Health Information Science

Siuly Siuly
Yan Li
Yanchun Zhang

EEG Signal Analysis and Classification

Techniques and Applications

Springer

Health Information Science

With the development of database systems and networking technologies, Hospital Information Management Systems (HIMS) and web-based clinical or medical systems (such as the Medical Director, a generic GP clinical system) are widely used in health and clinical practices. Healthcare and medical service are more data-intensive and evidence-based since electronic health records are now used to track individuals' and communities' health information. These highlights substantially motivate and advance the emergence and the progress of health informatics research and practice. Health Informatics continues to gain interest from both academia and health industries. The significant initiatives of using information, knowledge and communication technologies in health industries ensures patient safety, improve population health and facilitate the delivery of government healthcare services. Books in the series will reflect technology's cross-disciplinary research in IT and health/medical science to assist in disease diagnoses, treatment, prediction and monitoring through the modeling, design, development, visualization, integration and management of health related information. These technologies include information systems, web technologies, data mining, image processing, user interaction and interfaces, sensors and wireless networking, and are applicable to a wide range of health-related information such as medical data, biomedical data, bioinformatics data, and public health data.

More information about this series at http://www.springer.com/series/11944

Siuly Siuly · Yan Li · Yanchun Zhang

EEG Signal Analysis and Classification

Techniques and Applications

Springer

Siuly Siuly
Centre for Applied Informatics, College of Engineering and Science
Victoria University
Melbourne, Victoria
Australia

Yan Li
School of Agricultural, Computational and Environmental Sciences, Faculty of Health, Engineering and Sciences
University of Southern Queensland
Toowoomba, Queensland
Australia

Yanchun Zhang
Centre for Applied Informatics, College of Engineering and Science
Victoria University
Melbourne, Victoria
Australia

and

School of Computer Science
Fudan University
Shanghai
China

ISSN 2366-0988 ISSN 2366-0996 (electronic)
Health Information Science
ISBN 978-3-319-47652-0 ISBN 978-3-319-47653-7 (eBook)
DOI 10.1007/978-3-319-47653-7

Library of Congress Control Number: 2016959570

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To people who are suffering from
neurological diseases and disorders

Preface

Electroencephalography (EEG) has played a prominent role in brain studies, mental and brain diseases' and disorders' diagnosis, and treatments in medical fields. The examination of EEG signals has been recognized as the most preponderant approach to the problem of extracting knowledge of brain dynamics. At present, EEG recordings are widely used for *epilepsy diagnosis* and for *brain computer interfaces (BCIs)*. The main use of an EEG is to detect and investigate epilepsy that causes seizures. In BCI systems, EEG signals help to restore sensory and motor functions in patients who have severe motor disabilities. Generally, the vast amounts of multi-channel EEG signals are visually analysed by an expert to identify and understand abnormalities within the brain and how they propagate.

The visual inspection approach for such huge EEG data is not a satisfactory procedure for accurate and reliable diagnosis and interpretation as this process is time-consuming, burdensome, reliant on expensive human resources, and subject to error and bias. An extensive amount of research is required for automatic diagnosis of epileptic seizures and also for the automatic identification of mental states to help motor disabled people through BCI systems. Hence, in this monograph, we aim to develop advanced methods for the analysis and classification of epileptic EEG signals and also for the identification of mental states in BCI applications.

A number of edited books have been published in these two areas (but never have they been presented together in one book) and those books present common signal processing techniques such as, wavelet transformation, Fourier transformation for EEG data analysis. This book, however, presents some different EEG signal analysis approaches; combining statistical techniques (e.g. random sampling, optimum allocation, etc.) and machine learning methods. In this book, the authors present their methods that provide better performance compared to the existing methods.

The book consists of four parts with 13 chapters. Part I provides a basic overview of EEG signals including concept, generation procedure, characteristics, nature and abnormal patterns. This part also provides a discussion of the different applications of EEG signals for the diagnosis of brain diseases and abnormalities. In addition, we provide the aims of this book, description of analyzed datasets used in

the research, performance evaluation measures and a short review of commonly used methods in EEG signal classification in Part I. Part II presents our developed techniques and models for the detection of epileptic seizures through EEG signal processing. Implementation of these proposed methods in real-time databases will also be highlighted. In Part III, we introduce the methods for identifying mental states from EEG data designed for BCI systems and their applications in several benchmark datasets. We also report the experimental procedure with the results of each methodology. Finally, we provide an overall discussion on EEG signal analysis and classification in Part IV. This part gives a summary discussion on the developed methods, future directions in the EEG signal analysis area and conclusions with suggestions for future research.

Melbourne, Australia — Siuly Siuly
Toowoomba, Australia — Yan Li
Melbourne, Australia and Shanghai, China — Yanchun Zhang

Contents

Part I
Introduction

Chapter 1
Electroencephalogram (EEG) and Its Background

1.1 What Is 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's electrical activity, through the EEG records, is one of the most important tools for the diagnoses of neurological diseases, such as epilepsy, brain tumours, head injury, sleep disorders, dementia and monitoring the depth of anaesthesia during surgery (Hazarika et al. 1997; Adeli et al. 2003). It is also helpful for the treatment of abnormalities, behavioural disturbances (e.g., Autism), attention disorders, learning problems and language delay.

The first EEG recording machine was 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 representations 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, anaesthesia and epilepsy. This was a revolutionary idea that helped create a new branch of medical science called neurophysiology.

During the EEG test a number of small disks called electrodes are placed in different locations on the surface of the scalp with temporary glues. 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. Figure 1.1 presents an example of how electrodes are placed on the scalp during the recording of EEG signals and EEG signals are displayed on a computer screen. The electrodes detect tiny electrical charges that result from the activity of the brain cells. The charges are amplified and appear as a graph on a computer screen, or as a recording that may be printed out on paper. An expert then interprets the reading. EEG recordings,

S. Siuly et al., *EEG Signal Analysis and Classification*,
Health Information Science, DOI 10.1007/978-3-319-47653-7_1

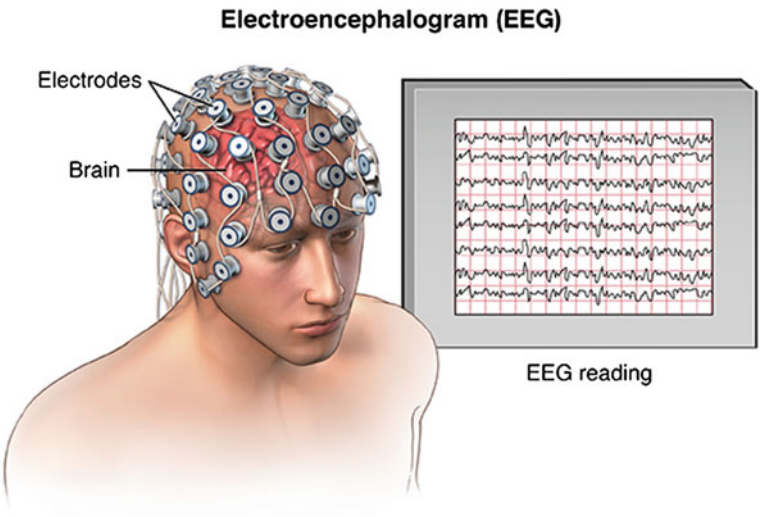


Fig. 1.1 An illustration of EEG recording (Ref. EEG, Saint Luke's Health System)

depending on their use can have from 1 to 256 electrodes recorded in parallel. This is called multichannel EEG recordings. One pair of electrodes usually makes up a channel. Each channel produces a signal during an EEG recording.

There are two types of EEGs, 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 10–20 mV when measured with subdural electrodes such as needle electrodes. Since the architecture of the brain is nonuniform and the cortex is functionally organized, 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 1958). The "10" and "20" represent actual distances between neighbouring electrodes are either 10 or 20% of the total front-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 *inon* which is the bony prominence at the base skull on the midline at the back of the head. Figure 1.2 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

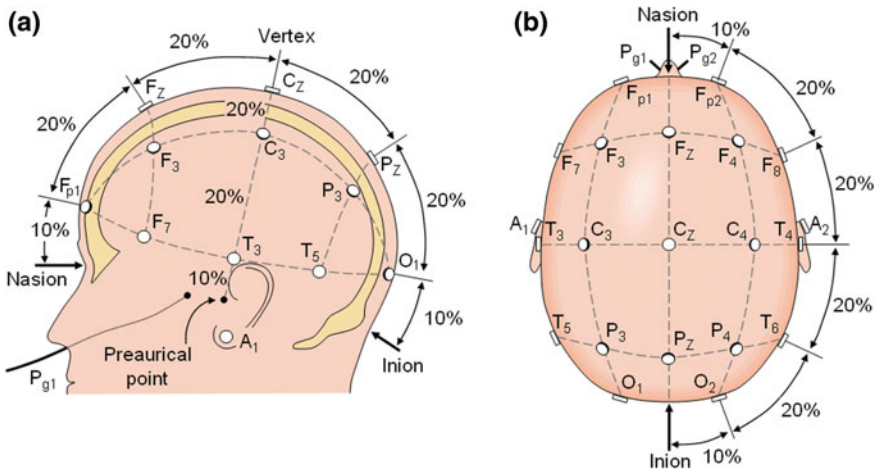


Fig. 1.2 The international 10–20 electrode placement system (Campisi 2012; Jasper 1958; 10/20 positioning/DIY tDCS)

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.

Bipolar montage: One pair of electrodes usually makes up a channel as shown in Fig. 1.3. Each channel (waveform) represents the difference between two adjacent electrodes (Niedermeyer and Lopes Da Silva 2005; Fisch 1999) as shown by Fig. 1.4. The entire montage consists of a series of these channels. For example, we present a diagram of a bipolar montage in Fig. 1.5, where 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). In Fig. 1.6, electrode A2 is considered as the reference electrode. 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). An illustration of the average reference montage is given by Fig. 1.7.

