will be incoherent with anything in the reference. In this work, any other signal, that is not a reference signal, is treated as a non-reference signal.

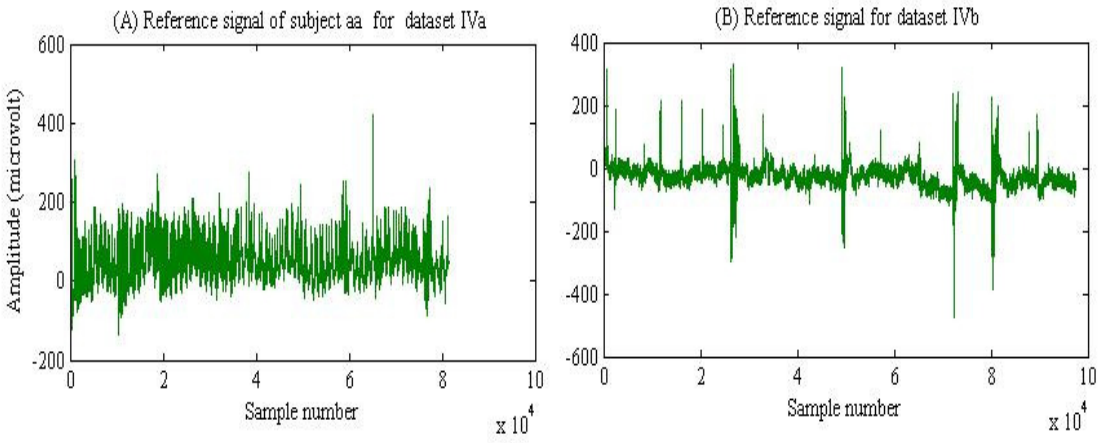


Figure 7.3: Typical reference signals: (A) Dataset IVa and (B) Dataset IVb.

In this study, we use two datasets, IVa and IVb of BCI Competition III. Dataset IVa consists of MI tasks EEG signals from the right hand class and the right foot class. Dataset IVb consists of MI tasks EEG signals of the left hand class and the right foot class. Both datasets contain 118 channel data in each class of a subject. we consider the electrode position Fp1 in the international 10/20 system as the reference signal for the cross-correlation technique for all the subjects. For dataset IVa, the Fp1 is selected from the right hand class while it is from the right foot class for dataset IVb. Figures 7.3(A) and 7.3(B) show the typical reference signals of subject **aa** for dataset IVa and dataset IVb, respectively.

7.3.2 Computation of a cross-correlation sequence

A cross-correlation sequence, denoted by 'R_{xy}', is calculated recursively using a reference signal and any other non-reference signal using the cross-correlation technique as shown in Figure 7.2. In Chapter 5, Equation (5.2) of the cross-correlation method (Dutta et al., 2010; Chandaka et al., 2009a) is used to compute a cross-correlation sequence. The graphical presentation of a cross-correlation sequence is called a cross-correlogram. The reference signal of a class is cross-correlated with the data of the remaining signals of this class and the data of all signals of another class. If we have two classes of EEG signals, and class 1 has n signals and class 2 has m signals, and a reference signal is chosen from class 1, then a total of (n-1) cross-correlation sequences are obtained from class 1 and a total of m cross-correlation sequences from class 2.

As there are 118 signals in each of the two classes of a subject in datasets, IVa and IVb, in each subject of both datasets, the reference signal is cross-correlated with the data from the remaining 117 signals of the reference signal class. This reference signal is also cross-correlated with the data of all 118 signals of the non-reference signal class. Thus, for each subject, a total of 117 cross-correlation sequences/cross-correlograms are obtained from the reference signal class and 118 from the non-reference signal class. For example, in subject **aa**, the signal of the Fp1 channel is a reference signal, which comes from the right hand class and this reference signal is cross-correlated with the data from the remaining 117 signals of the right hand class. In the right foot class of subject **aa**, this reference signal is also

cross-correlated with the data of all 118 signals of this class. Thus, for subject **aa**, a total of 117 cross-correlation sequences/cross-correlograms are obtained from the right hand class and 118 from the right foot class. The same process is followed for subjects **al**, **av**, **aw** and **ay** and the subject of dataset IVb in this study.

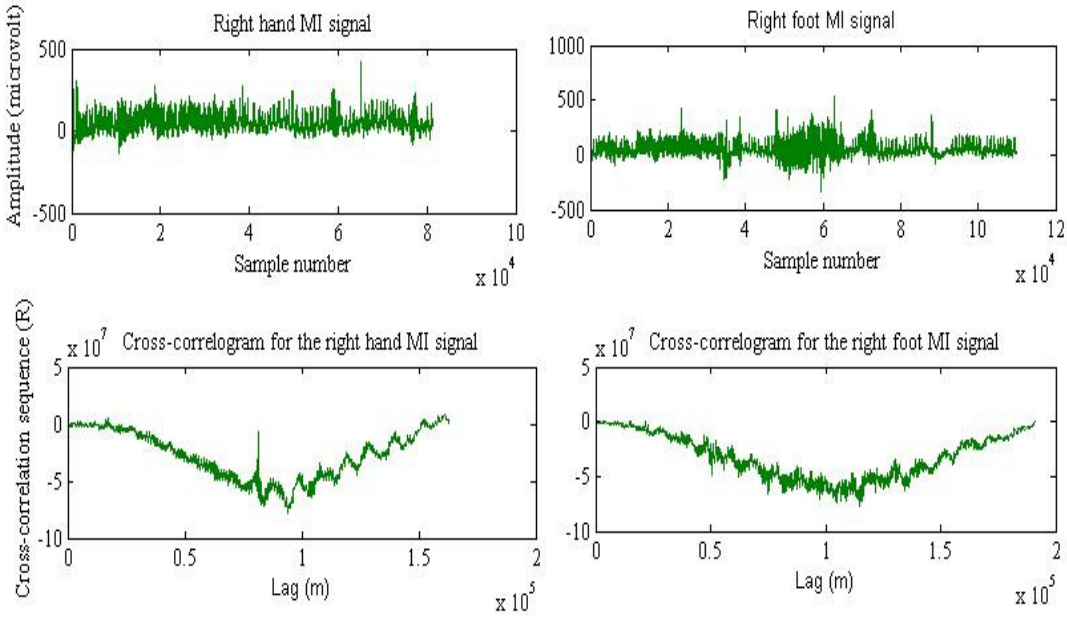


Figure 7.4: Typical right hand and right foot MI signals and their respective cross-correlograms for subject **aa** in dataset IVa.

Figure 7.4 presents typical signals of the right hand and the right foot MI data for subject **aa** of dataset IVa. The typical cross-correlograms for the right hand and the right foot MI signals of the same subject are also shown in Figure 7.4. The cross-correlogram of the right hand signal is obtained using the reference signal and the right hand MI signal and the cross-correlogram of the right foot signal is acquired using the reference signal and the right foot MI signal as depicted in Figure 7.4.

Figure 7.5 shows typical signals of dataset IVb for the right foot MI and the left hand MI. This figure also presents typical results of the cross-correlation for the right foot MI signal and the left hand MI signal. As shown in Figure 7.5, the cross-correlogram of the right foot MI signal is obtained using the reference signal and the right foot MI signal and the left hand cross-correlogram is generated by the reference signal and the left hand MI signal.

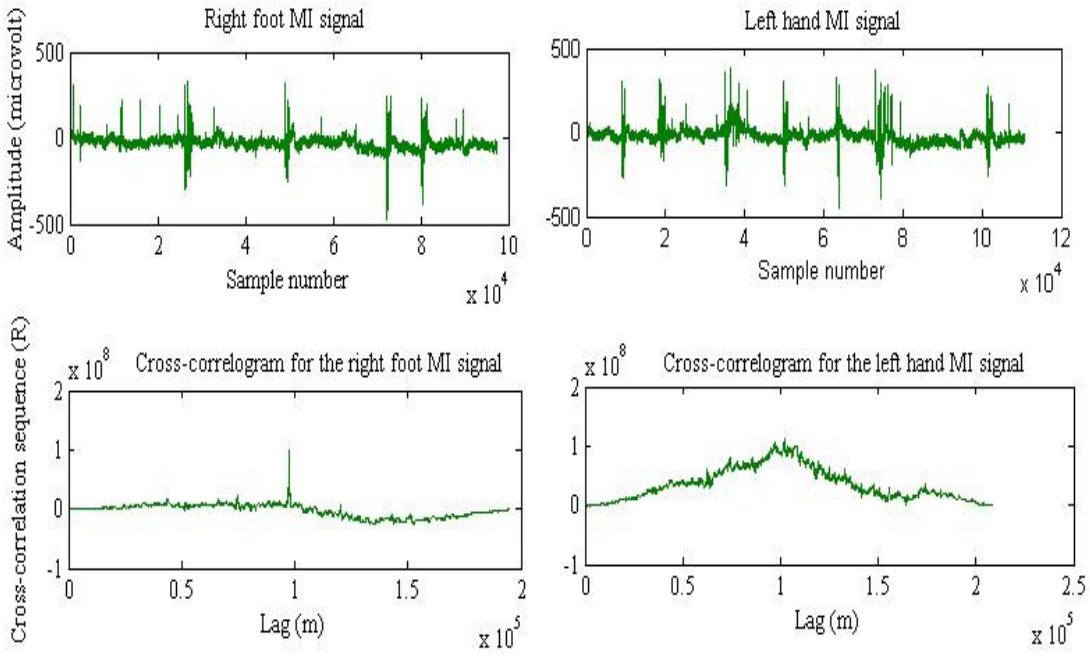


Figure 7.5: Typical right foot and left hand MI signals and their respective cross-correlograms for dataset IVb.

It is known that if two curves have exactly the same shape, this means, they are highly cross-correlated with each other and cross-correlation is around 1. From Figures 7.4 and 7.5, one can see that the shapes of the two curves are not exactly same, which indicates the statistical independency. That means, there is more of a chance to achieve better separation.

As mentioned in Chapter 5, if each of x and y signals has a finite number of samples N, the resulting cross-correlation sequence has $(2N\text{-}1)$ samples. Hence in each of Figures 7.3, 7.4 and 7.5, the scale for each signal is shown over the range of 0 to 10×10^4 samples and the scale for the corresponding cross-correlogram is shown over the range of 0 to 2×10^5 samples. The cross-correlogram signals convey greater signal information and consist of low level noises compared to the original signal. It is worthy to mention that a cross-correlogram contains information about the frequencies which are common to both waveforms, one of which is usually the signal and the other a reference wave (Dutta et al., 2009). Six statistical features, *mean, median, mode, standard deviation, maximum* and *minimum* are extracted from each cross-corrologram as discussed in the following section.

7.3.3 Statistical feature extraction

To reduce the dimensions of the cross-correlation sequences, this study considers six statistical features, *mean, median, mode, standard deviation, maximum* and *minimum* as the representatives ideally containing all important information of the original signal patterns like Chapter 5. These features are calculated from each cross-correlation sequence or cross-correlogram to create feature vector sets. The six traits of the cross-correlation sequences are found to serve as important indicators of the neurological state of the subjects (Hieftje et al., 1973; Wren et al., 2006). The reasons of choosing theses feature sets are described in detail in Section 5.31 of Chapter 5.

In this study, we obtain 117 cross-correlation sequences from the reference signal class and 118 from the non-reference signal class for a subject in both datasets. We calculate the mentioned six features from each cross-correlation sequence. For example, subject **aa** contains 117 cross-correlation sequences for the right hand class (reference signal class) and 118 cross-correlation sequences for the right foot class (non-reference signal class). As we calculate the six features from each cross-correlation sequence, so we obtain 117 feature vectors of six dimensions from the right hand class and 118 features vectors of the same dimensions from the right foot class for subject **aa**. Thus we acquire a total of 235 feature vectors with six dimensions for this subject. We follow the same process for the other subjects in both datasets. We use MATLAB 'mean', 'median', 'std', 'max', 'min' function for calculating *mean*, *median*, *standard deviation*, *maximum* and *minimum* values, respectively, from each cross-correlation sequence. In the *mode* calculation, we compute a histogram from a cross-correlation sequence and then the peak of the histogram is considered as an estimate of the *mode* for that cross-correlation sequence. These feature vector sets are divided into a training set and a testing set using a 10-fold cross validation method, which is discussed in Section 7.3. These feature vectors are inputs for the LS-SVM and also for the logistic regression and the kernel logistic regression classifiers in the classification stage.

7.3.4 Classification

This study employs the LS-SVM with radial basis function (RBF) kernel as a

classifier to distinguish the features obtained from the cross-correlation technique. The decision function of the LS-SVM in Equation (3.8) is derived directly from solving a set of linear equations (Thissen et al., 2004; Suykens et al., 2002; Siuly et al. 2009). A detailed description of the LS-SVM algorithm could be found in Chapter 3.

In this study, the obtained training feature vector of six dimensions are used as the input in Equation (3.8) to train the LS-SVM classifier and the testing feature vector sets are employed to verify the performance and the effectiveness of the trained LS-SVM for the classification of two-class of EEG signals in the both datasets. For dataset IVa, y_i is treated as right foot=+1and right hand= -1 and for the dataset IVb, y_i is considered as right foot= +1 and left hand= -1. As the result of the LS-SVM lies largely on the choice of a kernel, the RBF kernel is chosen after many trials. The two important parameters (γ, σ^2) of the LS-SVM are selected by a two-step grid search technique for getting reliable performance of the method, that is discussed in Section 7.4.1. The solution of Equation (3.8) provides the prediction results that directly assign the samples with a label +1 or -1 to identify that which category it belongs to.

To compare the performance of the proposed LS-SVM classifier, we employ the logistic regression classifier instead of the LS-SVM for the same feature sets as its inputs. In Chapter 5, the logistic regression (Caesarendra et al., 2010; Hosmer and Lemeshow, 1989) in Equation (5.2) is applied for the classification of the MI features. The details description of the logistic regression is available in Chapter 5.

In this work, we consider the mentioned six features of a feature vector set (training/testing) as the six input variables (x_1=*mean* values, x_2=*maximum* values, x_3=*minimum* values, x_4=*standard deviation* values, x_5=*median* values and x_6=*mode* values.) in Equation (5.2) for the both dataset set. We treat the dependent variable y as right hand= 0 and right foot= 1 for dataset IVa and right foot=0 and left hand=1for dataset IVb. Finally we obtain the prediction results that directly provide the class label 0 or 1 with the samples.

We also compare the performance of the LS-SVM with a kernel logistic regression classifier. Kernel logistic regression is a non-linear form of logistic regression. It can be achieved via the so-called “kernel trick” which has the ability to classify data with non-linear boundaries and also can accommodate data with very

high dimensions. A detailed description of the kernel logistic regression is available in (Cawley and Talbot, 2008; Rahayu et al., 2009). A final solution of the kernel logistic regression could be achieved using the following equation.

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x) + b \tag{7.1}$$

where x_i represents ith input feature vector of d dimensions, n is the number of feature vectors and b is the model parameter. The vector α_i contains the parameters which define decision boundaries in the kernel space and $K(x_i, x)$ is a kernel function. The most commonly used kernel in practical applications is the radial basis function (RBF) kernel defined as $K(x_i, x\ ) = \exp\left(-(\|x_i - x\|)^2 / 2\sigma^2\right)$ which is also used in this study. Here σ is a kernel parameter controlling the sensitivity of the kernel. The parameters of this method are automatically estimated by iteratively re-weighted least square procedure (Rahayu et al., 2009).

In the kernel logistic regression, we utilize the feature vectors and their class labels as the same process of the logistic regression for the inputs. Finally we acquire the output of the kernel logistic regression as an estimate of a posterior probability of the class membership. In Section 7.4, the classification results of these three classifiers are presented for datasets IVa and IVb. The following section discusses how the performance of the proposed algorithm is evaluated through a 10-fold cross-validation procedure in this paper.

7.4 Performance Evaluation

The classification accuracy has been one of the main pitfalls in the developed BCI systems. It directly affects the decision made in a BCI output. The classification accuracy from an experiment is calculated by dividing the number of correctly classified samples by the total number of samples (Siuly et al., 2011d). In this study, a *k*-fold cross-validation (Abdulkadir, 2009; Ryali et al., 2011c) is used to calculate classification accuracy for assessing the performance of the proposed method.

In *k*-fold cross-validation procedure, a data set is partitioned into *k* mutually exclusive subsets of approximately equal size and the method is repeated *k* times (folds). Each time, one of the subsets is used as a test set and the other *k-1* subsets are

put together to form a training set. Then the average accuracy across all *k* trials is computed.

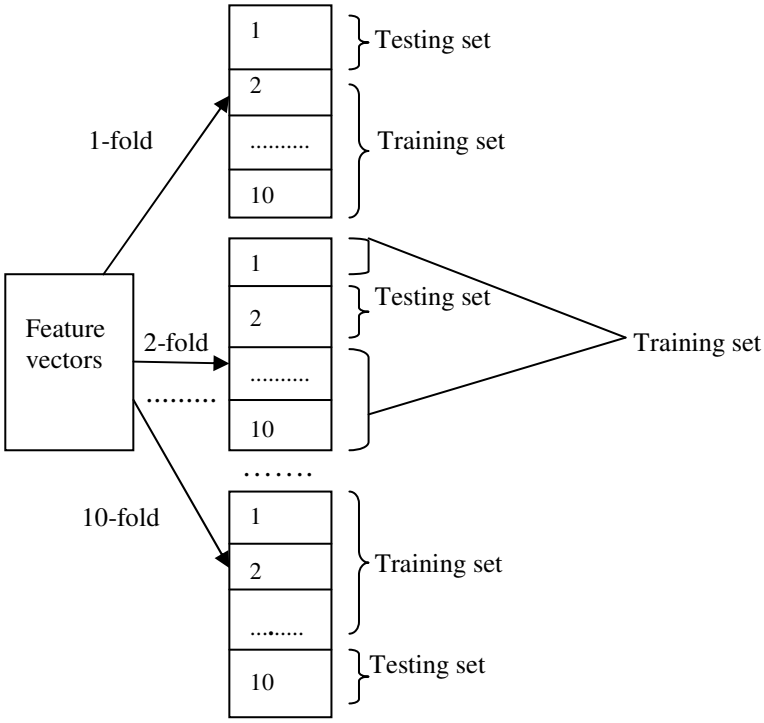


Figure 7.6: Partitioning design of the obtained feature vectors for the 10-fold cross-validation method.