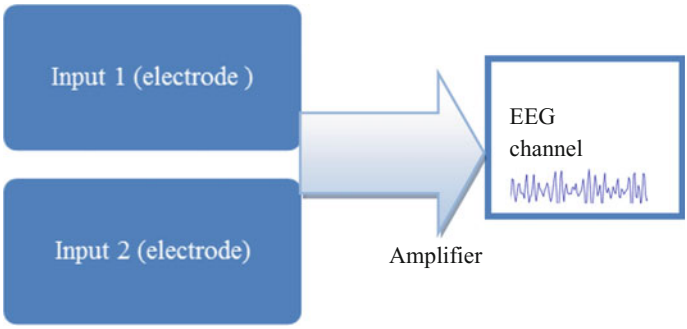


Fig. 1.3 A model of an EEG channel

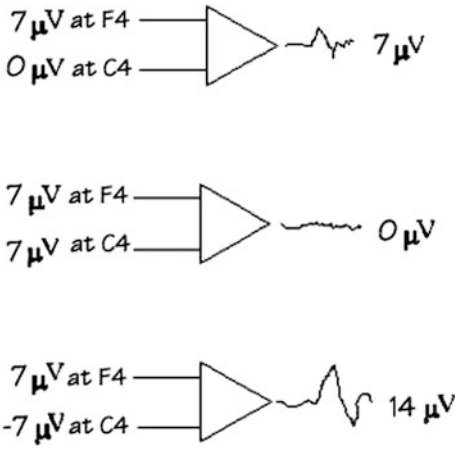


Fig. 1.4 A channel (waveform) is the difference between two adjacent electrodes (Ref. EEG Recording, medical electronics III)

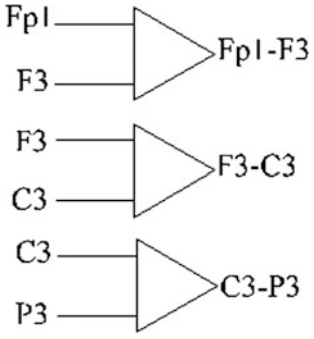


Fig. 1.5 An illustration of the bipolar montage (Ref. EEG Recording, medical electronics III)

Laplacian montage: Each channel represents the difference between an electrode and a weighted average of the surrounding electrodes (Fisch 1999). With digital EEGs, all signals are typically digitized and stored in a particular (usually

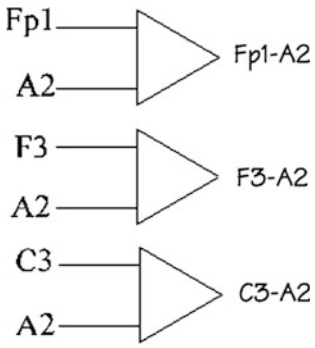


Fig. 1.6 An example of the referential montage (Ref. EEG Recording, medical electronics III)

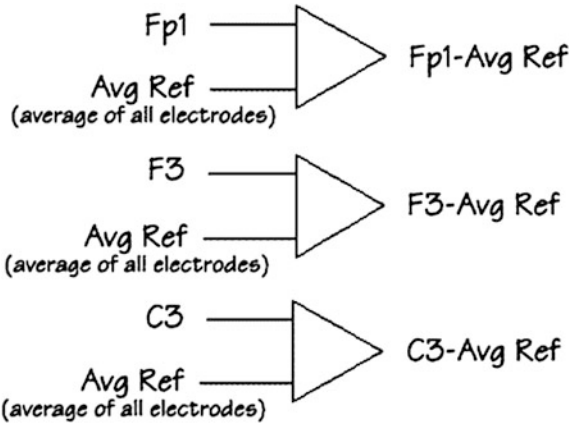


Fig. 1.7 An illustration of the average reference montage (Ref. EEG Recording, medical electronics III)

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.

1.2 Generation Organism of EEG Signals in the Brain

The human brain consists of about 100 billion nerve cells called neurons and the electrical charges of the brain are maintained by these neurons. Neurons share the same characteristics and have the same parts as other cells, but their electrochemical character 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 Fig. 1.8.

The cell nucleus is the heart of the cell giving instructions to the cell. 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

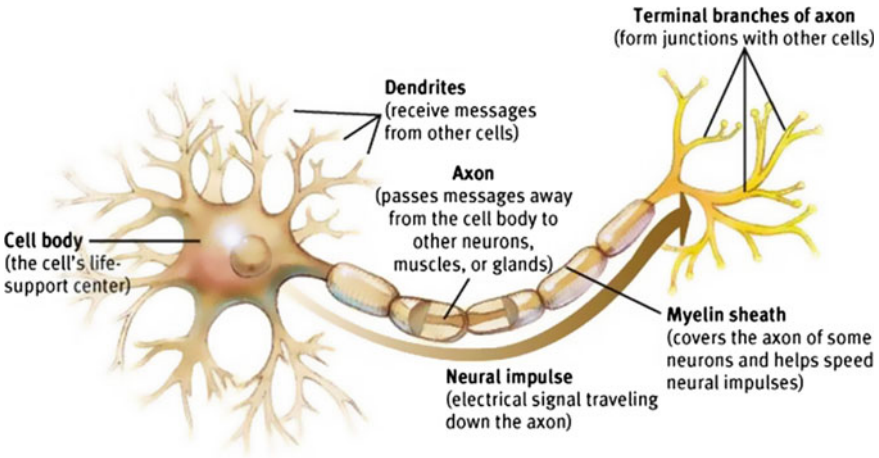


Fig. 1.8 A simple structure of a neuron (Sanei and Chambers 2007; Neuroscience, http://www.appsychology.com/Book/Biological/neuroscience.htm)

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 (Carlson 2002b; Sanei and Chambers 2007; Purves et al. 2004) called neurotransmitters. The interneuron communication system is depicted in Fig. 1.8. Figure 1.9 presents the current flow that contributes to the surface EEG during a net excitatory input. 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 PSPs from pyramidal neuron cells that create electrical dipoles between the soma and apical dendrites (see Fig. 1.9). These PSPs summate in the cortex and extend to the scalp surface where they are recorded as the EEG. Nerve cell APs have a much smaller potential field distribution and are much shorter in duration than PSPs. APs 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;

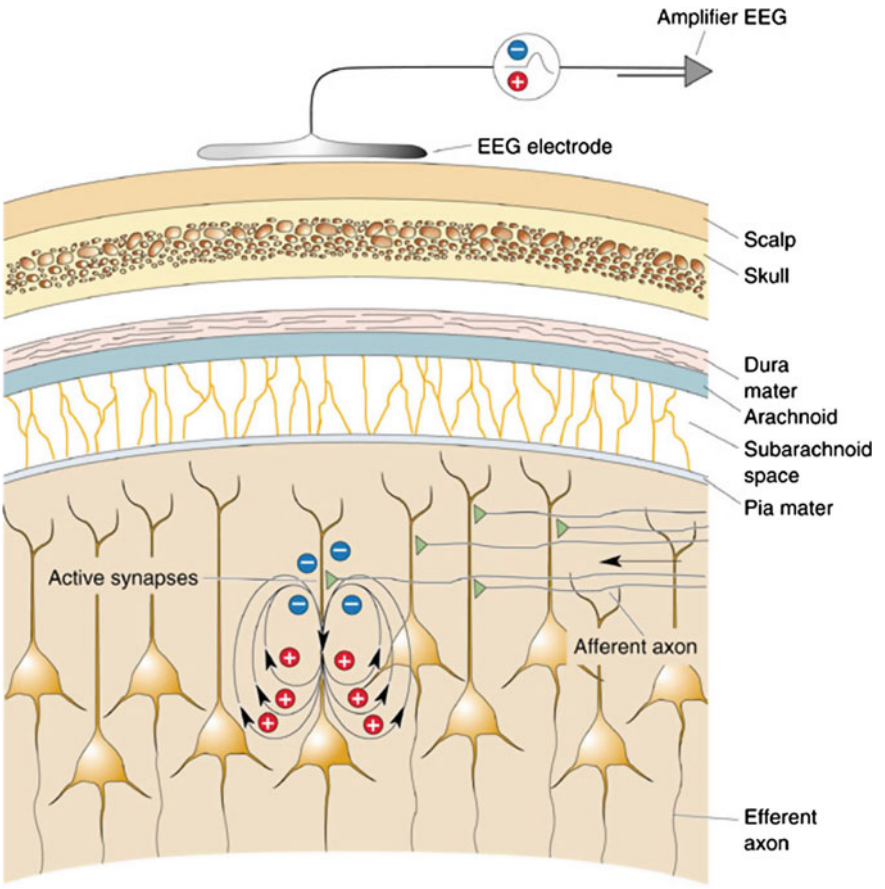


Fig. 1.9 Illustration of generation of very small electrical fields by synaptic currents in pyramidal cells. The EEG electrode measures the signal through the thick layers of tissues. Only if thousands of cells simultaneously their small voltages can the signals become large enough to be seen at the surface of the scalp (Freeman 2004a, b)

Sanei and Chambers 2007; Purves et al. 2004). The voltage, when generated by a single cell, is typically too small to be accurately measured with present-day technology.

Anatomically the brain can be divided into three major parts; *cerebrum*, *cerebellum* and *brainstem* (Gray 2002) as illustrated in Fig. 1.10. 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: the right and left hemispheres. Each hemisphere is divided into four lobes: *frontal*, *parietal*, *occipital* and *temporal* (Purves et al. 2004). These lobes are responsible for a variety of bodily functions.