In this study, we select k=10 as it is a common choice for the k-fold cross-validation. As mentioned in Section 7.2.3, we obtain a total of 235 feature vectors of six dimensions from the two-class MI EEG signals of a subject in each of the two datasets, IVa and IVb. Figure 7.6 presents a design of how the extracted feature vectors of this study are partitioned into 10 mutually exclusive subsets according to the k-fold cross-validation system. As shown in Figure 7.6, the feature vector set of each subject is divided into 10 subsets and the procedure is repeated 10 times (the folds). Each time, one subset is used as a testing set and the remaining nine subsets are used as a training set, which is illustrated in Figure 7.6. The obtained classification accuracy of each of 10 times on the testing set is averaged called '10-fold cross-validation accuracy' in this paper.

7.5 Experiments and Results

Before classification, the hyper parameters of the LS-SVM classifier is tuned by a

two-step grid search algorithm discussed in Section 7.4.1 as the classification performance of the LS-SVM depends on the parameters and the values chosen of the parameters significantly affect the classification accuracy. Section 7.4.2 discusses how the variables are set up in the logistic regression classifier and the kernel logistic regression classifier. The results obtained by the LS-SVM, the logistic regression and the kernel logistic regression for the cross-correlation based features are compared to each other for datasets IVa and IVb in Section 7.4.3. Section 7.4.4 presents a comparative study for our proposed method with eight existing methods in the literature. In this research, the classification by the LS-SVM is carried out in MATLAB (version 7.7, R2008b) using the LS-SVMlab toolbox (version 1.5) (LS-SVMlab toolbox (version 1.5)-online) and the classification of the logistic regression and kernel logistic regression are executed through MATLABArsenal (MATLAB Classification Wrapper 1.00 (Debug version)) package (MATLABArsenal-online).

In this study, the training set is applied to train the classifier and the testing vectors are used to verify the accuracy and the effectiveness of the classifiers for the classification of the two-class MI data. Our proposed algorithm is employed on each subject of both datasets separately, as the MI EEG signals are naturally highly subject-specific depending on physical and mental tasks. All experimental results are presented based on the testing set in this paper.

7.5.1 Tuning the hyper parameters of the LS-SVM classifier

To improve the generalization performance of the LS-SVM classifier, it is needed to select its two parameters (γ, σ^2) through an appropriate procedure. These parameters play an important role in the classification performance. The regularization parameter γ (gamma) determines tradeoff between minimizing the training error and minimizing the model complexity. The parameter σ^2 (sig2) is the bandwidth and implicitly defines the nonlinear mapping from the input space to a high dimensional feature space. Large values of γ and σ^2 may lead to an over-fitting problem for the training data (Chandaka et al., 2009a; Suykens and Vandewalle, 1999), so the values must be chosen carefully. This study applies a two-step grid search technique in order to obtain the optimum values of the hyper parameters for the LS-SVM. This section describes the process on how the parameters are selected

through a two-step grid search algorithm. A grid search is a two dimensional minimization procedure based on an exhaustive search in a limited range (Xie et al., 2009). It tries values of each parameter across a specified search range using geometric steps. In each iteration, one leaves a point, and fits a model on the other data points. The performance of the model is estimated based on the one-point-left-out. This procedure is repeated for each data point. Finally, all the different estimates of the performance are combined. The two-step grid search procedure is provided in the free LS-SVM toolbox (LS-SVMlab toolbox (version 1.5)-online), to develop the LS-SVM model (Li et al., 2008).

In this research, the two-step grid search method is applied in each of the 10 folds of a subject of the both datasets for selecting the optimal parameter values of the LS-SVM. The obtained values of the parameters for each fold are used in the LS-SVM algorithm to obtain the reliable performance of the proposed method. Figures 7.7 (A) and 7.7 (B) show the process of the two-step grid search for optimizing the parameters γ (gamma) and σ^2 (sig2) of the LS-SVM classifier for dataset IVa (for the 1-fold of subject **aa**) and dataset IVb (for the 1-fold), respectively.

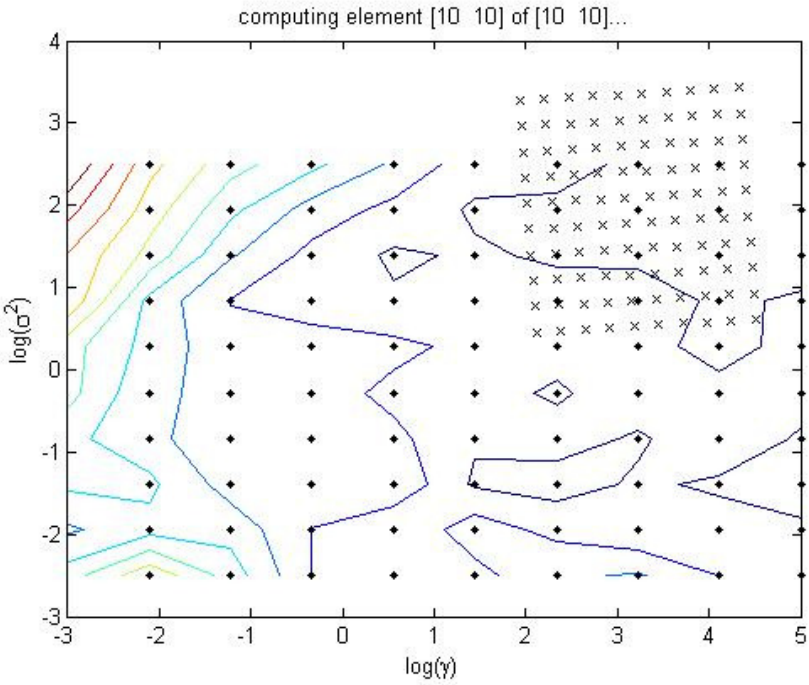


Figure 7.7 (A): Process of the two-step grid search for optimizing the parameters γ (gamma) and σ^2 (sig2) of the LS-SVM classifier in the 1-fold of subject **aa** of dataset IVa.

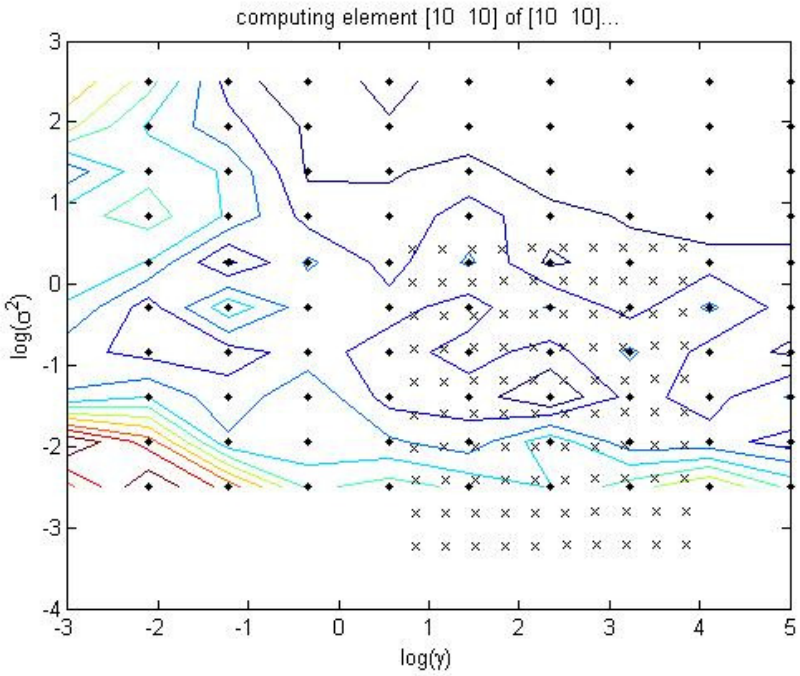


Figure 7.7 (B): Process of the two-step grid search for optimizing the parameters γ (gamma) and σ^2 (sig2) of the LS-SVM classifier in the 1-fold of dataset IVb.

Firstly, the optimal range of the parameters is determined in the first step of a grid search. The grids denoted as " ◆" in the first step is 10×10, and the searching step is for a crude search with a large step size. The optimal search area is determined by the error contour line.The grids denoted as "×" in the second step is also 10×10, and the searching step is the specified search with a small step size. The contour lines indicate the value levels of the cost function in the grid search. Using this method, the obtained optimal combinations of γ and σ^2 for the LS-SVM is presented in Table 7.1 for dataset IVa and in Table 7.2 for dataset IVb.

Table 7.1: Optimal values of the parameters γ and σ^2 of the LS-SVM for dataset IVa.

	Obtained optimal parameter values of γ and σ^2 of the LS-SVM									
Subject	**aa**		**al**		**av**		**aw**		**ay**	
Parameters	γ	σ^2	γ	σ^2	γ	σ^2	γ	σ^2	γ	σ^2
1-fold	82.13	8.84	65.89	5.52	22.16	5.06	260.30	11.74	58.12	4.69
2-fold	31.59	7.04	60.05	8.32	220.64	7.88	181.25	11.19	72.15	1.53
3-fold	202.90	16.09	245.97	2.24	7.33	6.46	401.84	14.81	45.85	1.88
4-fold	128.22	18.62	80.27	5.66	921.84	1.78	343.21	12.92	92.86	1.05
5-fold	30.79	9.48	315.18	14.67	4.94	8.66	200.97	0.79	42.59	0.65
6-fold	58.76	20.79	632.36	13.27	10.11	3.92	141.36	12.81	78.74	1.39
7-fold	46.71	7.73	261.91	11.72	719.69	5.42	193.22	8.44	26.18	0.73
8-fold	29.75	10.85	64.29	2.21	9.23	3.99	179.95	13.72	149.09	0.84
9-fold	208.08	32.27	49.22	1.65	10.25	4.04	75.48	7.41	246.47	1.52
10-fold	29.49	10.57	79.59	1.99	16.63	5.25	605.63	13.71	49.67	1.35

Note: up to two digits decimal considered.

Table 7.2: Optimal values of the parameters γ and σ^2 of the LS-SVM for dataset IVb.

	Obtained optimal parameter values γ and σ^2 of the LS-SVM	
Parameters	γ	σ^2
1-fold	7.0451	5.3714
2-fold	6.9404	1.5538
3-fold	47.7992	3.7401
4-fold	107.1772	1.7566
5-fold	10.0366	1.8417
6-fold	320.4905	3.3605
7-fold	57.1816	2.7994
8-fold	820.5462	1.3092
9-fold	569.3277	2.1852
10-fold	31.9349	1.8465

As shown in Table 7.1 and Table 7.2, the optimal values of hyper parameters for the LS-SVM are obtained in each of 10-folds for each subject of the two datasets through the two-step grid search algorithm. In this study, the classification results of each fold are achieved using the optimal parameter values in each subject of the both datasets.

7.5.2 Variable selections in the logistic regression and kernel logistic regression classifiers

Although the parameters of the logistic regression are obtained automatically through maximum likelihood estimation (MLE) method, the variable selections are an important task for the logistic regression model. In this research, the logistic regression presented in Equation (5.2) is used to estimate the probability of the dependent variable using independent variables as the input. For each of the two datasets, we consider the MI tasks as a dependent variable, termed *y,* and the six statistical features are treated as six independent variables. The six independent variables used in Equation (5.2) are x_1=*mean* values, x_2=*maximum* values, x_3=*minimum* values, x_4=*standard deviation* values, x_5=*median* values and x_6=*mode* values. It is known that the dependent variable *y* has two values, 0 and 1, in the logistic regression. For dataset IVa, the right hand MI class is treated as 0 and the right foot MI class as 1. For dataset IVb, we denote the right foot MI class as 0 and the left hand MI class as 1. In the kernel logistic regression, the model parameters in Equation (7.1) are automatically anticipated by the iteratively re-weighted least square procedure (Rahayu et al., 2009). The feature vectors and class labels of the

kernel logistic regression in Equation (7.1) are considered as the same as those in the logistic regression.

7.5.3 Performances on both datasets

Table 7.3 presents the classification results for the LS-SVM, the logistic regression and the kernel logistic regression classifiers for the five subjects of dataset IVa. In Table 7.3, the results of each subject are reported in terms of mean ± standard deviation of the accuracy over a 10-fold cross-validation method on the testing set. It is observed from Table 7.3 that the proposed LS-SVM classifier for the cross-correlation features produces an accuracy of 97.88% for subject **aa**, 99.17% for subject **al**, 98.75% for subject **av**, 93.43% for subject **aw** and 89.36% for subject **ay**; while these values are 95.31%, 87.26%, 94.89%, 94.93%, 75.33%, respectively, for the logistic regression classifier; and 97.03%, 96.20%, 95.74%, 94.51%, 83.42%, respectively, for the kernel logistic regression with the same features. Based on the experimental results, the classification success rates of the proposed LS-SVM classifier are higher than those of the logistic regression and the kernel logistic regression in four out of five subjects.

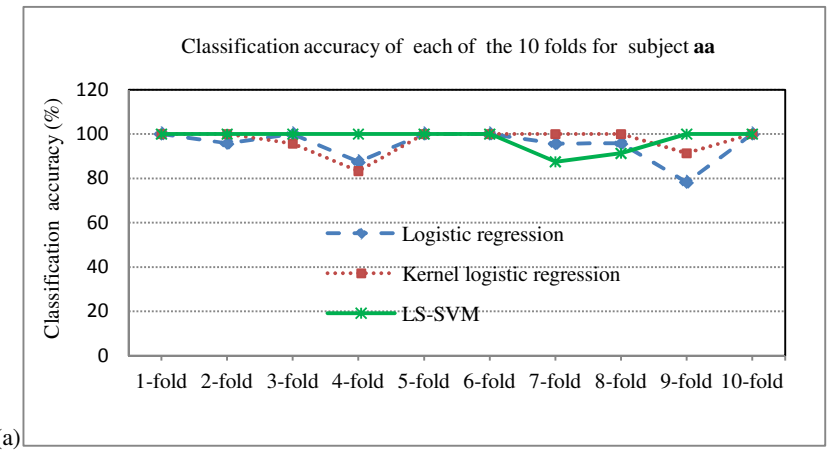
Table7. 3: Classification results by the 10-fold cross-validation method on testing set of dataset IVa.

Subject	10-fold cross-validation accuracy (%) (mean ± standard deviation)		
	LS-SVM	Logistic regression	Kernel logistic regression
aa	97.88±4.56	95.31±7.17	97.03±5.62
al	99.17±1.76	87.26±10.07	96.20±4.99
av	98.75±2.81	94.89±7.77	95.74±7.32
aw	93.43±6.87	94.93±5.14	94.51±4.88
ay	89.36±5.74	75.33±12.92	83.42±10.97
Average	95.72±4.35	89.54±8.61	93.38±6.76

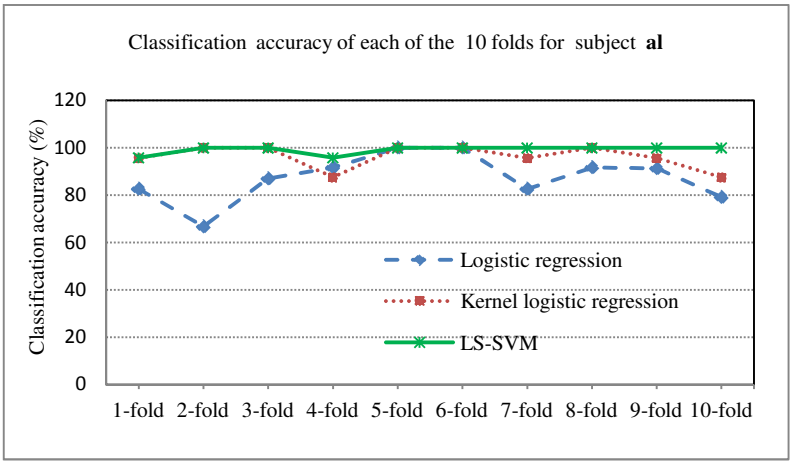
Table 7.3 also reports that the standard deviations for the proposed approach are much lower compared to those of the logistic regression and the kernel logistic regression in those four subjects. The lower values of standard deviation indicate the consistency of the proposed method. As seen from Table 7.3, the proposed LS-SVM provides the best results with an average classification accuracy of 95.72% whereas this value is 89.54% for the logistic regression and 93.38% for the kernel logistic regression classifier. The average classification accuracy for the proposed method

increases by 6.18% in comparison to the logistic regression model and 2.34% to the kernel logistic regression.

In what follows, we provide the details on how the 10-fold cross-validation system produces the classification accuracy in each of the 10-folds for one subject applying the LS-SVM, the logistic regression and the kernel logistic regression classifiers. Figures 7.8(a)-(e) plot the comparative results of each of the 10-folds for the five subjects for dataset IVa. The figures show the individual classification accuracies against each of the 10-folds for the logistic regression, kernel logistic regression and the proposed LS-SVM on the testing sets for subjects **aa**, **al**, **av**, **aw**, **ay**, respectively. From these figures, it is observed that in most of the cases, the proposed LS-SVM classifier yields a better performance for each of the 10-folds compared to the logistic regression and the kernel logistic regression. An increasing tendency of prediction accuracy in every fold of all subjects for the LS-SVM is shown in these figures.

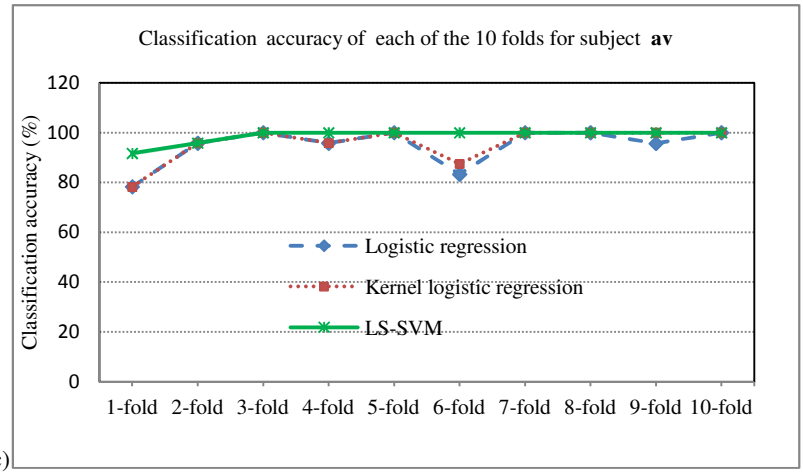


(a)

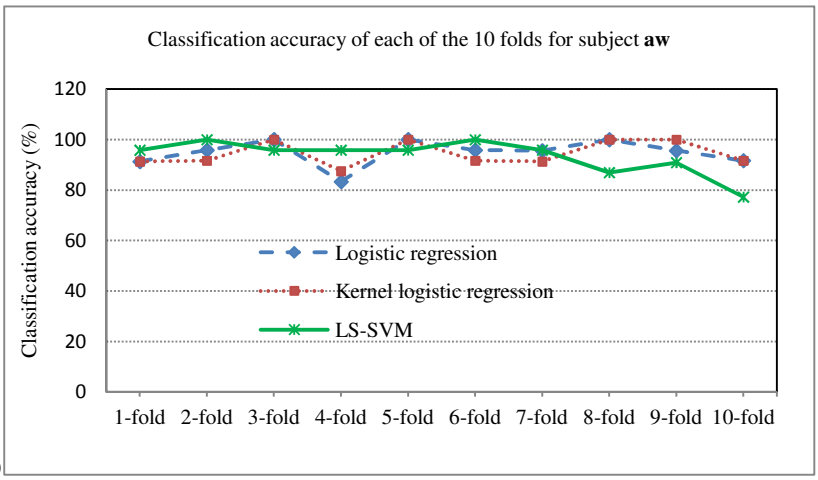


(b)

(c)



(d)



(e)

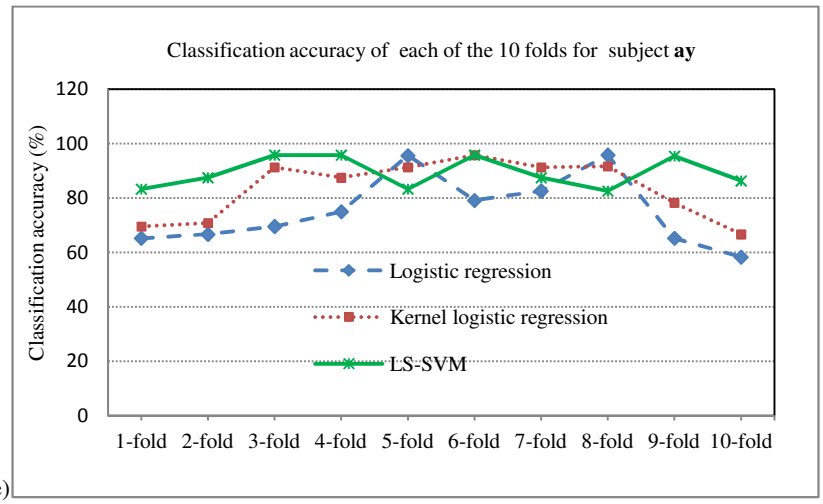


Figure 7.8: Comparisons of the individual classification accuracies among the logistic regression, kernel logistic regression and the LS-SVM for each of the 10-folds: (a) subject **aa** (b) subject **al** (c) subject **av** (d) subject **aw** (e) subject **ay** in dataset IVa.

From Figures 7.8(a)-(e), the fluctuations of the performance of the proposed method are smaller among the 10-folds for each subject compared to the logistic regression

model and the kernel logistic regression model, indicating that the proposed method is fairly stable.

In Table 7.4, we provide the classification accuracy for the proposed LS-SVM, the logistic regression and the kernel logistic regression models using the 10-fold cross validation procedure for dataset IVb. As shown in Table 7.4, the classification accuracy is 97.89% for the LS-SVM while this value is 95.31% for the logistic regression and 94.87% for the kernel logistic regression. The results show a 2.58% improvement in the proposed LS-SVM compared to the logistic regression and 3.02% over the kernel logistic regression for the same inputs.

Table 7.4: Classification results by the 10-fold cross-validation method on testing set of dataset IVb.

Methods	10-fold cross-validation accuracy (%) (mean ± standard deviation)
LS-SVM	97.89±2.96
Logistic regression	95.31±5.88
Kernel logistic regression	94.87±6.98

The standard deviation value is also smaller in the LS-SVM compared to the logistic regression and the kernel logistic regression, which reflects the consistency of the LS-SVM. The results in terms of the 10-fold cross-validation accuracy on both datasets, displayed in Table 7.3 and Table 7.4, demonstrate that the proposed LS-SVM classifier is superior compared to the logistic regression and the kernel logistic regression methods for the same features.

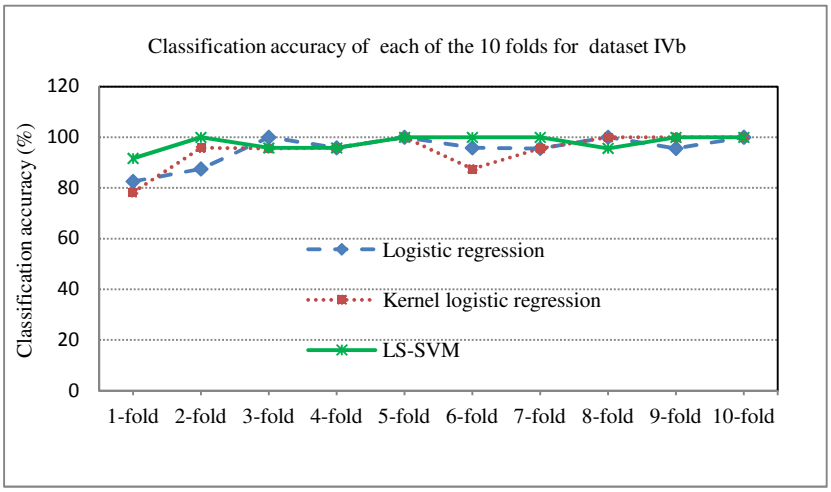


Figure 7.9: Comparisons of the individual classification accuracies among the logistic regression, kernel logistic regression and the LS-SVM for each of the 10-folds in dataset IVb.

From Figure 7.9, it is observed that in most of the 10-folds, the proposed LS-SVM generates higher accuracies and also the variations of the performance among the 10-folds are smaller compared to those of the logistic regression and the kernel logistic regression. These indicate that the proposed method is more reliable for the MI signal classification. From Figures 7.8(a)-(e) and Figure 7.9, it is clear that the proposed algorithm achieves a better classification performance both individually and overall, compared to the logistic regression and the kernel logistic regression.

In order to report the performance with a different channel data as a reference signal, we use the electrode position C3 (according to the 10/20 system) as a reference signal instead of Fp1 in the present algorithm for each subject of the both datasets. Like Fp1, C3 is also selected from the right hand class from dataset IVa, while it is from the right foot class for dataset IVb. Using the 10-fold cross-validation procedure, the proposed LS-SVM classifier yields the classification accuracy of 99.58%, 94.94%, 98.64%, 93.26% and 91.06% for subjects **aa**, **al**, **av**, **aw**, **ay**, respectively, for the reference signal of channel C3; whereas these values are 97.88%, 99.17%, 98.75%, 93.43%, and 89.36% for the reference signal of Fp1 in dataset IVa as shown in Table 7.5. The overall accuracy for the LS-SVM is 95.49% for channel C3 and 95.72% for channel Fp1. For the same dataset, the logistic regression generates the 10-fold cross-validation accuracy of 97.88% for subject **aa**, 78.32% for subject **al**, 97.83% for subject **av**, 96.18% for subject **aw** and 68.91% for subject **ay** using channel C3 as the reference signal; whereas those values are 95.31%, 87.26%, 94.89%, 94.93%, 75.33% for the reference signal of channel Fp1. The average classification accuracy of the logistic regression reaches at 87.82% for the reference signal C3 and 89.54% for the reference signal Fp1. For the same dataset, the kernel logistic regression produces classification accuracy of 96.16%, 89.78%, 96.58%, 93.15%, 85.09% for subjects **aa**, **al**, **av**, **aw**, **ay**, respectively, for the reference channel C3; whereas these values are 97.03%, 96.20%, 95.74%, 94.51%, 83.42% for the reference channel Fp1as provided in Table 7.5. Thus the kernel logistic regression achieves the overall accuracy for five subjects as 92.15% for the reference signal C3 and 93.38% for the Fp1. For dataset IVb, we obtain an accuracy of 97.88% for the proposed LS-SVM algorithm with the reference signal C3; while this value is 97.89% for the reference signal Fp1as shown in Table 7.6. On the other hand, the logistic regression classifier produces a classification accuracy of

86.41% for the channel C3 and 95.31% for the channel Fp1 for the same dataset. For reference channel C3, the kernel logistic regression is able to generate the classification performance of 95.36% where this value is 94.87% for the reference channel Fp1. From the results of both the reference signals C3 and Fp1, it is observed that the performance of the proposed algorithm does not differ significantly from changing the reference signal. It proves the robustness of the method.

In addition, to investigate the performance of the proposed six features, we add other three features, *inter quartile range* (*IQR*), *1/4 percentile* (P_{25}) (*first quartile*, $Q_1 = P_{25}$) and *3/4 percentile* (P_{75}) (*third quartile*, $Q_3 = P_{75}$), into our existing feature set. As mentioned before, our existing feature set consists of six features which are {*mean*, *median*, *mode*, *standard deviation*, *maximum* and *minimum*}. Adding the three features {*IQR*, P_{25} and P_{75}} into the existing feature set, we get a feature set of nine features, which is {*mean*, *median*, *mode*, *standard deviation*, *maximum* and *minimum*, *IQR*, P_{25} and P_{75}}. Table 7.5 and Table 7.6 present the classification results of three classifiers, the LS-SVM, the logistic regression and the kernel logistic regression for the nine features set in comparing with the results of the existing six features set for two reference signals, Fp1 and C3, for datasets IVa and IVb, respectively. As shown in Table 7.5, the proposed LS-SVM based algorithm with the nine features achieves the classification accuracy of 96.51%, 97.05%, 97.48%, 95.21% and 90.63% for subjects **aa**, **al**, **av**, **aw**, **ay**, respectively, in dataset IVa, for the reference signal Fp1. These values are 97.88%, 99.17%, 98.75%, 93.43% and 89.36 % for the existing feature set with the same reference signal.

Table 7.5: Classification results of the three classifiers for the nine features and the six features for the reference signals, Fp1 and C3, in dataset IVa.

Sub	Fp1 ref signal		C3 ref signal		Fp1 ref signal		C3 ref signal		Fp1 ref signal		C3 ref signal	
	SVM_9	SVM_6	SVM_9	SVM_6	LR_9	LR_6	LR_9	LR_6	KLR_9	KLR_6	KLR_9	KLR_6
aa	96.51	97.88	98.29	99.58	97.90	95.31	97.05	97.88	98.30	97.03	94.87	96.16
al	97.05	99.17	95.78	94.94	85.60	87.26	88.93	78.32	96.20	96.20	90.18	89.78
av	97.48	98.75	98.22	98.64	95.31	94.89	97.83	97.83	95.74	95.74	96.16	96.58
aw	95.21	93.43	94.91	93.26	94.89	94.93	92.00	96.18	95.74	94.51	94.86	93.15
ay	90.63	89.36	91.02	91.06	82.14	75.33	66.36	68.91	89.35	83.42	86.36	85.09
avg	95.38	95.72	95.65	95.49	91.17	89.54	88.43	87.82	95.07	93.38	92.49	92.15

Note: Sub=subject; avg=average; ref= reference; SVM_9=LS-SVM with the nine features; SVM_6=LS-SVM with the six features; LR_9=logistic regression with the nine features; LR_6=logistic regression with the six features; KLR_9=kernel logistic regression with the nine features; KLR_6=kernel logistic regression with the six features.

For the reference signal C3, the proposed method with nine features is able to provide accuracy of 98.29% for subject **aa**, 95.78% for subject **al**, 98.22% for subject

av, 94.91% for subject **aw** and 91.02% for subject **ay**; while these values are 99.58%, 94.94%, 98.64%, 93.26% and 91.06%, respectively, for the six features set. From Table 7.5, it is also seen that the average classification rates of the proposed method are 95.38% for the nine features and 95.72% for the six features with the reference signal Fp1; while these values are 95.65% and 95.49%, respectively, with the reference signal C3.

For the same dataset, Table 7.5 reports that the logistic regression with the reference signal Fp1 obtains 97.90%, 85.60%, 95.31%, 94.89% and 82.14% classification accuracy for subjects **aa**, **al**, **av**, **aw**, **ay**, respectively, for the nine features; whilst those values are 95.31%, 87.26%, 94.89%, 94.93% and 75.33%, respectively, for the six features. With the reference signal C3, the logistic regression produces 97.05% for subject **aa**, 88.93% for subject **al**, 97.83% for subject **av**, 92.00% for subject **aw** and 66.36% for subject **ay** for the nine features; while these are 97.88%, 78.32%, 97.83%, 96.18% and 68.91% for the six features. As shown in Table 7.5, the overall performance of the logistic regression model is 91.17% for the nine features and 89.54% for the six features with the reference signal Fp1, and 88.43% and 87.82%, respectively, for the reference signal C3.

On the other hand, it can be seen from Table 7.5 that the kernel logistic regression with the reference signal Fp1 yields the classification accuracy of 98.30%, 96.2%, 95.74%, 95.74% and 89.35% for subjects **aa**, **al**, **av**, **aw**, **ay**, respectively, for the nine features; whereas these values are 97.03%, 96.20%, 95.74%, 94.51% and 83.42 %, respectively, for the six features. With the reference signal C3, this algorithm achieves 94.87% for subject **aa,** 90.18% for subject **al**, 96.16% for subject **av**, 94.86% for subject **aw** and 86.36% for subject **ay**, respectively, for the nine features; while those values are 96.16%, 89.78%, 96.58%, 93.15% and 85.09%, respectively, for the six features. The average accuracies of this algorithm are 95.07% for the nine features and 93.38% for the six features with the reference signal Fp1; where the values are 92.49% and 92.15%, respectively, for the reference signal C3.

In dataset IVb, the LS-SVM classifier with the three added features {IQR, P_{25} and P_{75}} generates 97.48% accuracy for the reference signal Fp1 and 97.88% for the reference signal C3; where these values are 97.89% and 97.88%, respectively, for the six features as shown in Table 7.6. For the reference signal Fp1, the classification

accuracies of the logistic regression are obtained as 94.47% for the nine features and 95.31% for the six features, while these values are 87.70% and 86.41% for the reference signal C3. On the other hand, the kernel logistic regression with the nine features provides the classification performance of 95.31% for the reference signal Fp1 and 96.65% for the reference signal C3; while these values are 94.87% and 95.36% for the six features.

Table 7.6. Classification results of the three classifiers for the nine features and the six features for the reference signals, Fp1 and C3, in dataset IVb.

Methods	Nine features		Six features	
	Fp1 ref signal	C3 ref signal	Fp1 ref signal	C3 ref signal
LS-SVM	97.48	97.88	97.89	97.88
Logistic regression	94.47	87.70	95.31	86.41
Kernel logistic regression	95.31	96.65	94.87	95.36

From the discussions, we can see that there is no significant difference of performance between the nine features and the six features. If there are outliers (an *outlier* is an observation that lies an abnormal distance from other values in a set of data) in the data, the *IQR i*s more representative than the standard deviation as an estimate of the spread of the body of the data. The *IQR* is less efficient than the *standard deviation* as an estimate of the spread when the data is approximately normally distributed. For the same type of distribution (normal distribution), P_{25} and P_{75} are not good measures to represent a distribution. As the datasets used in this study are almost symmetric and there are no obvious outliers, that is why, we do not get significantly better performance when the three features {*IQR,* P_{25} *and* P_{75}*}* are added into the six features.

7.5.4 Performance comparisons with the existing techniques

In order to further examine the efficiency of the proposed algorithm, this section provides the comparisons of our approach with other eight recently reported techniques. Those eight existing algorithms for dataset IVa are discussed in Section 7.2. Table 7.7 reports the comparison results of the classification accuracy rates for the proposed method and the eight algorithms for dataset IVa. This table shows the classification performance for the five subjects as well as the overall mean accuracy

values. The highest classification accuracy rate among the nine algorithms is highlighted in bold font for each subject and their averages.

From Table 7.7, it is noted that the proposed CC-LS-SVM algorithm provides better classification accuracies than the other eight algorithms in all of the five subjects. The highest classification rates are 97.88% for subject **aa**, 99.17% for subject **al,** 98.75% for subject **av**, 93.43% for subject **aw** and 89.36% for subject **ay** produced by our proposed approach. Further looking at the performance comparisons in Table 7.7, it is noted that the proposed algorithm is ranked first in terms of the average accuracy (95.72%), while the CT-LS-SVM algorithm (Siuly et al., 2011c) comes second (88.32%), R-CSP with aggregation (Lu et al., 2010) is third (83.9%) and so on.

Table 7.7: Performance comparisons for dataset IVa.