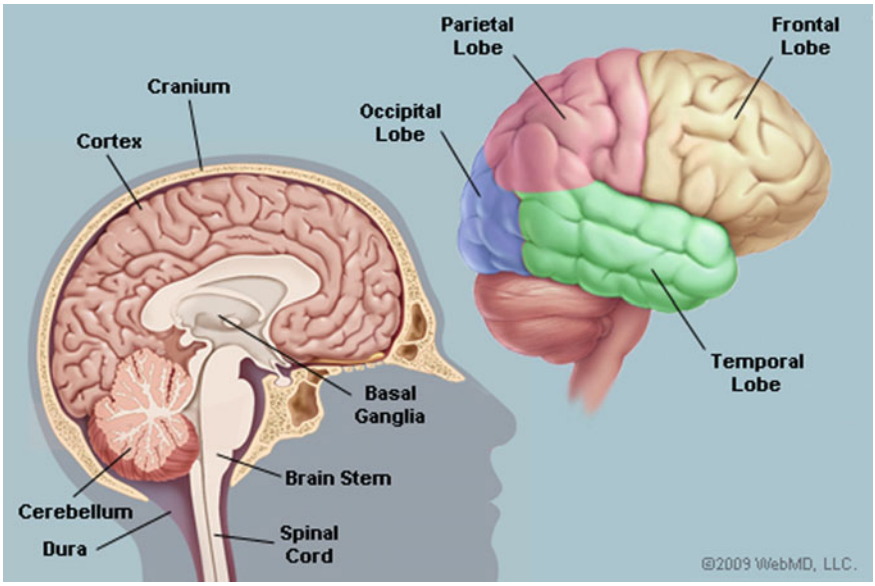


Fig. 1.10 Anatomical areas of the brain (Ref. Brain & Nervous System Health Center, WebMD)

The *frontal lobe* is involved with personality, emotions, problem solving, motor development, reasoning, planning, parts of speech and movement. The *parietal lobe* is responsible for sensation (e.g., pain, touch), sensory comprehension, recognition, perception of stimuli, orientation and movement. The *occipital lobe* is responsible for visual processing. The *temporal lobe* is involved in dealing with the recognition of auditory stimuli, speech, perception and memory. 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's 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. 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. 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.

Thus, the study of EEGs paves the way for diagnosis of many neurological disorders and other abnormalities in the human body. The EEG signals acquired

from a human (and also from animals) can be used for investigation of the following clinical problems (Birbaumer et al. 2006; Schröder et al. 2005):

- Distinguish epileptic seizures
- Characterizing seizures for the purposes of treatment
- Investigating epilepsy and locate seizure origin;
- Testing epilepsy drug effects;
- Assisting in experimental cortical excision of epileptic focus;
- Monitoring cognitive engagement;
- Monitoring the depth of anaesthesia, coma and brain deaths
- Monitoring for non-convulsive seizures/non-convulsive status epilepticus
- Locating areas of damage following head injury, stroke and tumour;
- Producing biofeedback situations;
- Controlling anaesthesia depth (servo anaesthesia);
- Monitoring the brain development;
- Testing drugs for convulsive effects;
- Investigating sleep disorders and physiology;
- Investigating mental disorders;
- Providing a hybrid data recording system together with other imaging modalities.

This list confirms the rich potential for EEG analysis and motivates the need for advanced signal processing techniques to aid clinicians in their EEG interpretation. 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 experiencing extreme emotions.

1.3 Characteristics and Nature of EEG Signals

Frequency is one of the most important criteria for assessing abnormalities in clinical EEGs and for understanding functional behaviours in cognitive research. Frequency refers to rhythmic repetitive activity (in Hz). The number of cycles in second is counted as frequency. With billions of oscillating communities of neurons as its source, human EEG potentials are manifested as aperiodic unpredictable oscillations with intermittent bursts of oscillations. In healthy adults, the amplitudes and frequencies of such signals change from one state to another, such as wakefulness and sleep. There are five major brain waves distinguished by their different frequency ranges. These frequency bands from low to high frequencies, respectively, 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 >30 Hz (*gamma*, *γ*)

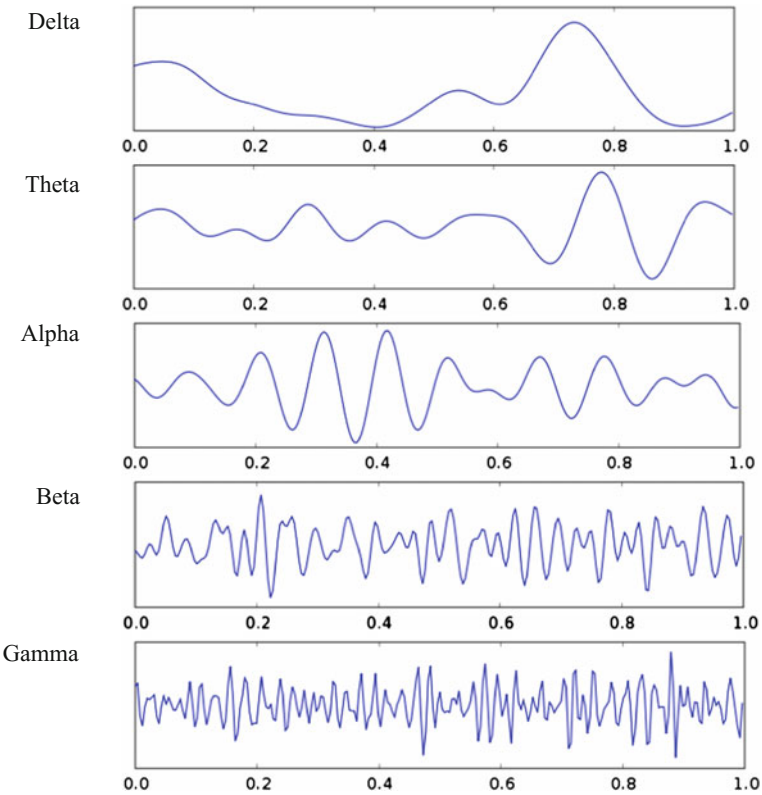


Fig. 1.11 Example of different types of normal EEG rhythms (Ref. Brainwave Entrainment, Itsu sync)

(Niedermeyer and Lopes Da Silva 2005; Fisch 1999). Higher frequencies are often more common in abnormal brain states such as epilepsy. Figure 1.11 illustrates examples of these EEG rhythms.

The *delta* wave lies between the range of 0.5–4 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 disorders and the waking state. The *theta* wave lies between 4 and 8 Hz with an amplitude usually greater than 20 μV. The *theta* arises from emotional stress, especially frustration or disappointment and unconscious material, creative inspiration and deep meditation. The *alpha* contains the frequency range from 8 to 13 Hz, with 30–50 m 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 the intense mental activity, stress and tension. The *alpha* activity recorded from sensorimotor areas is also called *mu* activity. The *beta* is in the frequency range of 13–30 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. These waves are characteristics of a strongly engaged mind. The *beta* is the brain wave usually associated with active things, active attentions and focusing on the outside world or solving concrete problems. The *gamma* waves have the frequency from 30 Hz and up. This rhythm is sometimes defined as having a maximal frequency around 80 or 100 Hz. It is associated with various cognitive and motor functions. Figure 1.12 presents an illustration of normal EEG recording. This is an example of awake EEG showing the normal or usual amount of beta activity. As shown here, beta activity is often easier to identify during relaxed wakefulness or early drowsiness.

Electrical signals in the EEG that are originated from non-cerebral origin are called artefacts. EEG data is almost always contaminated by such artefacts. The amplitude of artefacts is largely relative to the size of amplitude of the cortical signals of interest. This is one of the reasons why it takes the considerable experience to correctly interpret EEGs clinically. Figure 1.13 displays an illustration of most common four types of artefacts in human EEG recordings. 1—Electrooculographic artefact caused by the excitation of eyeball's muscles (related to blinking, for example). Big amplitude, slow, positive wave prominent in frontal electrodes. 2—Electrode's artefact caused by bad contact (and thus bigger impedance) between P3 electrode and skin. 3—Swallowing artefact. 4—Common reference electrode's artefact caused by bad contact between reference electrode and skin. Huge wave similar in all channels.

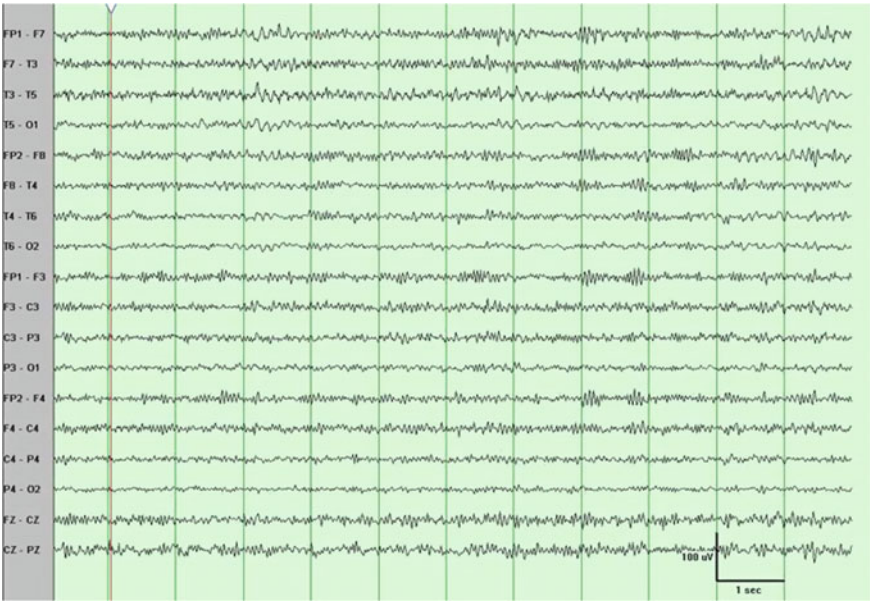


Fig. 1.12 An illustration of a normal EEG recording (Ref. Normal Awake EEG, Medscape 2015)

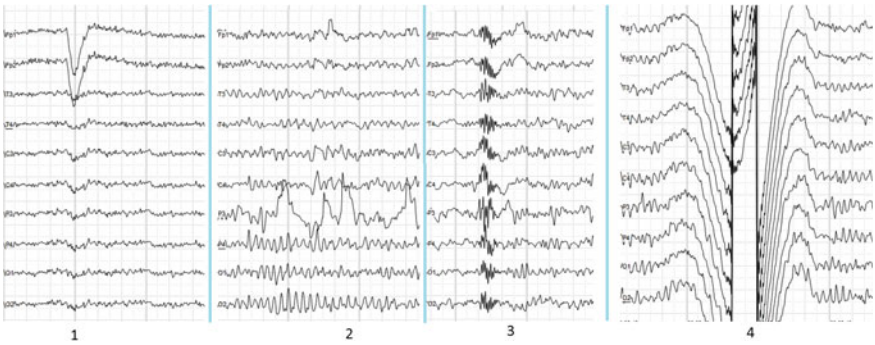


Fig. 1.13 An example of the main types of artefacts in human EEG (Picture Courtesy of Wikipedia: https://en.wikipedia.org/wiki/Electroencephalography)

1.4 Abnormal EEG Signal Patterns

EEG abnormalities typically signify dysfunction of the entire brain. Abnormal patterns of EEG signals appear as variations in the signals' patterns for certain states of the subject. Abnormal EEG findings can be divided into two categories such as, *epileptiform pattern activity* and *non-epileptiform pattern abnormalities*. *Epileptiform* refers to spike waves, sharp waves, spike and wave activity, which are the interictal marker of a patient with epilepsy and are the EEG signature of a seizure focus (Tedrus et al. 2012). *Non-epileptiform* abnormalities are characterized by alterations in normal rhythms or by the appearance of abnormal ones. They are associated with focal cerebral dysfunction, often due to a demonstrable structural lesion. By contrast, more widespread central nervous system (CNS) derangements, such as those due to metabolic disturbances, usually produce generalized EEG abnormalities. Identification of an abnormality requires the analysis of the EEG to determine various neurological conditions and any other available information. A precise characterization of abnormal patterns leads to clearer insight into some specific pathophysiologic reactions, such as epilepsy, or specific disease processes, such as subacute sclerosing panencephalitis.