Method	Classification accuracy rate (%)					
	aa	**al**	**av**	**aw**	**ay**	Average
CC-LS-SVM (proposed)	**97.88**	**99.17**	**98.75**	**93.43**	**89.36**	**95.72**
CT-LS-SVM (Siuly et al., 2011c)	92.63	84.99	90.77	86.50	86.73	88.32
R-CSP with aggregation (Lu et al., 2010)	76.8	98.2	74.5	92.9	77.0	83.9
SSRCSP (Lotte et al. 2011)	70.54	96.43	53.57	71.88	75.39	73.56
TRCSP (Lotte et al. 2011)	71.43	96.43	63.27	71.88	86.9	77.98
WTRCSP (Lotte et al. 2011)	69.64	98.21	54.59	71.88	85.32	75.93
SRCSP(Lotte et al. 2011)	72.32	96.43	60.2	77.68	86.51	78.63
R-CSP with generic learning (Lu et al., 2009)	69.6	83.9	64.3	70.5	82.5	74.20
Sparse spatial filter optimization (Yong et al., 2008)	57.5	86.9	54.4	84.4	84.3	73.50

Note: CC=cross-correlation technique; CT= Clustering technique; R-CSP= Regulized common spatial pattern; CSP= Common spatial pattern;

The sparse spatial filter optimization (Yong et al., 2008) is the last (73.50%). The results indicate that the proposed method achieves by 7.40% to 22.22% improvements over all the eight existing algorithms for BCI competition III, dataset IVa.

7.6 Conclusions

In this chapter, we present the CC-LS-SVM algorithm for improving the classification accuracy of the MI-based EEG signals in BCI systems. The proposed scheme utilizes a cross-correlogram based feature extraction procedure for the MI signals, and develops a LS-SVM classifier for the classification of the extracted MI features. We apply the same features as the inputs to a logistic regression and the kernel logistic regression models for comparing the performance of the proposed LS-

SVM classifier. In addition, we compare our proposed approach with eight other recently reported methods. As the parameters of the LS-SVM can significantly affect the classification performance, we use a two-step grid search algorithm for selecting optimal combinations of parameters of the LS-SVM classifier. The methods are tested on datasets IVa and IVb of BCI Competition III. All experiments on both datasets are evaluated through a 10-fold cross-validation process, which indicates the reliability of the obtained results. The main conclusions of this study are summarized as follows:

1. The proposed CC-LS-SVM method is promising for a two-class MI EEG signal classification. The feasibility of the approach has been verified with BCI competition III, datasets IVa and IVb.
2. The cross-correlation feature extraction procedure is effective for the classification performance even when the data size is very large. The experimental results from the three classifiers, the LS-SVM, the logistic regression and the kernel logistic regression, confirm that the extracted features are reliable for capturing the valuable information from the original MI signal patterns.
3. To further investigate the reliability of the obtained features, we add other three features {*IQR, P_{25} and P_{75}},* into the current six features and then employ the LS-SVM, the logistic regression and the kernel logistic regression algorithms as the inputs, separately. The results show that the performance of the nine features is not much improved in comparison with those of the six features for each of the three algorithms.
4. The experimental results using the proposed algorithm are consistent because the parameter values of the LS-SVM classifier are optimally selected through the two-step grid search algorithm rather than by the manual selection.
5. The results show that the proposed LS-SVM classifier achieves a better performance compared to the logistic regression and the kernel logistic regression classifiers for the same feature vectors in both datasets.
6. The experimental results also indicate that the proposed approach is better than the other eight recently reported methods in BCI Competition III, dataset IVa, by at least 7.40%. It demonstrates that our method performs the best for the MI signal classification in BCI applications.

This study concludes that the CC-LS-SVM algorithm is a promising technique for MI signal recognition and it offers great potentials for the development of MI-based BCI analyses which assist clinical diagnoses and rehabilitation tasks.

CHAPTER 8
AN INVESTIGATION STUDY

In Chapter 7, we introduced the three methods; the cross-correlation based least square support vector machine (CC-LS-SVM), the cross-correlation based logistic regression (CC-LR) and the cross-correlation based kernel logistic regression (CC-KLR) for the MI signals classification in BCI applications. In this chapter, we report a comparative study between motor area EEG and all-channels EEG for those three algorithms, particularly to investigate two issues: (i) which algorithm is the best one for the MI signal classification? (ii) Which EEG dataset is better for providing more information about the MI signal classification? Is it the motor area data or is it the all-channels data? In this chapter, we apply the 3-fold cross validation procedure to reduce the computation time and the number of experiments. The optimal values of the parameters of the LS-SVM method is selected by the two-step grid search algorithm in each of the three folds. Recently Wang et al. introduced a concept to select EEG channels for MI tasks over the motor cortex area in their article (Wang et al., 2007). In this study, we also follow Wang et al's idea during the consideration of motor area EEG data. We implement these three algorithms on the motor area EEG and the all-channels EEG to investigate how well they perform and also to test which area EEG is better for the MI EEG data classification. These three algorithms are also compared with other existing methods.

The work in this chapter is submitted for publication to the *International Journal of Intelligent Systems Technologies and Applications.*

8.1 Background

The ability to communicate with the outside world is one of the most indispensible assets that people have. Our hands, legs and other limbs are essential for performing our daily activities. Unfortunately, these abilities can be lost due to accidents or diseases (Kayikcioglu, & Aydemir, 2010). The neurological diseases can disrupt the neuromuscular channels through which the brain communicates with its environment

and exerts control. Therefore, it is impossible for the people who are motor disabled, to live and meet their daily needs without external help. The BCI is a well known emerging technology in which people are able to communicate with their environment and control prosthetic or other external devices by using only their brain activity (Vaughan et al., 2003). It promises to provide a way for people to communicate with the outside world using *thoughts* alone. A MI based BCI system translates a subject's motor intention into a command signal through real-time detection of motor imagery states, e.g. imagination of a left hand or right hand movement.

A BCI system, by extracting EEG signals directly from the brain, might help to restore abilities to patients who have lost sensory or motor functions because of their disabilities. The major purpose of BCIs is to translate a brain activity into a command to control an external device (Wolpaw et al., 2002). Users produce different brain activity patterns that will be identified by the system and translated into commands. In most existing BCIs, this identification relies on a classification algorithm (Wolpaw et al., 2002), i.e. an algorithm that aims at automatically estimating the class of data as represented by a feature vector. BCI applications are considered to be pattern recognition problems that signal processing, feature extraction and pattern classification techniques are attempting to solve.

It is known that EEGs record brain activities as multichannel time series from multiple electrodes placed on the scalp of a subject. The different signals from different scalp sites do not provide the same amount of discriminative information. In this study, we are interested in investigating the performance of the EEG channels of the motor cortex area and the all-channels EEG data. In the human brain, the motor cortex area is a very important area that controls voluntary muscle movements which are discussed in detail in Section 8.3. In this study, the EEG channels of motor cortex area are considered according to the suggestions of Wang et al. (2007). The major aim of this study is to investigate which area (motor area or the whole brain) is better for acquiring the MI information for classification and also to explore which algorithm performs better for the MI classification.

In this chapter, we employ the three algorithms; CC-LS-SVM, CC-LR and CC-KLR algorithms for the MI tasks classification. We implement these three algorithms considering the electrode C3 (according to the International 10-20

electrode placement system) as the reference signal on datasets, IVa and IVb, from BCI Completion III (Blankertz et al., 2006; BCI Competition III-online) as the same way like Chapter 7. There is only one difference, which is, in Chapter 7, we used the 10-fold cross validation method to evaluate the performances. But in this chapter, we consider the 3-fold cross validation procedure to reduce the computation time and the number of experiments. In both datasets, the experimental outcomes demonstrate that the LS-SVM classifier performs much better than the LR and KLR classifiers on the cross-correlation features in both areas and the all-channel EEG data generate better performance than the motor area EEG data in each three algorithms. The classification accuracy of the CC-LS-SVM algorithm is higher for the all-channels data than the channels of motor area. The experimental results also show that the CC-LS-SVM algorithm is superior to the existing methods for the motor area EEG and the all-channels EEG data.

8.2 Review of the existing research

Over the last two decades, there have been numerous studies performed on BCIs for MI tasks classification for dataset IVa of BCI Competition III. A number of research groups have developed BCIs that employ brain signals from the motor cortex area, for example, Wang et al. (2007) and Song et al. (2007). Some researchers introduced several methods for analysing the entire channels EEG data for BCI applications and investigated the physiological nature of the experimental paradigms, for instance, Blankertz et al. (2006) and Wu et al. (2008).

Wang et al. (2007) introduced a technique based on independent component analysis (ICA) with constraints, applied to the rhythmic EEG data recorded from a BCI system to isolate the rhythmic activity for MI tasks over the motor cortex area. Their algorithm includes three parts: spatial filter generation, power feature extraction and classification. They used a spatial filter through the technique of spectrally constrained ICA (cICA) and extracted power feature in μ-rhythm frequency band as the major classification pattern. An advanced SVM was applied to classify the power features. The results demonstrated that the more advanced SVM with cICA based power features did not show a significant improvement in performance.

Song et al. (2007) reported a framework to classify EEGs for BCI learning optical filters for dynamical system (DS) features. They used EEG signals as the output of a networked dynamical system (cortex) and exploited synchronization features from the DS for classification. They also proposed a new design for learning optical filters automatically from the data by employing a fisher ratio criterion on the motor cortex area. Experimental evaluations show that the dynamic system features combined with a filter learning approach is not enough to produce competitive performance on the motor cortex area for the MI signal classification in BCI applications. One of the disadvantages is that the parameters of their method were tuned manually.

The BCI III Winner algorithm (Blankertz, et al., 2006) involves an ensemble classifier based on three methods: (1) common special pattern (CSP) on even-related desynchronization (ERD) (2) autoregressive (AR) model on ERD and (3) Linear Discriminant Analysis (LDA) on temporal waves of readiness potential for dataset IVa. This algorithm was implemented for the all-channels of EEG data. For subjects, ***aa*** and ***aw*** in dataset of IVa of BCI Competition, all three features (ERD feature by CSP analysis, ERD feature by a AR model and ERP feature by LDA on temporal waves) have been used and combined by a bagging method. For the other three subjects, ***al***, ***av*** and ***ay***, of the same dataset, only the CSP based feature was used. Furthermore, the Winner algorithm used the bootstrap aggregation and employed formerly classified test samples in subjects, ***aw*** and ***ay***, to achieve the best performance.

Wu et al. (2008) reported an algorithm for classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL). In their adopted framework, feature extraction and feature classification are treated as independent stages: spectral filters and the classifier are parameterized jointly in the maximal margin hyperplane for optimization, and thereby their generalization performance can be controlled for the all-channels data. The results for the all-channels data demonstrated the efficacy of ISSPL and the resultant spectral filters did not suffer from the potential overfitting problem and only a few steps of iterations were needed to obtain a satisfactory classification performance.

Although many methods have been developed in the past decade that yield impressive results in interpreting BCI data, the BCI technologies are still not

adequate for identifying the MI tasks from original data. This study addresses two questions: (i) what algorithm is the best for the MI classification? (ii) Which EEG dataset is better for the MI signal classification? Is it the motor area data or is it the all-channels data? To answer these two questions, this chapter uses the three algorithms based on the CC technique as described in Chapter 7.

8.3 Data and Implementation

Two publicly available datasets, IVa and IVb of BCI Competition III (Blankertz et al., 2006; BCI competition III-online), are used in this chapter for experimental evaluation. We already provide the description of dataset IVa and IVb of BCI Competition III, in Chapter 4 and Chapter 5, respectively. All EEG data of these two sets were collected during motor imagery (MI) tasks.

In this study, we intend to implement our three methods on the electrodes of the motor cortex area of the brain and also on the all-channel electrodes for comparison. The channels recorded from the motor area are chosen to investigate the activities of the motor cortex area of the brain for the proposed algorithms and the all-channels are considered to see how the classification algorithms handle feature vectors of relatively high dimensions. Actually we are interested to see the performance of the three algorithms on the two areas (motor area and all-channels data) and also to decide which algorithm is better for given areas of the brain. We know that only a particular part of the brain is activated in response to an MI task which is called the motor cortex. Motor cortex is one of the important brain areas most involved in controlling and execution of voluntary motor functions and this area of the brain is typically associated to the MI movements.

As we are looking for a response specifically in the motor cortex area, we manually select the 18 electrodes around the sensorimotor cortex based on the placement of international 10-20 system which includes the channels C5, C3, C1, C2, C4, C6, CP5, CP3, CP1, CP2, CP4, CP6, P5, P3, P1, P2, P4 and P6 from each of the two datasets. Wang et al. (2007) also considered the same electrodes for their research and their experimental results suggested that these electrodes are the best channels for getting the MI information.

As described before, the both datasets are originally recorded from 118 electrodes. Figure 8.1 presents the locations of electrodes of datasets, IVa and IVb from BCI competition III. 118 electrodes are shown labelled according to the extended international 10-20 system. This figure was made in EEGLAB (MATLAB toolbox for processing data from EEG, magnetoencephalography (MEG), and other electrophysiological signals) and the electrode system is described in (Oostenveld & Praamstra, 2001). Wang et al. (2007) explained that the selected electrodes cover the motor cortex area. Thus, prior knowledge as well as the results of the following electrodes are investigated in this study.

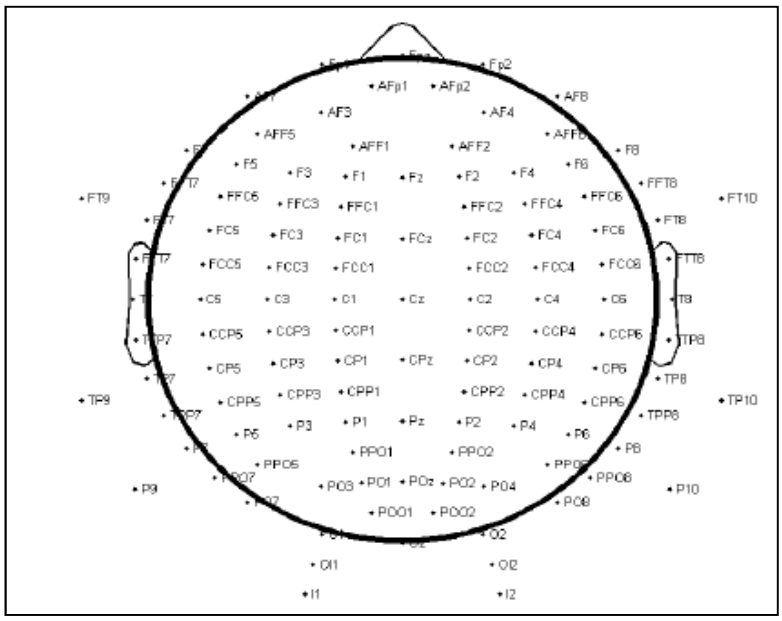


Figure 8.1: Locations of electrodes for datasets IVa and IVb in BCI Competition III. 118 electrodes are shown labelled according to the extended international 10-20 system described in (Oostenveld & Praamstra, 2001).

In this study we firstly consider the electrode position C3 of the RH class as a reference channel from each subject of both datasets for the CC technique. This study uses the channel of the C3 electrode in the international 10-20 system as the reference channel. The C3 electrode is the best candidate for supplying the MI information about brain activities during the MI tasks in the international 10-20 system (Sander et al., 2010). In each subject, the C3 channel is used as a reference channel for both the motor imagery EEG data and the all-channels EEG data.

Secondly, in the motor area data, the reference channel C3 of the RH class is cross-correlated with the data of the remaining 17 channels of that class and the data

of all 18 channels of the RF class for each subject of both datasets. Thus total 35 cross-correlation sequences are obtained from the two classes of each subject. Then the mentioned six statistical features, *mean*, *maximum*, *minimum*, *standard deviation* , *median* and *mode* values are calculated from each cross-correlation sequence and a feature vector set of 35×6 size is created. In the all-channels data, the reference channel C3 of the RH class is cross-correlated with 117 channels of this class and also 118 channels data of the RF class in each subject of both datasets. Thus we acquire a total of 235 cross-correlation sequences from the two-class MI data of a subject and then we extract previously mentioned six statistical features from each cross-correlation sequence to generate a feature vector set of 235×6 size.

Thirdly, we divide the feature vector set randomly as the training set and the testing set using the 3-fold cross-validation method (Siuly et al., 2011c; Abdulkadir, 2009) in both the motor cortex set and the all-channels data, separately. In the 3-fold cross-validation procedure, a feature vector set is partitioned into 3 mutually exclusive subsets of approximately equal size and the method is repeated 3 times (folds). Each time, one of the subsets is used as a test set and the other two subsets are put together to form a training set. Then the average accuracy across all 3 trials is computed.

Finally, we employ these feature vector sets as the input to the LS-SVM, the LR and also to the KLR. In the CC-LS-SVM algorithm, the training set is applied to train the LS-SVM classifier and the testing set is used to verify the effectiveness of the classifier for both datasets. As the result of the LS-SVM relies largely on the choice of a kernel, the RBF kernel is chosen after many trials. Before the classification, the two parameters (γ, σ^2) of the LS-SVM method are selected by applying a two-step grid search procedure (Xie et al., 2009) on each three folds for getting reliable performance of the method as these parameters play an important role in the classification performance. In the LS-SVM, the RF is treated as +1 and RH as -1 for dataset IVa, and the RF is considered as +1 and LH as -1 for dataset IVb.

In the CC-LR algorithm, we employ the training and testing sets as the inputs, separately, to the LR classifier; but we use the testing set to validate the classification accuracy of the classifier in both datasets. In the LR model, we consider independent variables x_1 as *mean* values, x_2 as *maximum* values, x_3 as *minimum* values, x_4 as

standard deviation values, x_5 as *median* values and x_6 *as mode* values. We treat the dependent variable y as RH= 0 and RF= 1 for dataset IVa, and RF=0 and LR=1 for dataset IVb. The parameters of the LR model are obtained automatically using the maximum likelihood estimation (MLE) method.

In the CC-KLR algorithm, we utilize the feature vectors and their class labels as the same process of the logistic regression for the inputs. In Section 8.4, the classification results of these three classifiers are presented for datasets IVa and IVb. The following section discusses about the performance of those three algorithms through a 3-fold cross-validation procedure.