The most significant category is *epileptiform pattern*, obtained in the epileptic seizure signal pattern, called ictal wave patterns, which appear during the onset of epilepsy. Epilepsy is defined as a brain disorder characterized by an enduring predisposition to generate epileptic seizures and by the neurobiologic, cognitive, psychological and social consequences of this condition. Seizures are caused by transient, paroxysmal and synchronous discharges of groups of neurons in the brain. A detailed discussion of epilepsy and epileptic seizures is provided in Chap. 2. Chatrian et al. (1974) defined the term *epileptiform* to describe distinct waves or complexes, distinguishable from the background activity, which resemble the waveforms recorded in a proportion of human subjects suffering from an epileptic disorder (Noachtar et al. 1974). Figures 1.14 and 1.15 display the

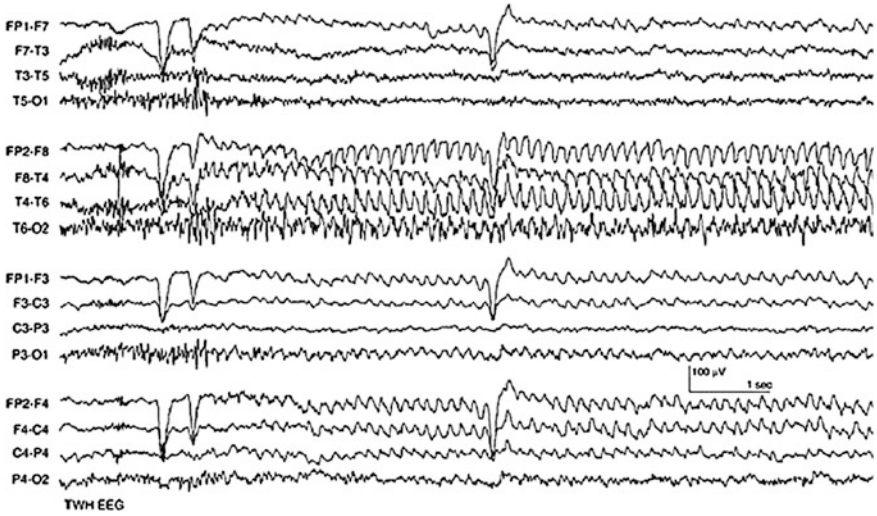


Fig. 1.14 EEG recordings of a partial seizure (Mani 2014)

Fig. 1.15 EEG recordings of a generalized epileptic seizure (Ref. Epilepsy and its Treatment for Providers, Angelman)

examples of e*pileptiform* patterns for a partial seizure and a generalized seizure, respectively. 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 a slow decrease in frequency (but increase in amplitude) during the seizure period. In the clinical practice, the ability to distinguish between epileptiform activity and an epileptic disorder may be challenging, as variability may be seen within each epilepsy syndrome and within a given child over time. The wave of the partial seizure may or may not be spiky in shape. Sudden desynchronization of electrical activity is found in electrodecremental seizures. The transition from the preictal to the ictal state, for a focal epileptic seizure, consists of a gradual change from chaotic to ordered waveforms. The amplitude of the spikes does not necessarily represent the severity of the seizure. Rolandic spikes in a child of 4–10 years, for example, are very prominent; however, the seizure disorder is usually quite benign or there may not be clinical seizure.

The demonstration of *non-epileptiform* EEG abnormalities in patients with an altered mental status or level of consciousness, for example, can be especially useful in guiding decision-making and the best treatment. In these situations, however, the EEG provides evidence of organic electrophysiological dysfunction and the patterns observed may orientate for the diagnostic possibilities. There are several types of *non-epileptiform* EEG abnormalities such as *focal slow activity*, *generalized or regional bisynchronous slow activity*, *generalized asynchronous slow activity*, *generalized attenuation/suppression*, *focal attenuation* and *other abnormal activity*.

Focal slow activity may be an indicative sign of focal cerebral dysfunction, especially in awake adults, and it seems to be the result of a cortex deafferentation from subcortical structures (Andraus and Alves-Leon 2011; Schaul 1998). It is the most common phenomenon encountered in clinical EEG and is indicative of a localized structural lesion (Schaul 1998). Slow activity is classified according to frequency in theta activity (ranging from 4.0 to 7.9 Hertz (Hz) or cycles per second), and in delta activity (around 0.5–3.9 Hz). Continuous slow activity suggests a more severe brain damage (likelihood of increased mass effect, large lesion or deep hemispheric lesion), whereas intermittent slow activity usually indicates a small lesion and the absence of a mass effect (Schaul et al. 1981). In general, fast-growing tumours, such as glioblastoma multiform or metastatic brain tumours, are associated with focal slow activity occurring in delta frequency as shown in Fig. 1.16. *Generalized or regional bisynchronous slow activity* may be intermittent or continuous, and seems to be due to disordered circuits between the cortex and thalamus, although there has been some controversy about its genesis and significance (Schaul 1998; Andraus et al. 2011). This type of abnormality can be found in conditions that affect both cortical and subcortical structures, as well as the presence of several toxic-metabolic encephalopathies, early stages of coma and deep midline lesions encephalopathy and is always abnormal in awake adults (Schaul 1998; Abou-Khalil and Missulis 2006). Some possibilities include degenerative processes such as, encephalitis, extensive multifocal vascular (Schaul 1998; Abou-Khalil and Missulis 2006). The abnormalities are associated with a large number of disorders,

Fig. 1.16 Polymorphic delta activity in the left hemisphere in a patient with a brain tumour (Lee and Khoshbin 2015)

including toxic-metabolic encephalopathies, early stages of coma (see Fig. 1.17), degenerative diseases, such as dementia (see in Fig. 1.18) and other conditions, affecting both cortical and subcortical structures (Abou-Khalil and Missulis 2006). *Generalized asynchronous slow activity* consisting of frequencies less than 4.0 Hz, is highly nonspecific and has a broad differential diagnosis (Schaul 1998). Its presence usually indicates diseases and toxic-metabolic encephalopathies. It is worth remembering that the correct interpretation of generalized asynchronous slow activity takes into account the age and condition of the patient, given that widespread slow activity may be normally present in drowsiness and sleep in all ages and in awake children, depending on the age (Abou-Khalil and Missulis 2006). It should be considered an abnormal activity when the pattern is inconsistent with age and stages of sleep. *Focal attenuation* indicates reduced amplitude of one type of activity that occurs at certain frequency, or of the entire EEG activity. Attenuation generally indicates focal cortical lesion or reversible cortical dysfunction (post-ictal state, for example), but may be related to the presence of a collection between the cortex and recording electrode (like an hematoma or subdural empyema) or a tumour (a dural based tumour, such as a meningioma, for example), leading to an increased distance between the cortex and the recording electrode (Abou-Khalil and Missulis 2006). *Generalized attenuation* may suggest cortical generalized injury or transitory dysfunction (Abou-Khalil and Missulis 2006). It is most often found in

Fig. 1.17 EEG patterns of *alpha* coma in an awake patient (Encephalopathic EEG Patterns, Medscape 2015)

Fig. 1.18 EEG findings in dementia (EEG in Dementia and Encephalopathy, Medscape 2015)

Fig. 1.19 EEG findings in autism (Moving Autism Forward 2015)

patients who have suffered a severe cerebral damage in postanoxic encephalopathy, or are under the effect of anaesthetic drugs or in a drug-induced coma. The term burst suppression refers to the presence of brain activity bursts of variable amplitude, duration and form, followed by a marked depression of the activity, which occur on a cyclical basis. It is most often found in patients who suffered severe cerebral damage in postanoxic encephalopathy, under the effect of anaesthetic drugs or drug-induced coma. In conclusion, the *non-epileptiform* EEG abnormalities provide evidence of brain dysfunction, which may be focal or generalized. Figure 1.19 shows abnormality in the EEG recoding of autism children. Many pathological processes can lead to their appearance, which, when properly analyzed, could help the diagnosis.

References

10/20 positioning/DIY tDCS; http://www.diytdcs.com/tag/1020-positioning/.

Abou-Khalil B, Missulis KE. Abnormal EEG: non-epileptiform abnormalities. In: Abou-Khalil B, Misulis KE (Eds). Atlas of EEG and seizure semiology. Philadelphia: Elsevier, 2006:99–123.

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.

Andraus M. E. C., Alves-Leon S. V., Non-epileptiform EEG abnormalities, *Arq Neuropsiquiatr* 2011; 69(5):829–835.

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

Birbaumer, N., Weber, C., Neuper, C., Buch, E., Haagen, K., and Cohen, K., 'Brain–computer interface research: coming of age', *Clin. Neurophysiol.*, 117, 2006, 479–483.

Brain & Nervous System Health Center, WebMD, http://www.webmd.com/brain/picture-of-the-brain.

Brainwave Entrainment, Itsu sync, http://itsusync.com/different-types-of-brain-waves-delta-theta-alpha-beta-gamma.

Carlson, N. R. (2002a) *Foundations of physiological psychology*, 5th ed., Boston, Mass. London: Allyn and Bacon.

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.

Campisi, P. "EEG for Automatic Person Recognition", *Computer*, vol. 45, no. 7, pp. 87–89, July 2012, doi:10.1109/MC.2012.233.

Chatrian et al. (1974) A glossary of terms most commonly used by clinical electroencephalographers. *Electroencephalogr Clin Neurophysiol*. 1974 Nov. 37(5):538–48.

Collura, T. F. (1993) 'History and evolution of electroencephalographic instruments and techniques', Clinical Neurophysiology, Vol. 10, Iss. 4, pp. 476–504.

EEG in Dementia and Encephalopathy, Medscape, 2015; http://emedicine.medscape.com/article/1138235-overview#showall.

EEG Recording, medical electronics III, http://medical-electronics-iii.blogspot.com.au/2007_10_01_archive.html.

EEG, Saint Luke's Health System; http://www.saintlukeshealthsystem.org/health-library/electroencephalogram-eeg.

Encephalopathic EEG Patterns, Medscape, 2015. http://emedicine.medscape.com/article/1140530-overview#a2 Alejandro L. Escalaya and Jorge G. Burneo, Surgical Treatment of Neurocysticercosis-Related Epilepsy, Epilepsy Program, Western University, London, Ontario, Canada, DOI:10.5772/54275.

Epilepsy and its Treatment for Providers, Angelman, http://www.angelman.org/what-is-as/medical-information/epilepsy-and-its-treatment-for-providers/.

Fisch B J (1999) *EEG premier*: *Basic principles of digital and analog EEG* (third edition), Elsevier publication.

Freeman WJ. (2004a) Origin, structure, and role of background EEG activity. Part 1. Analytic amplitude. Clinical Neurophysiology. 115(9):2077–88.

Freeman WJ. (2004b) Origin, structure, and role of background EEG activity. Part 2. Analytic phase. Clinical Neurophysiology. 115(9):2089–107.

Gray, F. J. (2002) *Anatomy for the medical clinician*, first edition, Shannon Books Pty Ltd., Victoria, Australia.

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.

Jasper, H. H. (1958) 'The ten-twenty electrode system of the International Federation', *Electroencephalogram. Clinical. Neurophysiology*. Vol. 10, pp: 367–380.

Lee, J.E., and Khoshbin, S. (2015) Clinical Neurophysiology and Electroencephalography, http://clinicalgate.com/75-clinical-neurophysiology-and-electroencephalography/.

Mani, J. (2014), Video electroencephalogram telemetry in temporal lobe epilepsy, Annals of Indian academy of Neurology, Volume: 17(5) pp. 45–49.

Moving Autism Forward, 2015; https://tacanowblog.com/2015/04/17/15-years-later-ready-for-autism-answers/.

Neuroscience, http://www.appsychology.com/Book/Biological/neuroscience.htm.

Niedermeyer E. and Lopes da Silva F. (2005) *Electroencephalography*: *basic principles*, *clinical applications*, *and related fields*, Lippincott Williams & Wilkins, ISBN 0781751268, 5th edition, 2005.

Noachtar S., Binnie C., Ebersole J., MauguieÁre F., Sakamoto A. and Westmoreland B., A glossary of terms most commonly used by clinical electroencephalographers, *Electroencephalogr Clin Neurophysiol*, 1974 Nov. 37(5):538–48.

Normal Awake EEG, Medscape, 2015; http://emedicine.medscape.com/article/1140143-overview.

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.

Sanei, S. and Chambers, J. A. (2007) *EEG Signal Processing*, John Wiley & Sons, Ltd., 2007.

Schaul N. The fundamental neural mechanisms of electroencephalography. Electroencephalogr Clin Neurophysiol 1998;106:101–107.

Schaul N, Gloor P, Gotman J. The EEG in deep midline lesions. Neurology 1981; 31:157–167.

Schröder, M. I., Lal, T. N., Hinterberger, T., Bogdan, M., Hill, N. J., Birbaumer, N., Rosenstiel, W., and Schölkopf, B., 'Robust EEG channel selection across subjects for brain–computer interfaces', *EURASIP J. Appl. Signal Proces.*, **19**, 2005, 3103–3112.

Tedrus GM, Fonseca LC, Nogueira Junior E, Pazetto D. Epilepsy with onset at over 50 years of age: clinical and electroencephalographic characteristics. *Arq Neuropsiquiatr*. 2012 Oct. 70 (10):780–5.

Chapter 2
Significance of EEG Signals in Medical and Health Research

EEG is becoming increasingly important in the diagnosis and treatment of mental and brain neuro-degenerative diseases and abnormalities. The role of the EEG is to help physicians for establishing an accurate diagnosis. In neurology, a main diagnostic application of EEGs is in the case of epilepsy, as epileptic activity can create clear abnormalities on a standard EEG study (Abou-Khalil and Misulis 2006). EEG is also used as a first-line method for diagnosing many neurological disorders, such as dementias, Alzheimer's disease, brain tumours, strokes, Parkinson's disease, migraine, neuroinfections, sleep disorders and traumatic disorders of the nervous system, such as brain trauma and autism. Furthermore, EEG can also be used in the diagnosis of coma, encephalopathies and brain deaths. EEG has become very popular in brain–computer interface (BCI) applications. EEG and its derivatives, event-related potentials (ERPs), are used extensively in neuroscience, cognitive science, cognitive psychology, and psycho physiological research. ERPs refer to averaged EEG responses that are time-locked to more complex processing of stimuli.

Many techniques used in research are not sufficiently standardized to be used in the clinical context. As EEG recordings contain a huge amount of data, the development of computer-aided analysis systems is essential for classifying abnormal EEG signals from normal EEGs to support the diagnosis of brain diseases and to contribute to a better understanding of mental states for BCI applications. 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.

In this chapter, we provide a brief discussion of various uses and the significance of EEGs in brain disorder diagnosis and also in BCI systems. In this chapter, we also discuss why EEG signal analysis and classification are required for medical and health practice and research. Then, we provide the key concepts of EEG signal classification and a brief description of computer-aided diagnostic (CAD) systems.

S. Siuly et al., *EEG Signal Analysis and Classification*,
Health Information Science, DOI 10.1007/978-3-319-47653-7_2

2.1 EEG in Epilepsy Diagnosis

Epilepsy is one of the most common and devastating neurological diseases experienced worldwide (Supriya et al. 2016). The hallmark of epilepsy is recurrent seizures termed "epileptic seizures" (Alexandros et al. 2012). Seizures are defined as sudden changes in the electrical functioning of the brain, resulting in altered behaviours such as losing consciousness, jerky movements, temporary 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.

Figure 2.1 shows an image of abnormal electrical impulses during a seizure. 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. 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 (1–100 μV). In some people, this happens only occasionally; for others, it may happen up to hundreds of times a day.

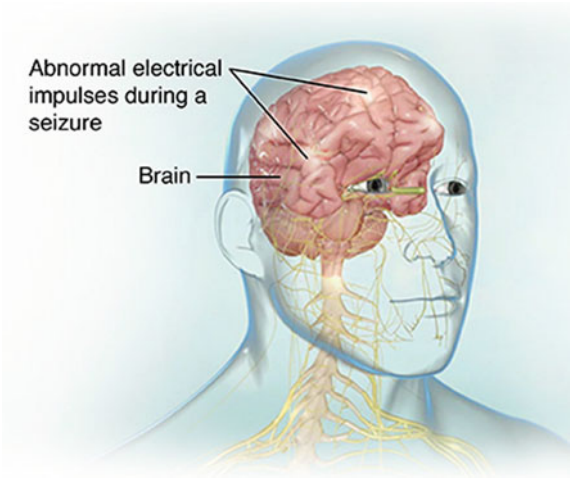


Fig. 2.1 An illustration of abnormal electrical impulses during a seizure (Ref. Johns Hopkins Medicine Health Library)

EEG is an essential component for the diagnosis and analysis of epilepsy (Alotaiby et al. 2014; Kabir et al. 2016). EEG continues to play a central research role in the diagnosis and management of patients with seizure disorders. One of the main reasons behind this is that it is a convenient and relatively inexpensive way to demonstrate the physiological manifestations of abnormal cortical excitability that underlies epilepsy. EEG's chief manifestation is the epileptic seizure. The seizures can encompass a discrete part of the brain partial or the complete cerebral mass generalized. The recorded EEG represents electrical activity produced by firing of neuron within the brain along the scalp. When epilepsy is present, seizure activity will appear as rapid spiking waves on the EEG. Epileptic activity can create clear abnormalities on a standard EEG. Epilepsy leaves its signature in the EEG signals (Siuly et al. 2015). 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 *interictal* (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 a slow reduction in frequency but increases in amplitude during the seizure period. It may or may not be spiky in shape. To assist the diagnosis and treatment of epilepsy or neurological disease, this book aims to develop methods that can identify the epileptic EEG signals during seizure activity and also during seizure-free time.

EEG helps to determine seizure types and epilepsy syndrome in patients with epilepsy, thereby helping to determine antiepileptic medication choice and the prediction of prognosis. Another important contribution of EEG findings is to determine the multi-axial diagnosis of epilepsy, in terms of whether the seizure disorder is focal or generalized, idiopathic or symptomatic, or part of a specific epilepsy syndrome (Smith 2005). Figure 2.2 displays usual patterns of EEG signals for normal, partial seizure and generalized seizure. The EEG provides important information about background EEG and epileptiform discharges and is required for the diagnosis of specific electroclinical syndromes (Nolan et al. 2004). Such diagnoses carry important prognostic information, guide selection of antiepileptic medication, and suggest when to discontinue medication. Neurologic examination and imaging in the essential idiopathic, typically genetic, epilepsies are normal (Urbach 2005). Following a seizure, the EEG background may be slow. However, interictal background EEG frequencies that are slower than normal for age usually suggest symptomatic epilepsy. Thus, EEG background offers important prognostic and classification information. Epileptiform discharges help clinicians to separate generalized from focal (i.e. partial) seizures.