8.4 Experiments and Results

This section discusses the experimental results of the three algorithms for the motor area EEG and the all-channels EEG in datasets, IVa and IVb, and also reports a comparative study with the existing methods. As accuracy is a major concern in BCI systems, this study uses the classification accuracy as the criterion to evaluate the performance of the proposed method. The classification accuracy is calculated by dividing the number of correctly classified samples by the total number of samples (Siuly et al., 2010; Siuly et al., 2011c; Siuly et al., 2011d). It is worthy to mention that all experimental results in both datasets, are presented based on the testing set. In this study, MATLAB (version7.7, R2008b) is used for mathematical calculations of the CC technique. The classification by the LS-SVM is carried out in MATLAB using the LS-SVMlab toolbox (LS-SVMlab toolbox (version 1.5)-online), the classification by the LR is performed using PASW (Predictive Analytics SoftWare) Statistics 18 and the KLR algorithm is executed through MATLABArsenal (MATLAB Classification Wrapper 1.00 (Debug version)) package (MATLABArsenal-online).

8.4.1 Results for dataset IVa

The complete experimental results for dataset IVa are summarized in Table 8.1. The table provides the classification performance as well as the overall mean of the CC-LS-SVM, CC-LR and CC-KLR algorithms for the motor area EEG and the all-

channels EEG. The results of each subject are reported in terms of mean ± standard deviation of the accuracy over a 3-fold cross-validation method on the testing set. In the motor area, the CC-LS-SVM algorithm yields the classification accuracy 100%, 94.19%, 100%, 96.97%, 94.45% for subjects ***aa***, ***al***, ***av***, ***aw*** and ***ay***, respectively while these values are 88.90%, 77.0%, 75.0%, 100%, 100% for the CC-LR algorithm and 85.61%, 97.22%, 100%, 100% and 93.94% for the CC-KLR algorithm. The average accuracy rate is 97.12% for the CC-LS-SVM algorithm, 88.18% for the CC-LR algorithm and 95.35% for the CC-KLR algorithm in the motor area data. So, the CC-LS-SVM algorithm provides an 8.940% of improvement over the CC-LR method and 1.77% over the CC-KLR method on average.

Table 8.1: Experimental results of the three algorithms reported in percentage (mean ± standard deviation) for dataset IVa.

Subject	Motor area data			All-channels data		
	CC-LS-SVM	CC-LR	CC-KLR	CC-LS-SVM	CC-LR	CC-KLR
aa	100±0.0	88.90±19.22	85.61±12.92	99.57±0.74	100±0.0	99.57±0.74
al	94.19±5.04	77.0±21.18	97.22±4.81	94.88±4.45	95.67±4.45	91.47±8.26
av	100.0±0.0	75.0±22.05	100±0.0	99.16±1.46	98.7±2.25	97.86±3.7
aw	96.97±5.25	100±0.0	100±0.0	97.45±1.26	100.0±0.0	98.73±1.27
ay	94.45±4.81	100±0.0	93.94±10.49	98.72±1.28	73.6±3.20	95.75±2.64
Average	97.12±3.02	88.18±12.49	95.35±5.99	97.96±1.84	93.59±1.98	96.68±3.24

The standard deviation value of a subject describes the variation of the classification accuracies among the three folds. If the variation of the accuracies among the three folds is less, it indicates robustness of the method. For the motor area data, we can see that the standard deviation among the three folds in each subject is relatively small in the CC-LS-SVM algorithm, which indicates the strength of the CC-LS-SVM algorithm.

For the EEG data recorded from the all-channels, the CC-LS-SVM algorithm produced a classification accuracy of 99.57% for subject ***aa***, 94.88% for subject ***al***, 99.16% for subject ***av***, 97.45% for subject ***aw*** and 98.72% for subject ***ay***, whereas these values are 100%, 95.67%, 98.7%, 100% and 73.6%, respectively, for the CC-LR algorithm and 99.57%, 91.47%, 97.86%, 98.73%, 95.75% for the CC-KLR algorithm, respectively. The average accuracy was 97.96% for the CC-LS-SVM algorithm, 93.59% for the CC-LR method and 96.68% for the CC-KLR algorithm. Thus the average accuracy of the CC-LS-SVM algorithm was increased by 4.37%

from the CC-LR method and 1.28% from the CC-KLR algorithm for the all-channels data. In the all-channels data, the standard deviation value in each subject was relatively low in the three algorithms. So, it can be claimed that the performance of the three algorithms are reliable in the all-channel data. The results reveal that the CC-LS-SVM algorithm performs better on the both motor area and all-channels data than the CC-LR approach and the CC-KLR algorithm. The performance of the CC-LS-SVM method is better for the all-channels data than the motor area data.

Figure 8.2 presents a comparison of the classification accuracy between the motor area EEG data and the all-channels EEG data for the CC-LS-SVM algorithm. From the figure, it may be seen that the CC-LS-SVM algorithm produces a higher performance for subject ***aa*** and subject ***av*** in the motor area EEG data than the all-channels data. On the other hand, the performance of the all-channels data is better for subject ***al***, subject ***aw*** and subject ***ay*** compared to the motor area data.

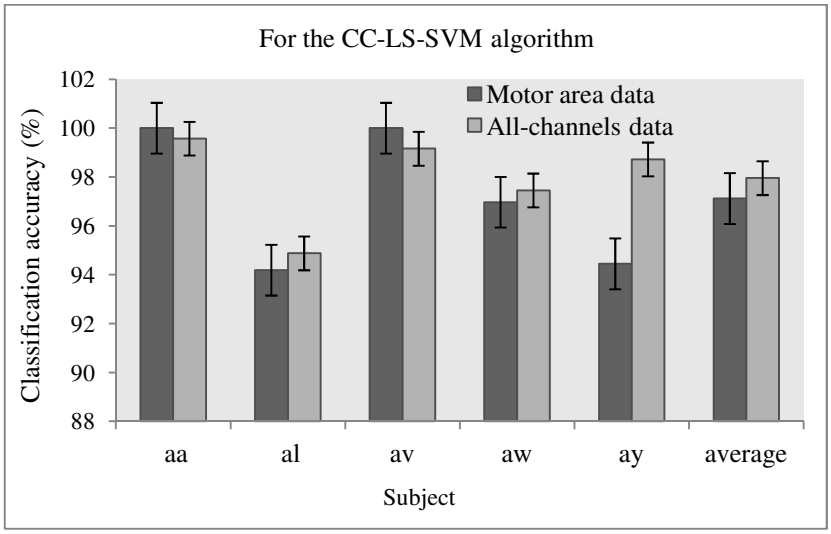


Figure 8.2: Comparison of the performance between the motor area EEG and the all-channels EEG data for the CC-LS-SVM algorithm. The vertical lines show the standard errors of the test accuracies.

Figure 8.2 also illustrates that the overall classification performance of the algorithm is much better for the all-channels data than for the motor area data. Error bars of the motor area EEG data are also higher than the all-channels data. The error bars indicate the superiority of the CC-LS-SVM algorithm for the all-channels EEG data over the motor area data.

Figure 8.3 displays the comparison of the classification accuracy between the motor area EEG and the all-channels EEG data for the CC-LR algorithm. It can be observed from the figure that the classification accuracy rates for the all-channels data are substantially higher for subjects, ***aa***, ***al*** and ***av*** and the same for subject ***aw*****,** compared to the motor area data.

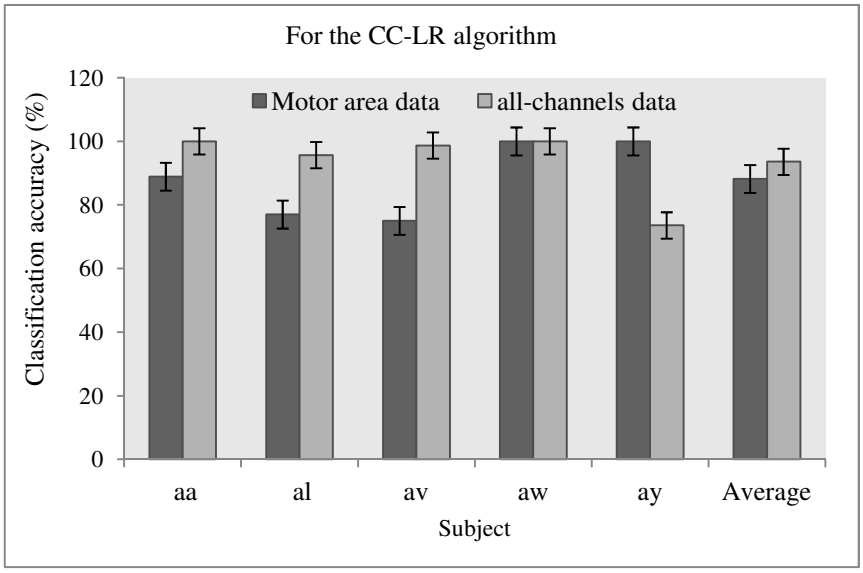


Figure 8.3: Comparison of the performance between the motor area EEG and the all-channels EEG data for the CC-LR algorithm. The vertical lines show the standard error of the test accuracies.

The motor area data provided better results only for subject ***ay*** over the all-channels data. The overall accuracy for the all-channels data is significantly higher than the motor area data for the CC-LR method.

Figure 8.4 shows the comparison between the motor area EEG and the all-channels EEG data for the CC-KLR algorithm in dataset IVa. From this figure, we can see that the motor area EEG data produce a bit better results in subjects ***al***, ***av*** and ***aw*** than the all-channels EEG data for the CC-KLR algorithm. The all-channels EEG data provide higher performance in subjects ***aa*** and ***ay*** compared to the motor channel data. The average performance of the all-channels is better than the motor area data for this algorithm. Figure 8.2, Figure 8.3 and Figure 8.4 depict that the EEG data recorded from the all-channels give the best result for both algorithms when compared to the data recorded from the motor cortex area.

Table 8.2 presents a comparison of the performances for the motor cortex area of the CC-LS-SVM, CC-LR and CC-KLR algorithms with the previously

existing methods; SVM on constraints independent component analysis (cICA) power features (Wang et al., 2007) and SVM on dynamical system (DS) features (Song et al., 2007).

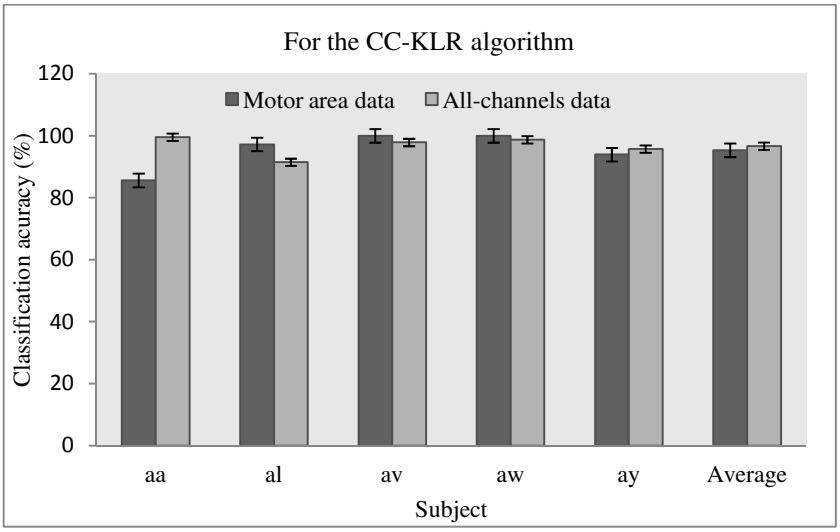


Figure 8.4: Comparison of the performance between the motor area EEG and the all-channels EEG data for the CC-KLR algorithm. The vertical lines show the standard error of the test accuracies.

These two existing methods are also implemented on the motor cortex area data for dataset IVa as discussed in Section 8.2. From Table 8.2, it can be seen that the highest accuracy was obtained by the CC-LS-SVM algorithm for subject ***aa*** and subject ***av***. The CC-LR method achieved a better performance for subject ***aw*** and subject ***ay*** than the other methods. The CC-KLR method produced the best performance for subjects ***al, av*** and ***aw***.

Table 8.2: The comparison of our three algorithms with two existing methods for the motor area data in dataset IVa.

Subject	Classification accuracy on the motor area data (%)				
	CC-LS-SVM	CC-LR	CC-KLR	SVM on cICA power features (Wang et al., 2007)	SVM on DS features (Song et al., 2007)
aa	100.0	88.9	85.61	85.7	83.3
al	94.19	77.0	97.22	89.3	96.3
av	100.0	75.0	100	75.0	72.7
aw	96.97	100.0	100	85.3	86.9
ay	94.45	100.0	93.94	85.0	89.0
Average	97.12	88.18	95.35	84.06	85.64

In Table 8.2, it is noted that the CC-LS-SVM algorithm provided the best result with an average classification accuracy of 97.12% while this value is 88.18% for the CC-LR algorithm, 95.35% for CC-KLR algorithm, 85.64% for the SVM on DS algorithm and 84.06% for the SVM based on cICA approach. The CC-LS-SVM method achieves by 1.77% to 13.06% improvements for the motor area data over the four algorithms for dataset IVa.

Table 8.3 lists a comparison study for the all-channels data of our three algorithms with BCI III Winner (Blankertz, et al., 2006) and iterative spatio-spectral patterns learning (ISSPL) (Wu et al., 2008) for dataset IVa. A detailed description of BCI III Winner (Blankertz, et al., 2006) and ISSPL (Wu et al., 2008) methods are provided in Section 8.2. The CC-LS-SVM algorithm produced an excellent result for subject ***av*** and subject ***ay,*** while the CC-LR algorithm achieved the best results for subject ***aa*** and subject ***aw***. Our CC-KLR method also provided better results for subjects, ***aa*** and ***ay,*** than the two popular existing methods, the BCI III Winner algorithm and the ISSPL algorithm. The BCI III Winner method gave the best performance for subject ***al*** and subject ***aw***. Both BCI III Winner and ISSPL methods achieved a 100% accuracy for subject ***al***. Obviously, the average classification accuracy of the CC-LS-SVM method is excellent for the all-channels data.

Table 8.3: The comparison of our three algorithms with two existing methods for the all-channels data in dataset IVa.

Subject	Comparison of accuracy on the all-channels data (%)				
	CC-LS-SVM	CC-LR	CC-KLR	BCI III Winner (Blankertz, et al., 2006)	ISSPL (Wu et al., 2008)
aa	99.57	100	99.57	95.50	93.57
al	94.88	95.67	91.47	100.0	100.0
av	99.16	98.7	97.86	80.6	79.29
aw	97.45	100	98.73	100	99.64
ay	98.72	73.6	95.75	97.6	98.57
Average	97.96	93.59	96.68	94.20	94.21

Table 8.3 depicts that the CC-LS-SVM algorithm is able to increase the classification accuracy by 4.37% from the CC-LR algorithm, 1.28% from the CC-KLR algorithm, by 3.76% from BCI III Winner and by 3.75% from the ISSPL.

Generally, it can be observed from Table 8.2 and Table 8.3 that there is an improvement in the performance of the CC-LS-SVM algorithm for both the motor cortex area data and the all-channels data over the previously existing methods. Based on these results, it can be concluded that the LS-SVM method is more potential than the existing methods for the MI tasks EEG signal classification on the motor cortex area data and the all-channels data, and the all-channels data performs better than the motor area data in the MI classification.

8.4.2 Results for dataset IVb

Table 8.4 reports the classification results of the CC-LS-SVM algorithm, the CC-LR algorithm and CC-KLR algorithm on the motor cortex area data and the all-channels data for dataset IVb. These results are listed in Figure 8.5. For the CC-LS-SVM algorithm, the classification accuracy reaches 94.45% in the motor cortex area data while this value is 88.90% for the CC-LR algorithm and 83.08% for CC-KLR algorithm. For the all-channels data, the CC-LS-SVM method is able to yield the accuracy of 98.72%, where the CC-LR method produces 96.83% and the CC-KLR method produces 97.04%.

Table 8.4: Experimental results of the three algorithms reported in terms of the 3-fold cross validation accuracy (mean ± standard deviation) for dataset IVb.

Method	Classification accuracy (%)	
	Motor area data	All-channels data
CC-LS-SVM	94.45±4.81	98.72±1.28
CC-LR	88.90±19.22	96.83±0.72
CC-KLR	83.08±21.91	97.04±3.18

Therefore the performance for the all-channels data is 4.27% higher for the CC-LS-SVM, 7.93% higher for the CC-LR method and 13.96% higher for the CC-KLR method than the performance of the motor area data. For the three algorithms, the standard deviations among the three folds are relatively lower for the all-channels data than for the motor cortex area data. The lower value of the standard deviation proves the reliability of those three methods in the all-channels data.

Figure 8.5 shows a clearer picture of the performance for the CC-LS-SVM, the CC-LR and the CC-KLR algorithms applied to the motor cortex area and the all-channels data for dataset IVb. From Figure 8.5, it is observed that the three algorithms produce better results on the all-channels data than on the motor area data and the classification accuracy of the CC-LS-SVM method is higher for the all-channels data than for the motor area data. Note that we could not compare the results of the CC-LS-SVM, CC-LR and CC-KLR algorithms with any other previously existing methods for this dataset because there are no reported research results available.

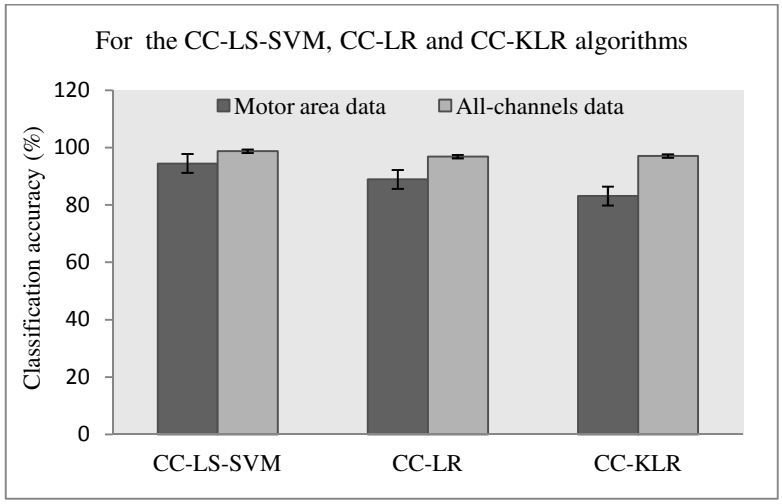


Figure 8.5: The comparison of the performance for the CC-LS-SVM, CC-LR and CC-KLR algorithms between the motor area data and the all-channels data. The vertical lines show the standard errors of the test accuracies.