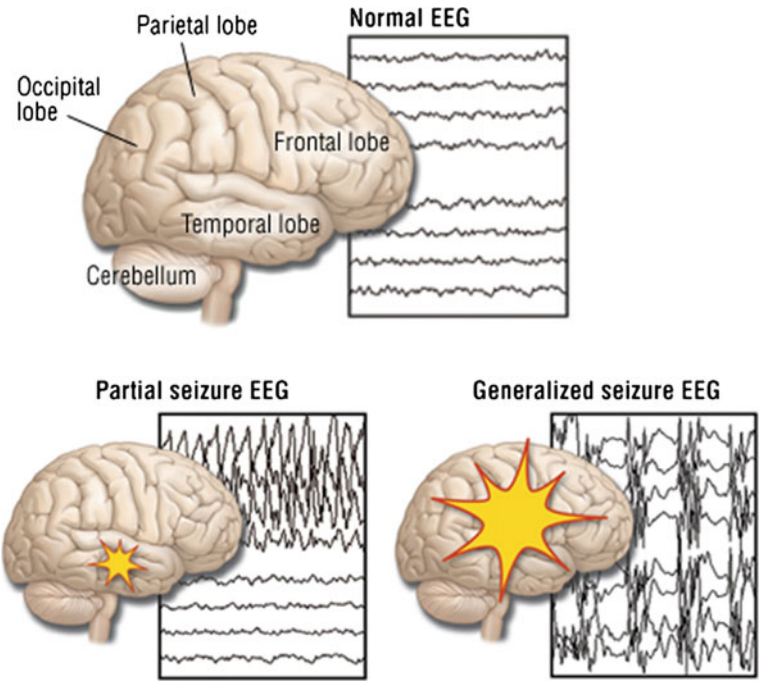


Fig. 2.2 Typical patterns of EEG signals for normal, partial seizure and generalized seizure (Ref. Seizure, Drugs.com)

2.2 EEG in Dementia Diagnosis

Dementia refers to a group of neuro-degenerative disorder diversity caused by the gradual neuronal dysfunction and death of brain cells. This disorder can be defined clinically as a syndrome that causes a decline in cognitive domain (i.e. attention, memory, executive function, visual-spatial ability and language) (Al-Qazzaz et al. 2014), which is common in the elderly. Dementia is classified into Alzheimer's disease (AD), Parkinson's disease (PD), dementia with lewy bodies, Creutzfeldt–Jakob disease, normal pressure hydrocephalus, vascular dementia and frontotemporal dementia (Minguez and Winblad 2010; DeKosky and Marek 2003). AD is the most well-known and common type of dementia. Out of all the mentioned types of dementia, two-third of the demented patients suffer from AD.

EEG plays an important role in detecting and classifying dementia. The resting alpha frequency declines, especially in early dementia. It is known that the lower limit of normal alpha frequency is 8 Hz (cycles per second). In assessing the frequency of the alpha rhythm, alerting manoeuvres are essential to ensure that the patient is in the best awake state and not drowsy. Although the EEG may be normal or minimally disturbed in a number of patients in the initial stages of AD, an abnormal EEG usually is recorded later in the course. A large percentage of patients

with moderately severe-to-severe AD exhibit abnormal EEGs. Computerized methods, such as EEG spectral analysis (Neto et al. 2015), coherence and complexity (i.e. correlation dimension), have been demonstrated to correspond to cognitive function (Staudinger and Polikar 2011). EEG is useful for clinical evaluation because of its ease of use, non-invasiveness and capability to differentiate types and severity of dementia at a cost lower than that of other neuro-imaging techniques.

2.3 EEG in Brain Tumour Diagnosis

A brain tumour is a collection, or mass, of abnormal cells in the brain. Different types of brain tumours are featured in the literature, where some are noncancerous (benign), and the others are cancerous (malignant). Sometimes, brain tumours can begin in the brain (primary brain tumours), alternatively cancer can begin in other parts of the body and spread to the brain (secondary, or metastatic, brain tumours). EEG is primarily used to complement advanced imaging, for example, CT scanning, and MRI. EEG can also be used for localization diagnosis of brain tumours. Old neuro-imaging techniques consisted of skull radiography (Ref. Medscape (EEG in Brain Tumours)). EEG was first introduced by Walter in 1936 by introducing the term "delta waves" that identified the association between localized slow waves on the EEG and tumours of the cerebral hemispheres (Walter 1936). Delta is the frequency of EEG that is less than 4 Hertz (Hz), whereas the normal alpha is between 8 and 12 Hz. This established EEG is known to be an important tool for localizing brain tumours. After this establishment, in the next four decades, electroencephalographers mounted an enormous effort to improve the accuracy of localization and seek clues to underlying pathological processes. EEG abnormalities in brain tumours depend on the stage at which the patient presents for evaluation and the changes observed with tumours result mainly from disturbances in bordering brain parenchyma as tumour tissue is electrically silent (Ref. EEG in Brain Tumours, Medscape).

It is known that brain tumours may be associated with various EEG findings. Fischer-Williams and Dike (Fischer-Williams and Dike 1993) mentioned that the following characteristics may be seen at the time of brain tumour diagnosis:

- Focal slow activity
- Focal attenuation of background activity
- Asymmetric beta activity
- Disturbance of the alpha rhythm
- Interictal epileptiform discharges (spikes and sharp waves)
- Normal EEG.

2.4 EEG in Stroke Diagnosis

A stroke is the sudden death of brain cells in a localized area which happens when the blood flow to an area of the brain is interrupted by either a blood clot or a broken blood vessel. A stroke is a medical emergency that kills many brain cells per minute and causes permanent brain damage. Depending on the region of the brain affected, a stroke may cause paralysis, speech impairment, loss of memory and reasoning ability, coma or death. A stroke is also sometimes called a brain attack or a cerebrovascular accident (CVA). There are two main types of stroke: *ischemic*, due to lack of blood flow, and *hemorrhagic*, due to bleeding. About four out of every five strokes are *ischaemic*. About one in every five strokes is *haemorrhagic*. After dementia, strokes are the second leading cause of disability. Disability may include loss of vision and/or speech, paralysis and confusion. Once the damage becomes clinically or radiographically apparent, proper neurological exams and imaging are very useful for detecting delayed cerebral ischemia stroke. Thus, EEG can be a useful way to detect and subsequently treat ischemia before the injury becomes irreversible. EEG is also very useful for identifying the stroke. In the operating room, EEG has an established role in identifying ischemia prior to the development of infarction during carotid endarterectomy (Foreman and Claassen 2012).

Brain function is represented on an EEG by oscillations of certain frequencies. Slower frequencies (typically delta [0.5–3 Hz] or theta [4–7 Hz]) are generated by the thalamus and by cells in layers II–VI of the cortex. Faster frequencies (or alpha, typically 8–12 Hz) derive from cells in layers IV and V of the cortex (Amzica et al. 2010). All frequencies are modulated by the reticular activating system, which corresponds to the observation of reactivity on the EEG (Evans 1976). Pyramidal neurons found in layers III, V and VI are exquisitely sensitive to conditions of low oxygen, such as ischemia, thus leading to many of the abnormal changes in the patterns seen on an EEG (Ordan 2004).

2.5 EEG in Autism Diagnosis

Autism and related autism spectrum disorders (ASD) are lifelong, often severely impairing neurodevelopmental syndromes involving deficits in social relatedness, language and behaviour. The diagnosis of autism is a difficult process that usually includes certain behavioural and cognitive characteristics. As EEG recording and analysis is one of the fundamental tools in diagnosing and identifying disorders in neurophysiology, researchers are trying to identify diagnostic approaches for autism, which is a disorder of neurophysiology (Hashemian and Pourghassem 2014). The EEG signal is characterized by a high temporal resolution (in the order of milliseconds) allowing for precise temporal examination of cortical activity. In

autism analysis, EEG rhythms are the most commonly used features based on the comparison technique.

Dr. Molholm and colleagues suggested that brainwave EEG recordings could potentially reveal how severely autistic individuals are affected. In addition, EEG recordings might help diagnose autism earlier. “Early diagnosis allows for earlier treatment-which we know increases the likelihood of a better outcome”, said Molholm (Ref. Autism Research 2014). Molholm hopes that one day, EEG recordings might contribute to a more objective diagnosis for autism, as well as better categorization of where people are on the spectrum. The increased prevalence of epilepsy and/or epileptiform EEG abnormalities in individuals with ASD may be an important clue to an underlying neurological abnormality, at least for a subset of autism patients. In general, autistic children appear to show poor short range connectivity in the left hemisphere of the brain which is responsible for language, while at the same time having an increased connectivity in regions that are further apart from each other, indicating that some compensation or misbalance is occurring. The EEG tests have proved more accurate, providing a clearer picture and able to identify these patterns.

2.6 EEG in Sleep Disorder Diagnosis

The term “sleep disorder” refers to a range of conditions that result in abnormalities during sleep. The most common disorder is sleep apnoea. Sleep apnoea occurs when the walls of the throat come together during sleep, blocking off the upper airway. The full name for this condition is Obstructive Sleep Apnoea (OSA). Another rare form of breathing disturbance during sleep is called central sleep apnoea. It is caused by a disruption to the mechanisms that control the rate and depth of breathing.

Sleep staging is crucial for the diagnosis and treatment of sleep disorders. It also relates closely to the study of brain function. The EEG monitors the various stages of sleep and is interpreted by clinicians. This information is very useful to get a clear picture of different types and causes of sleep disorders. For example, in an intensive care unit, EEG wave classification is used to continuously monitor patients’ brain activities. For newborn infants at risk of developmental disabilities, sleep staging is used to assess brain maturation. Many other applications acclimate the EEG wave classification techniques (originally developed for sleep staging) to their purposes. Besides being used to study human activities, sleep staging has also been used to study avian bird song systems and evolutionary theories about mammalian sleep.