The experimental results for both datasets demonstrate that the CC-LS-SVM algorithm is better for the motor cortex area data and the all-channel data than the CC-LR and CC-KLR algorithms. The results also indicate that the CC-LS-SVM algorithm provides the best performance for the all-channels data and the all-channel data produces better performance than the motor area data.

8.5 Conclusions and Contributions

In this chapter, we have employed the CC-LS-SVM, CC-LR and CC-KLR algorithms to compare performance between the motor imagery EEG data and the all-channels EEG data. The CC-LS-SVM algorithm assembles the CC technique and

the LS-SVM classifier; the CC-LR algorithm combines the CC technique and the LR model; and the CC-KLR algorithm mixtures the CC technique and the KLR classifier for MI tasks classification. In order to investigate the effectiveness of these three algorithms, we implemented them individually on the EEG data recorded from the motor cortex area and also the all-channels EEG data. The results on the two datasets, IVa and IVb of BCI Competition III, demonstrate that the CC-LS-SVM method produces a better accuracy for the all-channels EEG data and the motor area EEG data than the CC-LR and the CC-KLR algorithms. The performance of the CC-LS-SVM algorithm is higher for the all-channels data than for the motor area data for the MI EEG signal classification. The results also suggest that the CC-LS-SVM algorithm do better than some of the existing algorithms in the literature for both the motor area and the all-channels data. Thus, it can be concluded that the CC-LS-SVM algorithm is the best algorithm for the MI EEG signal classification and the all-channels EEG can provide better information than the motor area EEG for the MI classification.

In the next chapter, we present a summary of our research reported in this dissertation and also provide a direction for future work.

CHAPTER 9
CONCLUSIONS AND FUTURE WORK

9.1 Summary and Conclusions of the Dissertation

The EEG is an important measurement of brain activity and has great potential in helping in the diagnosis and treatment of mental and brain neuro-degenerative diseases and abnormalities. The classification of EEG signals is a key issue in biomedical research for identification and evaluation of the brain activity. Identification of various types of EEG signals is a complicated problem, requiring the analysis of large sets of EEG data. Representative features from a large dataset play an important role in classifying of EEG signals in the field of biomedical signal processing. In this dissertation, we studied and developed EEG signal processing and classification techniques in order to identify different types of EEG signals with three main objects:

(1) Develop methods for the classification of the epileptic EEG signals to improve the classification rate with less execution time.

(2) Introduce methods to identify EEG signals during MI tasks for the development of BCI systems.

(3) Investigate which algorithm and which EEG data (motor area data or the all-channels data) is better for the MI signal classification.

In order to reach these objectives, we first developed two methods; *simple sampling technique based least square support vector machine (SRS-LS-SVM)* and *clustering technique based least square support vector machine (CT-LS-SVM)* to contribute in the epileptic EEG signal classification.

In the SRS-LS-SVM method, we introduced a simple sampling (SRS) technique in two stages for feature extraction process (see Chapter 3). In the first stage, we selected ten ‘random samples’ from each EEG channel data and then we selected five ‘sub-samples’ from each random sample at the second stage. Finally, we calculated the nine features from each sub-sample set to represent the distribution

of the original EEG signals reducing the dimensionality of the data. In the experiments, the sample and sub-sample sizes were determined using a sample size calculator of the "Creative Research System" considering 99-100% confidence interval and 99% confidence level. We employed the LS-SVM with the RBF kernel for classification where those features were employed as inputs. Firstly, we implemented this method to the EEG epileptic data (EEG time series, 2005) to classify epileptic signals. To test the effectiveness of the method, we also employed this method to the mental imagery tasks EEG data (Chiappa & Millán**,** 2005) for the classification of different categories mental tasks. From the experimental evaluation, we achieved an average classification accuracy of 95.58% for EEG epileptic data and 98.73% for mental imagery tasks EEG data. We were able to achieve a classification accuracy of 100% on the EEG epileptic database for the pair of healthy subjects with eyes open and epileptic patients during seizure activity. In order to further test the efficacy of the method on non-EEG data, we also applied this method to a two-class synthetic data from Ripley (1996) and we obtained impressive results again. Therefore, the experimental results demonstrated that the SRS-LS-SVM is promising for capturing representative characteristics of EEG signals by the SRS technique, and for the classification of these signals by LS-SVM, which can be used as a new intelligent diagnosis system.

To reduce experimental time and also to improve the classification performance, we developed the CT-LS-SVM for the epileptic EEG signals classification (see Chapter 4). In this method, we proposed the clustering technique (CT) approach as a new process for feature extraction from the EEG data. In this procedure, each set of EEG channel data is divided into n (n=16) mutually exclusive groups named 'clusters' with a specific time duration. Each cluster is again partitioned m (m=4) into 'sub-clusters' over a specific time period and then nine statistical features; *minimum*, *maximum*, *mean*, *median*, *mode*, *first quartile*, *third quartile*, *inter-quartile range* and *standard deviation* are extracted from each sub-cluster, representing the distribution of EEG signals. These features are applied to the LS-SVM classifier as the input for classifying two-class EEG signals. We implemented this method to an epileptic EEG dataset. For further evaluation, we applied the method to a mental imagery tasks EEG database and a motor imagery EEG data to classify different pairs of two-class EEG signals. This proposed

approach has two main advantages compared to the SRS-LS-SVM. The first advantage is that this method uses all data points for the experiments. The second advantage is that by using the CT technique, it takes much less time to run the program. We evaluated the performance of this method through the 10-fold cross validation procedure. The performance of the CT-LS-SVM algorithm was compared in terms of the classification accuracy and execution (running) time with the SRS-LS-SVM method. We also compared the proposed method with other existing methods in the literature for the three databases. The experimental results showed that the proposed algorithm can produce a better classification rate than the SRS-LS-SVM method and takes much less execution time compared to the SRS-LS-SVM technique. The research findings indicate that this proposed approach is very efficient for the classification of two-class EEG signals that can distinguish the categories of EEG signals. The CT-LS-SVM algorithm can help to provide clinical information about patients who have a neurological disorder, mental or physiological problems.

Concerning MI task classification, in this dissertation, we further developed three algorithms:

(1) Cross-correlation based logistic regression (CC-LR).
(2) Modified CC-LR with diverse feature sets.
(3) Cross-correlation based least square support vector machine (CC-LS-SVM) for the classification of MI based EEG signals.

If the MI tasks are reliably distinguished through identifying typical patterns in EEG data, motor disabled people could communicate with a device by composing sequences of these mental states. These three methods were tested on two benchmarks datasets, IVa and IVb of BCI Competition III (BCI competition III, 2005). In both datasets, each subject was considered, separately, for experiments as the MI tasks EEG signals are naturally highly subject-specific, depending on physical and mental tasks.

In the CC-LR algorithm, we have combined the cross-correlation (CC) feature extraction and the logistic regression (LR) classification to identify MI tasks in BCI applications (see Chapter 5). In this algorithm, the CC technique follows three

steps to extract features from MI tasks data. At first, one of the EEG channels was selected randomly as a reference channel from a class of a subject, as there are no specific requirements for selecting a reference signal in the cross-correlation analysis. Secondly, the reference channel of a class was cross-correlated with the data of the remaining channels in this class and the data of all channels of another class. Thirdly, six statistical features: *mean, median, mode, standard deviation, maximum* and *minimum,* were calculated from each cross-correlation sequence to reduce the dimensions, which ideally represents the distribution of the signal containing important information. These values were employed to the LR model as input variables for the classification. The performance of this algorithm was measured in terms of classification accuracy using a 3-fold cross-validation method. The experimental results have demonstrated the effectiveness of the CC-LR algorithm, especially its superiority over some reported algorithms. Moreover, the CC-LR method is efficient for the identification MI tasks that can provide positive impacts to develop BCI systems.

In order to select a more suitable feature set to enhance the classification performance and also to make more reliable results for the BCI systems, we modified the CC-LR algorithm with three diverse feature sets (see Chapter 6). In the modified CC-LR method, we provide an important framework to classify the two-class MI based-EEG signals for BCI data. The building components of this proposed scheme are divided into two major parts where the feature extraction procedure is described in the first part and the classification technique in the second. In this algorithm, we provided an outline on how a reference channel is selected for the CC method considering the structure of the brain associated with MI tasks. To investigate which features are suitable for the representation of the distribution of the MI signals, three statistical feature sets are extracted from each cross-correlation sequence of a subject. The performance of each of the three feature sets is evaluated through the 3-fold cross validation method, and finally this study reached a conclusion on which features better characterize the distribution of EEG signals. The experimental results reported that the CC technique is suitable for the six statistical features: *mean*, *standard deviation*, *skewness*, *kurtosis*, *maximum* and *minimum* representing the distribution of MI tasks EEG data and C3 channel providing better classification results as a reference signal. The results also demonstrated that the method is an

improvement over some of the existing methods. The findings of this study also indicated that the CC technique has the capability to extract representative characteristics from MI tasks EEG data and the LR has the potential to identify MI tasks in BCI systems. The modified CC-LR algorithm can be used to properly identify MI tasks, which can help to generate control signals for BCI systems.

With tuning hyper parameters of the LS-SVM, we developed a CC-LS-SVM method where the CC technique was used for feature extraction and the LS-SVM classifier for the classification of the extracted features, (see Chapter 7). In order to evaluate the performance of the LS-SVM classifier, we tested a logistic regression classifier (LR) and a kernel logistic regression classifier (KLR), separately, on the same features extracted from the cross-correlation method as the inputs. Individually, we used two electrode positions, Fp1 and C3 (according to the international 10-20 system), as a reference signal for each subject of the both datasets to report the performance. From each cross-corrologram, we calculated six statistical features, *mean, median, mode, standard deviation, maximum* and *minimum* to reduce the dimensionality of the sequence. In addition, to investigate the performance of the proposed six features, we added a further three features, *inter quartile range* (*IQR*), *1/4 percentile* (P_{25}) and *3/4 percentile* (P_{75}), into our existing six feature set. From the experimental results, it was seen that there was no significant difference in performance between the existing six features set and the new feature set after adding three features. In this method, we used a two-step grid search process for selecting optimal hyper parameters of the LS-SVM as the parameters can significantly affect the classification performance. We used the 10-fold cross validation method for the evaluation of classification performance. The experimental results showed that the proposed LS-SVM classifier achieved a better performance compared to the logistic regression and the kernel logistic regression classifiers for the same feature vectors in both datasets. In order to further verify the effectiveness of the CC-LS-SVM algorithm, we also compared it with the eight most recently reported methods in the literature. The experimental results also indicated that this proposed approach do better than the other eight recently reported methods in BCI Competition III, dataset IVa, by at least 7.4%. It demonstrated that our method performed the best for the MI signal classification in BCI applications. This study concluded that the CC based LS-SVM algorithm is a promising technique for the MI signal recognition and it offers

great potential for the development of MI-based BCI analyses which assist clinical diagnoses and rehabilitation tasks.

Finally we investigated two issues for the MI tasks based EEG signal classification (see Chapter 8). Firstly, which algorithm performs better, and secondly which EEG data is more suitable for getting information about the MI. To answer these two questions, we applied the three algorithms: the CC-LS-SVM, the CC-LR and the CC-KLR. These three algorithms were implemented on the motor area EEG data and the all-channels EEG data to investigate how well they performed, and also to test which area EEG is better for the MI classification. We manually selected the 18 electrodes around the sensorimotor cortex based on the international 10-20 system. The following channels were considered from each of the two datasets which are C5, C3, C1, C2, C4, C6, CP5, CP3, CP1, CP2, CP4, CP6, P5, P3, P1, P2, P4 and P6. Wang et al. (2007) also considered the same electrodes for their research and their experimental results suggested that these electrodes are the best channels for getting the MI information. These algorithms were also compared with some existing methods, which revealed their competitive performance in classification. In this algorithm, we used C3 channel as the reference channel and employed the 3-fold cross validation procedure for the evaluation of the performance. The results on the both datasets, IVa and IVb from BCI Competition III, showed that the CC-LS-SVM algorithm performed better than the CC-LR and CC-KLR algorithms on both the motor area EEG and the all-channels EEG. Based on the results, it can be concluded that the CC-LS-SVM algorithm is the best algorithm for the MI EEG signal classification and the all-channels EEG can provide better information than the motor area EEG for the MI classification. Furthermore, the LS-SVM based approach can correctly identify the discriminative MI tasks, demonstrating the algorithm's superiority in classification performance over the most of existing methods.

Taken together, it can be concluded that the research presented in this dissertation has found new and successful methods for the reliable classification of EEG signals in biomedical signal processing. These techniques will enable neurologists to diagnose the brain degenerative diseases correctly and efficiently and are also helpful in the development of BCI systems to assist individuals with high level impairments. The outcomes will help brain disorder patients to improve the quality of their lives.

9.2 Future work

We believe that the methods presented in this dissertation would provide promising outcomes in the EEG signal classification area. Extensive future work will examine the possibility of using the methods in the application of the EEG signal classification. To facilitate the further development of those proposed methods, we have highlighted a few key issues which are addressed below.

Concerning the SRS-LS-SVM algorithm, it would be developed to study the distribution of different random samples, sub-samples and population (whole EEG channel data) using the hypothesis testing procedure. In this process, firstly the distribution of the random samples and sub-samples in each EEG channel data will be tested whether they are homogeneous or not. If they are homogenous, then sub-samples could be considered as representative of the random samples, meaning the sub-samples could be used for feature extraction instead of the random samples. After that, the distribution of the sub-samples would be compared with the distribution of population, whether they are homogeneous or not. If the sub-samples follow the same distribution as the population, it can be concluded that the sub-sample's features are most representative to explore the original EEG signal. The CT-LS-SVM method could be improved following the same process like the SRS-LS-SVM algorithm.

In addition, the SRS-LS-SVM and CT-LS-SVM algorithms would be improved tunning hyper parameter of the LS-SVM method through a two-step grid search technique. These two algorithms could be extended for multiclass EEG signal classification.

Concerning the CC-LR and modified CC-LR algorithms, they would be developed using multivariate logistic regression instead of binary logistic regression for multiclass classification purpose, and then compare with kernel logistic regression.

As frequency bands of EEG signals are important key issue in identifying the characteristics of behaviour, in the near future, the SRS-LS-SVM, CT-LS-SVM, the CC-LR, modified CC-LR and CC-LS-SVM algorithms would be developed for classification on basis of the frequency bands. To reach this aim, a slight

modification of the current methodologies may be required. The CC-LS-SVM algorithm will be extended for the classification of multiclass EEG signals.

This dissertation only studied offline classification techniques but it is desirable in practical applications for such processing to occur in real life, e.g. control a robot. This requires more work to train subjects and build a closed loop and real time processing system. Therefore, all of our methods would be developed for online classification.

Sometimes EEG signals are highly contaminated with various artifacts, both from the subject and from equipment interferences. In this dissertation, our proposed methods did not develop to remove artifacts from EEG data. Further study is needed to successfully remove artifacts without contaminating the EEG signals for our proposed algorithms. In the future, our algorithms would be developed so that there would be a significant improvement in signal classification removing artifacts.

REFERENCES

Abdulkadir, S. (2009) 'Multiclass least-square support vector machines for analog modulation classification', *Expert System with Applications*, Vol. 36, pp. 6681–6685.

Abdullah H, Maddage N C, Cosic I and Cvetkovic D (2010) 'Cross-correlation of EEG frequency bands and heart rate variability for sleep apnoea classification', *Med Bio Eng Computer* 48 1261-1269

Adeli, H., Zhou, Z. and Dadmehr, N. (2003) 'Analysis of EEG Records in an Epileptic Patient Using Wavelet Transform', *J. Neurosci. Methods*, Vol. 123, no. 1, pp. 69-87.

Agarwal, R., Gotman, J., Flanagan, D. and Rosenblatt, B. (1998) 'Automatic EEG analysis uring long-term monitoring in the ICU', *Electroencephalography and Clinical Neurophysiology*, Vol. 107, no. 1, pp. 44-58.

Akay, M. (1998) *Time Frequency and Wavelets in Biomedical Signal Processing,* New York: Institute of Electrical and Electronics Engineers, Inc.

Allison. B.Z., Wolpaw, E.W., and Wolpaw, J.R. (2007) 'Brain computer interface systems: progress and prospects. In. E.Poll (Ed.), *British Review of Medical Device*, 2007, Vol. 4, n., 4, pp. 463-474.

Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P., and Elger, C. E.(2001) 'Indication of Non Linear Deterministic and Finite-Dimensional Structures in Time Series of Brain Electrical Activity: Dependence on Recording Region and Brain State', *Physical Review E,* Vol. 64, 061907.

Atwood, H. L. and MacKay, W. A. (1989) *Essentials of neurophysiology,* Toronto, Philadelphia : B. C. Decker.

Barreto, A. B., Taberner, A. M. and Vicente, L. M. (1996) 'Classification of spatio-temporal EEG readiness potentials towards the development of a brain-computer interface', *In Southeastcon '96. 'Bringing Together Education, Science and Technology', Proceedings of the IEEE*, pp.99 -102.

Bashashati, A., Fatourechi, M., Ward, R. K. and Birch G. E. (2007) 'A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals', *Journal of Neural engineering*, Vol. 4, no. 2, pp. R35–57.
BCI competition III, 2005, http://www.bbci.de/competition/iii

Berger, H. (1929) 'Uber das Elektroenkephalogram des Menschen', *Archives of Psychiatry*, Vol. 87, pp. 527-70.