2.7 EEG in Alcoholism Diagnosis

Alcoholism is a severe disorder that affects the functionality of neurons in the central nervous system (CNS) and alters the behaviour of an affected person. Alcoholism causes a wide range of effects such as liver diseases, heart diseases, brain damage and certain cancers. It is also the cause of other harms such as road and other accidents, domestic and public violence and crime, and it contributes to family breakdown and broader social dysfunction (MCDS 2011). Alcoholics experience numerous cognitive deficiencies, for instance, learning and memory deficits, impairment of decision making and problems with motor skills, as well as suffering behavioural changes that include anxiety and depression (Harper 2007; Brust 2010; Bajaj et al. 2016). EEG signals can be used as a diagnostic tool in the evaluation of subjects with alcoholism. The neurophysiological interpretation of EEG signals in persons with alcoholism is based on observations and interpretations of the frequency and power in their EEGs compared to EEG signals from persons without alcoholism (Acharya et al. 2014). While undertaking cognitive tasks, alcoholics and people at risk of alcoholism manifest increased resting oscillations (e.g. theta, beta) and decreased "active" oscillations in the same frequency bands (Spence and Schneider 2009). Not only does this underlying central nervous system (CNS) disinhibition appear to be involved in the predisposition toward alcoholism, it is hypothesized that neuroelectric features related to CNS disinhibition may provide insights into the neurobiology of craving and relapse. The relationship between this underlying CNS hyperexcitability and the induction of alcohol abuses leading to alcohol dependence remains to be explained.

2.8 EEG in Anaesthesia Monitoring

Awareness during surgery is one of the most feared complications of anaesthesia. EEG is used to monitor anaesthetic depth. Monitoring the level of consciousness during general anaesthesia with processed EEG monitors has become an almost routine practice in the operating room, despite ambiguous research results regarding its potential benefits (Musialowicz and Lahtinen 2014). Using EEG signals to monitor the depth of anaesthesia reduces the incidences of intraoperative awareness, leads to a reduction in drug consumption, prevents anaesthesia-related adverse events and enables faster recovery. EEG patterns are known to change with the patient's depth of anaesthesia, and assessment of hypnosis requires measurements of electrical activity in the central nervous system (CNS). Anaesthetics act on the brain; thus, this organ should be monitored in addition to the patient's spinal cord reflexes and cardiovascular system signs, such as blood pressure and heart rate. EEG-based depth-of-anaesthesia (DoA) monitors use algorithms to continuously analyse EEG signals and translate any changes into simple numerical indices that correspond to the level of consciousness.

There are many different EEG-based DoA monitors, including the Bispectral Index System Monitor (BIS Monitor; Covidien, USA); E-Entropy Module (GE Healthcare, USA); Narcotrend Compact M (MonitorTechnik, Germany); AEP A-line Monitor, Cerebral State Monitor (Danmeter, Denmark); Patient State Index Monitor, SEDLine (Masimo, USA); and SNAPII (Stryker, USA). However, not all of them have been thoroughly validated by clinical studies. Among them, the BIS is the best described monitor of the depth of the hypnotic component of anaesthesia or sedation. The BIS index discriminates between awake and asleep states but with considerable overlap of values and no clear-cut transition between awake and asleep values at the end of surgery.

2.9 EEG in Coma and Brain Death

Coma is an eyes-closed state of unresponsiveness with severely impaired arousal and cognition. It represents a failure of neurologic function resulting from damage to a critical number of brainstem and diencephalic pathways, which regulate the overall level of cortical function (Sutter and Kaplan 2012). Coma has been identified as a major predictor of death and poor neurofunctional outcomes in patients with a variety of critical illnesses including ischemic strokes (Sacco et al. 1990), intracerebral haemorrhage (Tuhrim et al. 1988), traumatic brain injury (Teasdale and Jennett 1976; Perel et al. 2008), hypoxic encephalopathy after cardiac arrest (Sacco et al. 1990; Levy et all. 1985; Booth et al. 2004) and metabolic derangements or sepsis (Sacco et al. 1990). EEG has long been used in evaluating comatose patients, and is being increasingly found to uncover patterns of prognostic significance, reveal subclinical seizure activity and provide data during treatment in which patients are paralysed. Some EEG patterns reveal increasing degrees of cerebral compromise with a progressive slowing of the background frequencies, while others can be explored for reactivity to external stimuli for prognostic purposes. When a patient's brain falls completely silent and electrical recordings devices show a flat line reflecting a lack of brain activity, doctors consider the patient to have reached the deepest stage of a coma.

Brain death is referred to the complete, irreversible and permanent loss of all brain and brainstem functions (Chen et al. 2008). Brain death implies the termination of a human's life. EEG is often used in the confirmatory test for brain death diagnosis in clinical practice. Because EEG recording and monitoring is relatively safe for the patients in deep coma, it is believed to be valuable for either reducing the risk of brain death diagnosis (while comparing other tests such as the apnea) or preventing mistaken diagnosis. Generally, an EEG demonstrates electrocerebral silence reflecting the absence of electrical brain activity. Transcranial doppler studies reveal the absence of cerebral blood flow. EEG is used to diagnose brain death in order to terminate treatment or prepare for organ donation.

2.10 EEG in Brain–Computer Interfaces (BCIs)

A relatively new but emergent field for EEGs is brain–computer interfaces (BCIs). BCI is a technology that finds a new communicative way between a brain and a computer. A BCI is a computer-based system that acquires brain signals, analyses them, and translates them into commands that are relayed to an output device to carry out a desired action (see Fig. 2.6). The key aim of BCI research is to create a new communication pathway allowing direct transmission of messages from the brain by analysing the brain's mental activities for people suffering from severe neuromuscular disabilities (Siuly and Li 2015). To measure EEG signal, an electrode cap is placed on the head of a user. To command the machine a user imagines a specific task, such as the movement of limbs or 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 a cursor movement) or control a machine (e.g. wheelchair).

BCIs do not require actual physical movement, and hence they may be the only means of communications possible for people who have severe motor disabilities. BCIs could also help reduce symptoms resulting from stroke, autism, emotional and attention disorders. 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.

In general, BCI systems allow individuals to interact with the external environment by consciously controlling their thoughts instead of contracting muscles (e.g. human–machine interfaces controlled or managed by myoelectric signals). A BCI system typically 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) as shown in Fig. 2.3. The result of the classification allows external devices to control signals. Another aspect of BCI systems is that the user receives stimuli (visual, auditory or tactile) and/or performs mental tasks while the brain signals are captured and processed. Based on the stimulus or task performed by the user, several phenomena or behaviours extracted from the EEG signals can be detected.

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 book, we use EEGs as the measurement of brain activities.

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

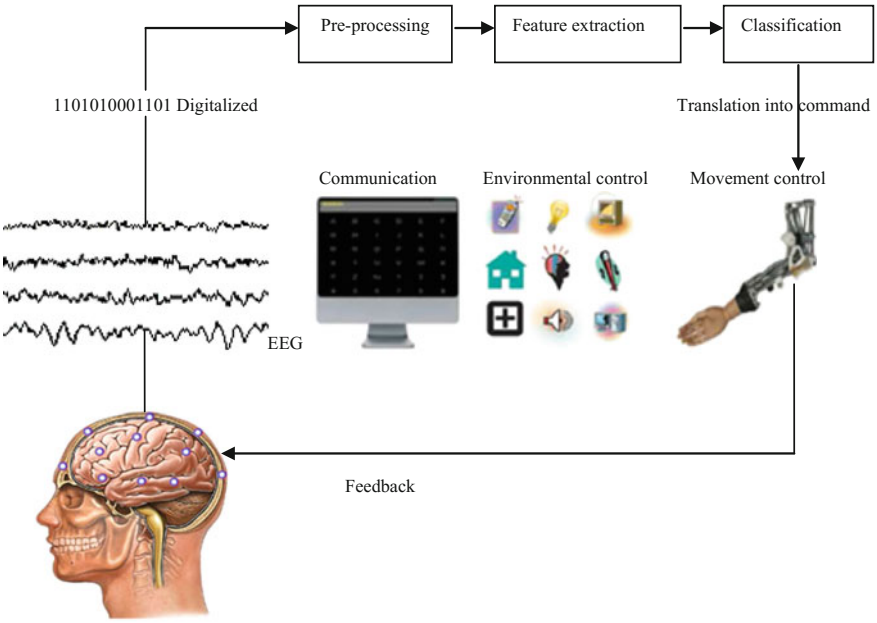


Fig. 2.3 The general architecture of a BCI system

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).

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".

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.

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 user's performance.

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 a 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.

2.11 Significance of EEG Signal Analysis and Classification

When measuring an EEG we often have large amounts of data with different categories, most particularly when the recordings are made 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 through 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 and Ercelebi 2005). Visual inspection of EEG signals is not a satisfactory procedure because there are no standard criteria for the assessments and it is a time-consuming process that can often result in errors due to interpreter 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 communication, it is a challenge for BCI systems to properly and efficiently recognize the intension's patterns of the brain using appropriate classification algorithms.

Given the high variability of the EEG signals in the presence of different subjects and target events (classes), the design of an effective classification system is complex. 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 is observed that most of the reported methods have a limited success rate. Some methods take more time to perform the required computation work and others are very complex for practical applications. Some methods used small sample setting (SSS) data points as a representative of a large number of data points of EEG recordings. Generally these were not representative enough for EEG signal classification. On the other hand, in most of the cases, the reported methods did not select their parameters using a suitable technique, even though parameters significantly affect the classification performance.

Since EEG signals provide significant contributions to biomedical science, a careful analysis of the EEG records is needed to provide valuable insight and to improve understanding of them. 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 task-based BCIs are still far from being fully understood. A considerable amount of neuroscience research is required to achieve this goal. Hence, in this book, we aim to develop methods to classify brain activities in these two areas.