Bian, Y., Zhao, L., Li, H., Yang, G., Shen, H. and Meng, Q. (2010) 'Research on Brain Computer Interface Technology Based on Steady State Visual Evoked Potentials, *Proceedings of the IEEE on Bioinformatics and Biomedical Engineering*, pp. 1- 4.

Blankertz, B., Muller, K..R., Krusienki, D. J., Schalk, G., Wolpaw, J.R., Schlogl, A., Pfurtscheller, S., Millan, J. De. R., Shrooder, M. and Birbamer, N. (2006) 'The BCI competition III: validating alternative approaches to actual BCI problems', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 14, no. 2, pp. 153-159.

Blankertz, B., Tomioka, R., Lemm, S., Kawanable, M. and Muller, K.R. (2008) 'Optimizing spatial filters for robust EEG single-trial analysis', *IEEE Signal Processing Magazine*, Vol. 25, no. 1, pp. 41-56.

Blankertz, B., Dornhege, G., Krauledat, M., Muller, K. and Curio, G. (2007) 'The non-invasive Berlin Brain-Computer interface: fast acquisition of effective performance in untrained subjects', *Neuroimage*, Vol. 37, pp. 539-550.

Blanchard G and Blankertz B (2004) 'BCI competition 2003-Data set IIa: Spatial patterns of self-controlled brain rhythm modulations' *IEEE Transactions on Biomedical Engineering*, Vol. 51, pp. 1062-1066.

Blume W, Lüders H, Mizrahi E, Tassinari C, van Emde Boas W, Engel J (2001) 'Glossary of descriptive terminology for ictal semiology: report of the ILAE task force on classification and terminology', *Epilepsia,* Vol. 42, no. 9, pp. 1212–1218.

Brown, M., Grundy, W., Lin, D., Cristianni, W., Sugnet, C., Furey, T., Ares, M. and Haussler, D. (2000) 'Analysis of gene expression data by using support vector machines', *Proc. Natl. Acad. Sci.* Vol. 97 (1), pp.262-267.

Brunelli, R. (2009) *Template Matching Techniques in Computer Vision: Theory and Practice*, Wiley, New York.

Burke D P, Kelly S P, Chazal P, Reilly R B and Finucane C (2005) 'A parametric feature extraction and classification strategy for brain-computer interfacing' *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* Vol. 13, pp. 12-17.

Burges, C. (1998) 'A tutorial on support vector machines for pattern recognition' *Data Mining and Knowledge Discovery,* Vol. 2. pp. 121-167.

Buzsaki, G. (2006) *Rhythms of the brain*, Oxford, Oxford University Press.

Caesarendra, W., Widodo, A. and Yang, B.S. (2010) 'Application of relevance vector machine and logistic regression for machine degradation assessment', *Mechanical Systems and Signal Processing*, Vol. 24, pp. 1161-1171.

Carlson, N. R. (2002a) *Foundations of physiological psychology*, 5th ed., Boston, Mass. London : Allyn and Bacon.

Cawley, G. C. and Talbot, N. L.C. (2008) 'Efficient approximate leave-one-out cross-validation for kernel logistic regression', *Mach Learn*, Vol. 71, pp. 243-264.

Carlson, N. R. (2002b) 'Structure and Functions of the Nervous System', *Foundations of physiological psychology*, Vol. 5th ed. Issue 3. Boston, Mass. London: Allyn and Bacon.

Chandaka, S., Chatterjee, A. and Munshi, S. (2009a) 'Cross-correlation aided support vector machine classifier for classification of EEG signals', *Expert System with Applications,* Vol. 36, pp. 1329-1336.

Chandaka, S., Chatterjee, A. and Munshi, S. (2009b) 'Support vector machines employing cross-correlation for emotional speech recognition', *Measurement*, Vol. 42, pp. 611-618.

Chiappa, S. and Millán, J.R. (2005) Data Set V<mental imagery, multi-class> [online]. Viewed 25 June 2009, http://ida.first.fraunhofer.de/projects/bci/competition_iii/desc_V.html

Cochran, W. G. (1977) *Sampling Techniques*, Wiley, New York.

Collura, T. F. (1993) 'History and evolution of electroencephalographic instruments and techniques', Clinical Neurophysiology, Vol. 10, Iss. 4, pp. 476-504.

Congedo, M., Lotte, F. and Lecuyer, A. (2006) 'Classification of movement intention by spatially filtered electromagnetic inverse solutions', *Physics in Medicine and Biology*, Vol. 51, no. 8, pp.1971-1989.

De Veaux, R. D., Velleman, P.F. and Bock, D.E. (2008) Intro Stats (3rd edition), Pearson Addison Wesley, Boston.

Duda, R.O., Hart, P.E. and Stork, D.G. (2001) *Pattern Classification,* 2nd edn. John Wiley & Sons, New York.

Dutta S, Chatterjee A and Munshi S (2009) 'An automated hierarchical gait pattern identification tool employing cross-correlation-based feature extraction and recurrent neural network based classification', *Expert systems,*Vol. 26, pp. 202-217

Dutta, S., Chatterjee, A. and Munshi, S. (2010) 'Correlation techniques and least square support vector machine combine for frequency domain based ECG beat classification', *Medical Engineering and Physics,* Vol. 32, no. 10, pp. 1161-1169.

EEG time series, 2005, [Online], Viewed 30 September 2008, http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html

Esen, H., Ozgen, F., Esen, M. and Sengur, A. (2009) 'Modelling of a new solar air heater through least-squares support vector machines', Expert System with applications, Vol. 36, pp.10673-10682.

Fan, J., Shao, C., Ouyang, Y., Wang, J., Li, S. and Wang, Z. (2006) 'Automatic seizure detection based on support vector machine with genetic algorithms', *SEAL 2006, LNCS 4247*, pp. 845-852.

Fawcett, T. (2006) 'An introduction to ROC analysis', *Pattern Recognition Letters,* Vol. 27, 861874.

Felzer, T. and Freisieben, B. (2003) 'Analysing EEG signals using the probability estimating guarded neural classifier', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 2, no. 4, pp. 361 -371.

Fisch B J (1999) *EEG premier: Basic principles of digital and analog EEG* (third edition), Elsevier publication.

Fletcher, R. (1987) *Practical methods of optimization,* Chichester and New York, John Wiley & Sons.

Garcia, G. N., Ebrahimi, T. and Vesin, J. M. (2003) 'Support vector EEG classification in the fourier and time-frequency correlation domains', *Proceedings of IEEE EMBS Conference on Neural Engineering*, pp 591-594.

Ghanbari A. A., Nazari Kousarrizi M .R., Teshnehlab M. and Aliyari, M. (2009) 'An evolutionary artefact rejection method for brain computer interface using ICA', *Int. J. of Elect. & Comp. Scien.*, Vol. 9, no. 9, pp. 461-466.

Graimann, B., Allison, B., Mandel, C., Luth, T., Valbuena, D. and Graser, A. (2009) 'Non-invasive brain-computer interfaces for semi-autonomous assistive devices', *Robust Intelligent Systems*, Chapter 6, pp. 113-138.

Graimann, B., Allison, B., Mandel, C., Luth, T., Valbuena, D. and Graser, A. (2008) 'Non-invasive brain-computer interfaces for semi-autonomous assistive devices', *Robust Intelligent Systems*, Spring London Ltd., pp. 113-137.

Gray, F. J. (2002) *Anatomy for the medical clinician*, first edition, Shannon Books Pty Ltd, Victoria, Australia.

Grosse-Wentrup, M., Liefhold, C., Gramann, K. and Buss, M. (2009) 'Beamforming in noninvasive brain-computer interfaces', *IEEE Transactions on Biomedical Engineering*, Vol. 56, no. 4, pp. 1209-1219.

Guger C, Schlogl A, Neuper C, Walterspacher C, Strein D, Pfurtscheller T and Pfurtscheller G (2001) 'Rapid prototyping of an EEG-based brain-computer interface (BCI)' *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* Vol. 9 pp.49-58.

Guler, N.F., Ubeylli, E. D. and Guler, I. (2005) 'Recurrent neural networks employing Lyapunov exponents for EEG signals classification', *Expert System with Applications* Vol. 29, pp.506-514.

Guler, I. And Ubeyli, E.D. (2007) 'Multiclass support vector machines for EEG-signal classification', *IEEE Transactions on Information Technology in Biomedicine*, Vol. 11, no. 2, pp. 117-126.

Guler, I. and Ubeyli, E. D. (2005) 'Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficient', Journal of Neuriscience Methods, Vol. 148, pp. 113-121.

Guo, X.C., Liang, Y.C., Wu, C.G. and Wang C.Y. (2006) 'PSO-based hyper-parameters selection for LS-SVM classifiers', *ICONIP 2006, Part II*, LNCS 4233,

pp. 1138-1147.

Guo, L., Rivero, D., Seoane, J.A. and Pazos, A. (2009) 'Classification of EEG signals using relative wavelet energy and artificial neural networks', *GCE,* 12-14.

Hanbay, D. (2009) 'An expert system based on least square support vector machines for diagnosis of the valvular heart disease', *Expert System with Applications,* Vol. 36, pp.4232-4238.

Hammon, P.S. and de Sa. V.R. (2007) 'Preprocessing and meta-classification for brain-computer interfaces', *IEEE Transactions on Biomedical Engineering*, Vol. 54, no. 3, pp. 518–525.

Haselsteiner, E. and Pfurtscheller, G. (2000) 'Using Time-Dependant Neural Networks for EEG classification', *IEEE Transactions on Rehabilitation Engineering*, Vol. 8, pp. 457-463.

Hazarika, N., Chen, J.Z., Tsoi, A.C., and Sergejew, A. (1997) 'Classification of EEG Signals Using the Wavelet Transform', *Signal Process*, Vol. 59 (1), pp.61-72.

Hieftje G M, Bystroff R I and Lim R (1973) 'Application of correlation analysis for signal-to-noise enhancement in flame spectrometry: use of correlation in determination of rhodium by atomic fluorescence' *Analytical Chemistry,* Vol. 45, pp. 253-258.

Hosmer, D.W. and Lemeshow, S. (1989) *Applied logistic regression*, Wiley, New York.

Hoya, T., Hori, G., Bakardjian, H., Nishimura, T., Suzuki, T., Miyawaki, Y., Funase, A. and Cao, J. (2003) 'Classification of Single Trial EEG signals by a combined Principal and Independent Component Analysis and Probabilistic Neural Network Approach', *In Proceedings ICA2003*, pp. 197-202.

Islam, M. N. (2004) *An introduction to statistics and probability*, 3rd ed., Mullick &

brothers, Dhaka New Market, Dhaka-1205, pp. 160-161.

Islam, M. N. (2007) *An introduction to sampling methods: theory and applications*, revised ed., Book World, Dhaka New Market & P.K. Roy road, Bangla Bazar, Dhaka-1100.

Jahankhani, P., Kodogiannis, V. and Revett, K. (2006) 'EEG Signal Classification Using Wavelet Feature Extraction and Neural Networks', *IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing (JVA'06).*

Joshi D and Anand S 2010 Cyclogram and cross correlation: A comparative study to quantify gait coordination in mental state *J. Biomedical Science and Engineering* 3 322-326.

Jasper, H. H. (1958a) 'The ten-twenty electrode system of the International Federation', *Electroencephalogram. Clinical. Neurophysiology*. Vol. 10, pp: 367-380.

Jasper, H. H. (1958b) 'Report of the committee on methods of clinical examination in electroencephalography: The ten-twenty electrode system of the International Federation' *Electroencephalography and Clinical Neurophysiology*, Vol. 10, Iss. 2, pp. 371-375.

Kamousi, B., Liu, Z. and He., B. (2005) 'Classification of motor imagery tasks for brain-computer interface applications by means of two equivalent dipoles analysis', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 13, pp.166-171.

Kaneswaran, K., Arshak, K., Burke, E., Condron, J. (2010) 'Towards a brain Controlled Assistive Technology for Powered Mobility', *Proceedings of the IEEE EMBS*, pp. 4176-4180.

Kang, H., Nam, Y. and Choi, S. (2009) 'Composite common spatial pattern for subject-to-subject transfer', *IEEE Signal Processing letters,* Vol. 16, no. 8, pp. 683-

686.

Kay, S. M. (1988) *Modern Spectral Estimation: Theory and Application,* New Jersey: Prentice Hall.

Kay, S. M., & Marple, S. L. (1981) 'Spectrum analysis – A modern perspective', *Proceedings of the IEEE*, Vol. 69, pp.1380–1419.

Kayikcioglu, T. and Aydemir, O. (2010) 'A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data', *Pattern Recognition Letters,* Vol. 31, pp. 1207-1215.

Kostov, A. and Polak, M. (2000) 'Parallel man-machine training in development of EEG-based cursor control', *IEEE Transactions on Rehabilitation Engineering*, Vol. 8, no. 2, pp. 203-205.

Kousarrizi, M. R. N., Ghanbari, A. A., Teshnehlab, M. , Aliyari, M. and Gharaviri, A. (2009) 'Feature Extraction and Classification of EEG Signals using Wavelet Transform, SVM and Artificial Neural Networks for Brain Computer Interfaces', *2009 International Joint Conference on Bioinformatics, System Biology and Intelligent Computing*, pp.352-355.

Kouijzer, M.E.J., et al. (2009) 'Long-term effect of neurofeedback treatment in autosim', *Research in Autism Spectrum Disorders*, Vol. 3, no. 2, pp. 469-501.

Kübler, A., Mushahwar, V.K., Hochberg, L.R. and Donoghue J.P. (2006) 'BCI meeting 2005-workshop on clinical issues and applications' *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol.14, no.2, pp.131–134.

Kutlu, Y., Kuntalp, M. and Kuntalp, D. (2009) 'Optimizing the Performance of an MLP classifier forthe Automatic detection of Epileptic spikes', *Expert System with applications*, Vol. 36, pp.7567-7575.

Last T, Nugent C D and Owens F J 2004 Multi-component based cross correlation beat detection in electrocardiogram analysis *Biomedical Engineering Online,* 3:26, doi:10.1186/1475-925X-3-26.

Lee, H. and Choi, S. (2003) 'PCA+HMM+SVM for EEG pattern classification', *Proceedings of the Seventh International IEEE Symposium on Signal Processing and Its Applications*, pp 541- 544.

Lemm S, Blankertz B, Curio G and Muller K R (2005) 'Spatio-spatial filters for improved classification of single trial EEG' *IEEE Transactions on Biomedical Engineering*, Vol. 52, pp. 1541-1548.

Li, X. L., He, Y. and Wu, C. Q. (2008) 'Least square support vector machine analysis for the classification of paddy seeds by harvest year', *ASABE*, Vol. 51, no. 5, pp. 1793-1799.

Liao, J.G. and Chin, K.V. (2007) 'Logistic regression for disease classification using microarray data: model selection in a large *p* and *n', Bioinformatics,* Vol. 23, no. 15, pp. 1945-1951.

Lima, C. A. M., Coelho, A. L.V. and Chagas, S. (2009) 'Automatic EEG signal classification for epilepsy diagnosis with Relevance Vector Machines', *Expert Systems with Applications*, Vol. 36, pp. 10054–10059.

Lin, C.J. and Hsieh, M.H. (2009) 'Classification of mental task from EEG data using neural networks based on particle swarm optimization', *Neurocomputing*, Vol. 72, pp. 1121-1130.

Long, J., Li Y. and Yu, Z. (2010) 'A semi-supervised support vector machine approach for parameter setting in motor imagery-based brain computer interfaces', *Cognitive Neurodynamics,* Vol. 4, pp. 207-216.

Lotte, F. and Guan, C. (2010) 'Spatially regularized common spatial patterns for EEG classification', *Inria-00447435* (25 Jan 2010) version 2.

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

Lotte, F. (2009) *Study of electroencephalographic signal processing and classification techniques towards the use of brain-computer interfaces in virtual reality applications*, PhD thesis.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F. and Arnaldi B. (2007) 'A review of classification algorithms for EEG-based brain-computer interfaces', *Journal of Neural Engineering*, Vol. 4, pp. R1–R13.

LS-SVMlab toolbox (version 1.5)-online, http://www.esat.kuleuven.ac.be/sista/lssvmlab/

Lu, H., Plataniotis, K.N. and Venetsanopoulos, A.N. (2009) 'Regularized common spatial patterns with generic learning for EEG signal classification', *31st Annual International Conference of the IEEE EMBS Minneapolis*, *Minnesota, USA, September 2-6*, 2009, pp.6599-6602.

Lu, H., Eng, H. L. , Guan, C., Plataniotis, K. N. and Venetsanopoulos, A. N. (2010) 'Regularized common spatial patterns with aggregation for EEG classification in small-sample setting', *IEEE Transactions on Biomedical Engineering,* Vol. 57, no. 12 pp. 2936-2945.

Mason S.G. and Birch G.E. (2003) 'A general framework for brain-computer interface design' *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 11, no. 1, pp.70–85.

MATLABArsenal-online, http://www.informedia.cs.cmu.edu/yanrong/MATLABArsenal/MATLABArsenal.zip

Makinac, M. (2005) 'Support Vector Machine Approach for Classification of Cancerous Prostate Regions', *World Academy of Science, Engineering and Technology*, 7.

McFarland, D.J. and Wolpaw, J. R. (2011) 'Brain-computer interfaces for communication and control', *Communications of the ACM*, Vol.54, no. 5, pp. 60-66.

Md Norani, N.A., Mansor, W. and Khuan L.Y. (2010) 'A Review of Signal Processing in Brain Computer Interface System', *2010 IEEE EMBS Conference on Biomedical Engineering & Sciences (IECBES 2010), Kuala Lumpur, Malaysia, 30th November - 2nd December 2010,* pp. 443-449.

Meng J, Liu G, Huang G and Zhu X 2009 Automated selecting subset of channels based on CSP in motor imagery brain-computer system *Proceedings of the 2009 IEEE International Conference on Robotics and Bioinformatics* December 19-23 2009 Guilin China 2290-2294

Millán, J.R (2004) 'On the need for online learning in brain-computer interfaces', *Proc. 2004 Int. Joint Conf. Neural networks*, Vol. 4, pp. 2877-2882.

Millan, J. R., Renkens, F., Mourino, J. and Gerstner, W. (2004) 'Noninvasive brain-actuated control of a mobile robot by human EEG', *IEEE Transactions on Biomedical Engineering*, Vol. 51, no. 6, pp.1026-1 033.

Ming, D., Zhu, Y. , Qi, H. , Wan, B., Hu, Y. and Luk, K. (2009) 'Study on EEG-based Mouse System by using Brain-Computer Interface', *Proceedings of the IEEE on Virtual Environments, Human-Computer Intel1aces and Measurements Systems*, pp 236 - 239.

Mrowski, P., Madhavan, D., LeCun, Y. and Kuzniecky, R. (2009) 'Classification of patterns of EEG synchronization for seizure prediction'**,** *Clinical Neurophysiology*, Vol. 120, pp. 1927-1940.

Muller K R, Krauledat M, Dornhege G, Curio G and Blankertz B (2004) 'Machine learning techniques for brain-computer interfaces', *Biomedical Engineering (Biomedizinische Technik)*, Vol. 49, pp.11-22.

Niedermeyer E. and Lopes da Silva F. (2005) *Electroencephalography: basic principles, clinical applications, and related fields*, Lippincott Williams & Wilkins, ISBN 0781751268, 5th edition, 2005.

Oostenveld, R. & Praamstra, P. (2001) 'The five percent electrode system for high-resolution EEG and ERD measurements', *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, Vol. 112, no. 4, pp.713-719.

Palaniappan, R., Paramesran, R., Nishida, S. and Saiwaki, N. (2002) 'A new brain-computer interface design using Fuzzy ART MAP', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 10, pp.140-148.

Penny, W. D., Roberts, S. J., Curran, E. A. and Stokes, M. J. (2000) 'EEG-based communication: A Pattern Recognition Approach', *IEEE Transactions on Rehabilitation Engineering*, Vol. 8, no. 2, pp. 214-215.

Polat, K. and Gunes, S. (2007) 'Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform', *Applied Mathematics and Computation,* 187 1017-1026.

Pfurtscheller, G., Neuper, C., Schlogl, A. and Lugger, K. (1998) 'Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters', *IEEE Transactions on Rehabilitation Engineering,* Vol. 6, no. 3, pp. 316-325.

Pfurtscheller G, Brunner C, Schlogl A and da Silva F H L (2006) 'Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks' *Neuroimage*, Vol. 31, pp. 153-159.

Pfurtscheller, G., Muller-Putz, G.R., Schlogl, A., Graimann, B., Leeb, R., Brunner, C., Keinrath, C., Lee, F., Townsend, G., Vidaurre, C. and Neuper, C. (2006b) '15 years of BCI research at Graz university of technology: current projects', *IEEE*

Transactions on Neural Systems and Rehabilitation Engineering: A publication of the IEEE Engineering in Medicine and Biology Society, Vol. 14, pp. 205-210.

Pfurtscheller, G., & Neuper, C. (2001) 'Motor imagery and direct brain-computer communication', *Proceedings of the IEEE,* Vol. 89, no. 7, pp.1123-1134.

Pfurtscheller, G., Flotzinger, D. and Kalcher, J. (1993) 'Brain-computer interface-a new communication device for handicapped persons', *Journal of Microcomputer Application*, Vol. 16, pp.293-299.

Pravin Kumar, S., Sriraam, N., Benakop, P.G. and Jinaga, B.C. (2010) 'Entropies based detection of epileptic seizures with artificial neural network classifiers', *Expert System with Applications*, Vol. 37, pp. 3284–3291.

Proakis, J. G., & Manolakis, D. G. (1996) *Digital Signal Processing Principles, Algorithms, and Applications*, New Jersey: Prentice Hall.

Purves, D., Augustine, G.J., Fitzpatrick, D., Katz, L.C. Lamantia, A.S. and McNamara, J.O. (2004) *Neuroscience*, Sinauer associates, third edition, Inc. Publishers, Sunderland, Massachusetts, USA.

Qin L and He B (2005) 'A wavelet-based time-frequency analysis approach for classification of motor imagery for brain-computer interface applications' *Journal of Neural Engineering*, Vol. 2, pp. 65-72.

Qin, L. Ding, L. and He., B. (2004) 'Motor imagery classification by means of source analysis for brain computer interface applications', *Journal of Neural Engineering*, Vol. 1, no. 3, pp. 135-141.

Quyen, M. L. V., Navarro, V., Baulac, M., Renault, B., Martinerie, J. (2000) 'Anticipation of epileptic seizures from standard EEG recordings', *The Lancet*, Vol. 361, Issue 9361, pp 970-971.

Rahayu, S.P., Purnami, S.W., Embong, A. and Zain, J. M. (2009) 'Kernel logistic regression-linear for leukemia classification using high dimensional data', *JUTI*, Vol.

7, no. 3, pp. 145-150.

Ripley, B.D. (1996) *Pattern recognition and neural networks. Cambridge*, Cambridge University Press.

Ripley data-online, http://www.stats.ox.ac.uk/pub/PRNN/

Rong H J , Sundararajan N, Huang G B and Saratchandran P (2006) 'Sequential adaptive fuzzy inference system (SAFIS) for nonlinear system identification and prediction', *Fuzzy Sets Syst.*, Vol. 157, pp. 1260-1275.

Ryali, S., Supekar, K., Abrams, D. A. and Menon, V. (2010) 'Sparse logistic regression for whole-brain classification of fMRI data', *NeuroImage*, Vol. 51, pp. 752-764.

Saggiol, G., Cavallo, P., Ferretti, A., Garzoli, F., Quitadamo, L. R., Marciani, M. G., Giannini, F. and Bianchi, L. (2009) 'Comparison of Two Different Classifiers for Mental Tasks-Based Brain-Computer Interface: MLP Neural Networks vs. Fuzzy Logic', *Proceedings of the IEEE WoWMoM*, pp. 1-5.

Sanei, S. and Chambers, J. A. (2007) *EEG Signal Processing*, John Wiley & Sons, Ltd., 2007.

Sander T H, Leistner S, Wabnitz H, Mackert B M, Macdonald R and Trahms L (2010) 'Cross-correlation of motor activity signals from dc-magnetoencephalography, near-infrared spectroscopy and electromyography', *Computational Intelligence and Neuroscience*, doi:10.1155/2010/78527.

Sanei, S. and Chambers, J. (2007) *EEG signal processing*, John Wiley & Sons, Ltd.
Sample size calculator-online, http://www.surveysystem.com/sscalc.htm

Satti, A. R., Coyle, D., Prasad, G. (2009) 'Spatio-Spectral & Temporal Parameter Searching using Class Correlation Analysis and Particle Swarm Optimisation for

Brain Computer Interface', *Proceedings of the IEEE on Systems, Man, and CybernetiCS*, pp 1731-1735.

Schlogl A, Neuper C and Pfurtscheller G (2002) 'Estimating the mutual information of an EEG-based brain-computer interface' *Biomed. Tech. (Berl)* Vol. 47, pp. 3-8.

Schlogl A, Lee F, Bischof H and Pfurtscheller G (2005) 'Characterization of four-class motor imagery EEG data for the BCI-competition' *Journal of Neural Engineering*, Vol. 2, pp. L14-22.

Sellers, E. W. and Donchin, E. (2006) 'A P300-based brain-computer interface: initial tests by ALS patients', *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, Vol. 117, pp. 538-548.

Silver, A.E., Lungren, M.P., Johnson, M.E., O'Driscoll, S.W.,An, K.N and Hughes, R.E.(2006) 'Using support vector machines to optimally classify rotator cuff strength data and quantify post-operative strength in rotator cuff tear patients', *Journal of Biomechanics,* Vol. 39, pp. 973-979.

Siuly, Li, Y. and Wen, P. (2009) 'Classification of EEG signals using Sampling Techniques and Least Square Support Vector Machines', *RSKT 2009*, LNCS 5589, pp. 375-382.

Siuly, Li, Y. and Wen, P. (2010) 'Analysis and classification of EEG signals using a hybrid clustering technique', *Proceedings of the 2010 IEEE/ICME International Conference on Complex Medical Engineering (CME2010)*, pp. 34–39.

Siuly, Li, Y., Wu, J. and Yang, J. (2011a) 'Developing a Logistic Regression Model with Cross-Correlation for Motor Imagery Signal Recognition', *The 2011 IEEE International Conference on Complex Medical Engineering (CME 2011)*, Harbin, Heilongjiang, China, 22-25 May 2011, pp.502-507.

Siuly, Li, Y. and Wen, P. (2011b) 'Identification of Motor Imagery Tasks through CC-LR Algorithm in Brain Computer Interface', *International Journal of Bioinformatics Research and Applications,* in Press.

Siuly, Li, Y. and Wen, P. (2011c) 'Clustering technique-based least square support vector machine for EEG signal classification', *Computer Methods and Programs in Biomedicine*, Vol. 104, Issue 3, pp. 358-372.

Siuly, Li, Y. and Wen, P. (2011d) 'EEG signal classification based on simple random sampling technique with least square support vector machines', *International journal of Biomedical Engineering and Technology*, Vol. 7, no. 4, pp. 390-409.

Siuly and Li, Y. (2012) 'Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain computer interface'**,** *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* DOI (identifier) 10.1109/TNSRE.2012.2184838, in press.

Sleigh, J.W., Steyn-Ross, D.A., Steyn-Ross, M.L., Grant, C. and Ludbrook, G. (2004) 'Cortical entropy changes with general anaesthesia: theory and experiment', *Physiol Meas,*Vol. 25, pp. 921-934.

Song, L. & Epps, J. (2007). Classifying EEG for brain-computer interface: learning optimal filters for dynamical features. *Computational Intelligence and Neuroscience,* Article ID 57180, 11 pages, doi:10.1155/2007/57180.

Stam, C.J., Pijn, J.M.P., Suffczynski, P. and da Silva, F.H. Lopes (1999) 'Dynamics of the human alpha rhythm: evidence for non-linearity', *Clin Neurophysiol*, Vol. 110, pp. 1801-1813.

Stoica, P., & Moses, R. (1997) *Introduction to Spectral Analysis,* New Jersey: Prentice Hall.

Subasi, A. and Ercelebi, E. (2005a) 'Classification of EEG signals using neural network and logistic regression', *Computer Methods and Programs in Biomedicine*, Vol. 78, pp. 87-99.

Subasi, A., Alkan, A., Kolukaya, E. and Kiymik, M. K. (2005b) 'Wavelet neural network classification of EEG signals by using AR model with MLE preprocessing', *Neural Networks*, Vol. 18, pp. 985-997.

Subasi, A. (2007) 'EEG signal classification using wavelet feature extraction and a mixture of expert model'**,** *Expert System with Applications, V*ol. 32, pp.1084-1093.

Sun, S. and Zhang, C. (2005) 'Learning on-line classification via decorrelated LMS algorithm: application to brain computer interfaces', A. Hoffman, H. Motoda, and T. Scheffer (Eds): *DS 2005 (2005), LNAI 3735*, pp. 215-226.

Sun, S. and Zhang, C. (2006) 'Adaptive feature extraction for EEG signal classification', Med. Biol. Eng. Comput., Vol. 44, no. 10, pp. 931-935.

Sun, S. (2007a) 'An improved random subspace method and its application to EEG signal classification', *MCS 2007*, LNCS 4472, pp. 103-112.

Sun, S., Zhang, C. and Zhang, D. (2007b) 'An experimental evaluation of ensemble methods for EEG signal classification', *Pattern Recognition Letters,* Vol. 28, 2157-2163.

Sun, S., Zhang, C. and Lu, Y. (2008) 'The random electrode selection ensemble for EEG signal', Pattern recognition, Vol. 41, pp. 1663-1675.

Suykens, J.A.K., Gestel, T.V., Brabanter, J.D., Moor, B.D. and Vandewalle, J. (2002) *Least Square Support Vector Machine*, World Scientific, Singapore.

Suykens, J.A.K., and Vandewalle, J. (1999) 'Least Square Support Vector Machine classifier', *Neural Processing Letters*, Vol. 9, no. 3, 293-300.

Thissen, U., Ustun, B., Melssen, W. J. and Buydens, L. M. C. (2004) 'Multivariate calibration with least-square support vector machines', *Analytical Chemistry*, Vol. 76, pp. 3099-3105.

Thomas, K. P., Guan, C., Lau, C. T., Vinod, A. P. and Ang, K. K. (2009) 'A new discriminative common spatial pattern method for motor imagery brain-computer interfaces', *IEEE Transactions on Biomedical Engineering,* Vol. 56, no.11, pp2730-2733.

Ting W, Guo-Zheng Y, Bang-Hua Y and Hong S (2008) 'EEG feature extraction based on wavelet packet decomposition for brain computer interface' *Measurement*, Vol. 41, pp. 618-625.

Ubeyli, E.D. (2010) 'Least Square Support Vector Machine Employing Model-Based Methods coefficients for Analysis of EEG Signals', *Expert System with Applications*. 37 233-239.

Ubeyli, E. D. (2009a) 'Decision support systems for time-varying biomedical signals: EEG signals classification', *Expert Systems with Applications,* Vol. 36, pp. 2275–2284.

Ubeyli, E. D. (2009b) 'Statistics over features: EEG signals analysis', *Computers in Biology and Medicine,* Vol. 39, pp.733 – 741.

Ubeyli, E.D. (2008) 'Wavelet/mixture of experts network structure for EEG signals classification' *Expert System with Applications* Vol. 34, pp.1954-1962.

Ubeyli, E. D., & Guler, I. (2004) 'Spectral broadening of ophthalmic arterial Doppler signals using STFT and wavelet transform. *Computers in Biology and Medicine*, Vol. 34, no. 4, pp. 345–354.

Ubeyli, E. D., & Guler, I. (2003) 'Comparison of eigenvector methods with classical and model-based methods in analysis of internal carotid arterial Doppler signals', *Computers in Biology and Medicine*, Vol. 33, no.6, pp. 473–493.

Vapnik, V. (1995) *The nature of statistical learning theory,* Springer-Verlag, New York.

Vapnik, V. (1998) *Statistical learning theory*, John Wiley, New York.

Vaughan, T. M. , Heetderks, W. J. , Trejo, L. J. , Rymer, W.Z., Weinrich, M., Moore, M.M., Kubler, A., Dobkin, B. H., Birbaumer, N., Donchin, E., Wolpaw, E. W. and Wolpaw, J. R. (2003) 'Brain-computer interface technology: a review the second international meeting', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol.11, no. 2, pp. 94-109.

Vellistte, M., Perel, S., Spalding, M.C., Whitford, A.S. and Schwartz, A.B. (2008) 'Cortical control of a prosthetic arm for self-feeding', *Nature*, Vol. 453, pp. 1098-1101.

Wang, B. Y., Gao, X., Hong, B., Jia, C. and Gao, S. (2008) 'Brain-computer interfaces based on visual evoked potentials', *IEEE Engineering in Medicine and Biology Magazine*, Vol. 27, no. 5, pp. 64-71.

Wang, S. & James, C. J., (2007) 'Extracting rhythmic brain activity for brain-computer interfacing through constrained independent component analysis', *Computational Intelligence and Neuroscience,* Article ID 41468, 9 pages, doi:10.1155/2007/41468.

Wang, T., Deng, J. and He, B. (2004a) 'Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns', *Clinical Neurophysiology,* Vol. 115, pp. 2744-2753.

Wang Y, Zang Z, Li Y, Gao S and Yang F (2004b) 'BCI competition 2003-Dataset iv: An algorithm based on CSSD and FDA for classifying single-trial EEG', *IEEE Transactions on Biomedical Engineering*, Vol. 51, pp. 1081-1086.

Welch, P. D. (1967) 'The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short modified periodograms', *IEEE Transactions on Audio and Electroacoustics*, AU-15, 70–73.

Wolpaw, J. R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G. and Vaughan, T.M.

(2002) 'Brain-computer interfaces for communication and control', *Clinical Neurophysiology,* Vol. 113, pp. 767-791.

Wolpaw, J. R. (2003) 'Brain-computer interface technology: a review the second international meeting', *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* Vol. 11, no. 2, pp. 94-109.

Wren, T. A. L., do, K. P., Rethlefsen, S. A. and Healy, B. (2006) 'Cross-correlation as a method for comparing dynamic electromyography signals during gait', *Journal of Biomechanics*, Vol. 39, pp. 2714-2718.

Wu, W., Gao, X. , Hong, B. and Gao, S. (2008) 'Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL)', *IEEE Transactions on Biomedical Engineering,* Vol. 55, no. 6, pp. 1733-1743.

Xie, X. J., Pendergast, J. and Clarke, W. (2008) 'Increasing the power: a practical approach to goodness-fit test for logistic regression models with continuous predictors', *Computational Statistics and Data Analysis,* Vol. 52, pp. 2703-2713.

Xie, L., Ying, Y. and Ying, T. (2009) 'Classification of tomatoes with different genotypes by visible and short-wave near-infrared spectroscopy with least-square support vector machines and other chemometrics', *Journal of Food Engineering*, Vol. 94, pp. 34-39.

Yamaguchi, T., Nagata, K., Truong, P. Q., Fujio, M. and Inoue, K. (2008) 'Pattern Recognition of EEG Signal During Motor Imagery by Using SOM', *International Journal of Innovative Computing*, *Information and Control*, Vol.4, no.IO, pp. 2617-2630.

Yong, X, Ward, R.K. and Birch, G.E. (2008) 'Sparse spatial filter optimization for EEG channel reduction in brain-computer interface', *ICASSP 2008*, pp.417-420.

Yue, L., Baojun, Y., Peng, L. & Shizhe, L. (2003) 'Detecting technique of weak periodic pulse signal via synthesis of cross-correlation and chaotic system', *Journal of Electronics*, Vol. 2, no. 5, pp. 397-400.