2.12 Concept 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 also for contributing to a better understanding of cognitive processes. An efficient classification technique helps to distinguish EEG segments in the decision making of 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, first, important to extract useful features from raw EEG signals, and then use the extracted features for 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 the given numbers of categories (Duda et al. 2001; Brunelli 2009). An example would be assigning an email 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. *Classifier*s 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 analysis and classification. Measuring brain activity through EEG leads to the acquisition of a large amount of data. In order to obtain the best possible performance, 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

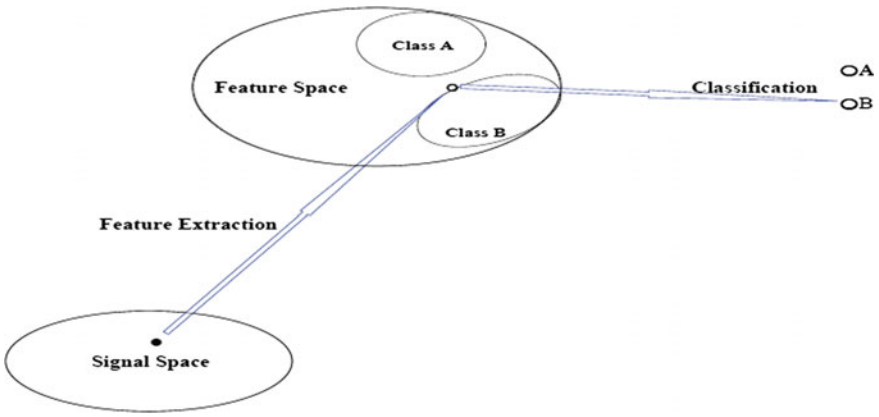


Fig. 2.4 Procedure of classification in biomedical signal processing

an operation which transforms one or several signals into a feature vector. The feature vector, which comprised 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.

The concept of signal classification is depicted in Fig. 2.4. This figure presents a structure of how signals with different categories are classified, extracting features from original data in a 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.

- **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).

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 $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$. Here, $x_1, x_2, \ldots, x_N$ are the observations and $y_1, y_2, \ldots, y_N$ 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. $y \varepsilon \{1, 2, \ldots, L\}$, 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 the *supervised classification* procedure predicting categorical labels are linear discriminant analysis (LDA), support vector machine (SVM), decision trees, naive Bayes classifier, logistic regression, K-nearest-neighbour (*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*. A classifier is constructed using the *training set*. Then the performance of the classifier is evaluated using the *testing* set. This evaluation is sometimes repeated for different parameters of the classifier constructed. This way the parameters of the classifier are 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*. Classification normally refers to a *supervised* procedure. In this research, we have used a *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 Fig. 2.5. The reason for separating the sets is related to

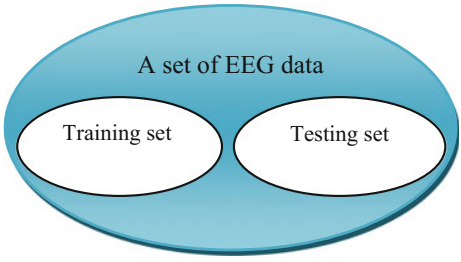


Fig. 2.5 Mutually exclusive training and testing set

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.

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).

2.13 Computer-Aided EEG Diagnosis

EEG signals are complex and analysing them in bulk obtained from a large number of patients makes assessment time consuming. Thus recently the computer-aided diagnosis system is introducing to make it possible to conduct an automatic neurophysiological assessment for detection of abnormalities from EEG signal data. The CAD system (Arimura et al. 2009; Siuly and Zhang 2016) consists of three main steps: *pre-processing*, *feature extraction,* and *classification* as shown in Fig. 2.6. In the *pre-processing*, acquired EEG data are processed for removing

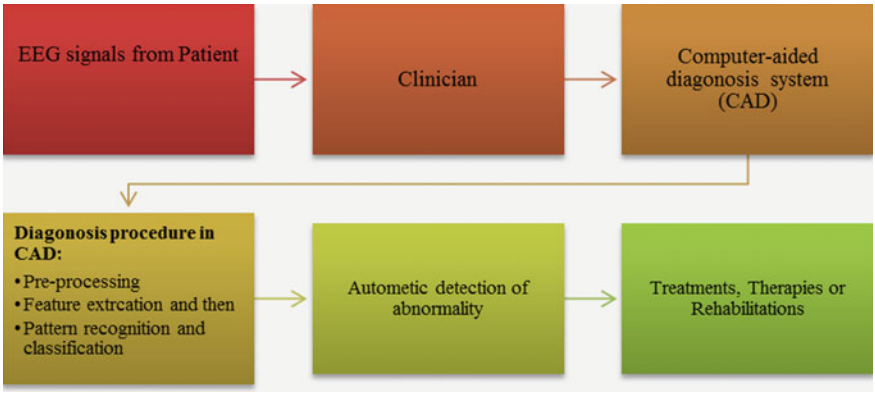


Fig. 2.6 A framework of computer-aided EEG diagnosis

noises, which reduces the complexity and computation time of the CAD algorithms. The *feature extraction* step of the CAD system is one of the most important parts where the biomarkers of disease identification are extracted from the original source data. In the *classification process* for CAD systems, the extracted feature vector is used in the classifier model as input for assigning the candidate to one of the possible categories (e.g. healthy or normal) according to the output of a classifier. Generally, a CAD system can be categorized one of the two types. When a CAD system involves classifying all candidates into two categories, such as abnormal and normal candidates, it is called a two-class categorization system. On the other hand, if a CAD system can classify unknown cases into several types of abnormalities, which are more than two, it is called a multi-class categorization system. Many researchers are working to develop CAD schemes for detection and classification of various kinds of abnormalities from medical data.

Like pattern recognition, the performance of CAD systems is assessed by a *k-fold cross validation test*, bootstrap method, leave-one-out (Jain et al. 2000), etc. The free-response receiver operating characteristic (FROC) and ROC curves are used for evaluation of the overall performance of the CAD systems for various operating points. The FROC curve shows the relationship between the sensitivity and the number of false positives, which can be obtained by thresholding a certain parameter of the CAD system or the output of the classifier (Arimura et al. 2009). Recently, there has been a lot of research performed on the development of CAD systems for detecting neurological problems such as epileptic seizures, dementia, Alzheimer's disease, autism, strokes, brain tumours, alcoholism related neurological disorders and sleeping disorders (Siuly and Zhang 2016).

In this book, our aim is to develop CAD methods for the detection of epilepsy and epileptic seizures from EEG signal data and also for identifying mental states for BCI applications.

References

Abou-Khalil, B., and Misulis, K.E. Atlas of EEG & Seizure Semiology, Elsevier, 2006.

Acharya, U.R., Vidya, S., Bhat, S., Adeli, H., and Adeli, A. Computer-aided diagnosis of alcoholism-related EEG signals, Epilepsy & Behavior 41 (2014) 257–263.

Alexandros T. Tzallas, Markos G. Tsipouras, Dimitrios G. Tsalikakis, Evaggelos C. Karvounis, Loukas Astrakas, Spiros Konitsiotis and Margaret Tzaphlidou, 'Automated Epileptic Seizure Detection Methods: A Review Study', book Published: February 29, 2012.

Alotaiby T N, Alshebeili S A, Alshawi T, Ahmad I, El-Samie F E A (2014) EEG seizure detection and prediction algorithms: a survey. EURASIP Journal on Advances in Signal Processing 2014:183.

Al-Qazzaz N, Ali S, Ahmad S. A., Chellappan K., Islam M. S., Escudero J (2014) Role of EEG as Biomarker in the Early Detection and Classification of Dementia. Scientific World Journal 2014, Article ID 906038, 16 pages.

Amzica F, Lopes da Silva FH. Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. 6. Niedermeyer E, Schomer DL, Lopes da Silva FH, editor.

Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health; 2010. Cellular substrates of brain rhythms; pp. 33–64.

Arimura H, Magome T, Yamashita Y, Yamamoto D (2009) Computer-Aided Diagnosis Systems for Brain Diseases in Magnetic Resonance Images. Algorithms 2: 925–952.

Autism Research, 2014; http://www.einstein.yu.edu/news/releases/1041/brainwave-test-could-improve-autism-diagnosis-and-classification/.

Bajaj, V., Guo, Y., Sengur, A., Siuly, Alcin, O. F. (2016) 'Hybrid Method based on Time-Frequency Images for Classification of Alcohol and Control EEG Signals' *Neural Computing and Applications,* pp 1–7.

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.

Booth CM, Boone RH, Tomlinson G et al. Is this patient dead, vegetative, or severely neurologically impaired? Assessing outcome for comatose survivors of cardiac arrest. JAMA 2004; 291: 870–879.

Brunelli, R. (2009) *Template Matching Techniques in Computer Vision: Theory and Practice*, Wiley, New York.

Brust, J.C. Ethanol and cognition: indirect effects, neurotoxicity and neuroprotection: a review. Int J Environ Res Public Health 7:1540–1557, 2010.

Chen, Z., Cao, Y. Cao, J., Zhang, Y., Gu, F., Guoxian Z., Hong, Z., Wang, and Cichocki, A. An empirical EEG analysis in brain death diagnosis for adults, Cogn Neurodyn. 2008 Sep; 2(3): 257–271.

DeKosky S. T., Marek K (2003) Looking backward to move forward: early detection of neurodegenerative disorders. Science, 302(5646): 830–834.

Duda, R.O., Hart, P.E. and Stork, D.G. (2001) *Pattern Classification*, 2nd edn. John Wiley & Sons, New York.

EEG in Brain Tumours, Medscape, http://emedicine.medscape.com/article/1137982-overview.

Evans BM. Patterns of arousal in comatose patients. J Neurol Neurosurg Psychiatry. 1976;16:392–402. doi:10.1136/jnnp.39.4.392.

Fischer-Williams M, Dike GL. Brain tumours and other space-occupying lesions. Niedermeyer E, DaSilva FL, eds. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. 3rd ed. Williams & Wilkins; 1993. 305–432.

Foreman, B., Claassen, J. Quantitative EEG for the detection of brain ischemia, *Critical Care* 2012, 16:216.

Harper, C. The neurotoxicity of alcohol. Hum Exp Toxicol 26: 251–257, 2007.

Hashemian, H., and Pourghassem, H. Diagnosing Autism Spectrum Disorders Based on EEG Analysis: a Survey, *Neurophysiology*, *Vol. 46*, *No. 2*, *April 2014.*

Jain A K, Duin R P, Mao W, (2000) J. Statistical pattern recognition: Review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22: 4–37.

Johns Hopkins Medicine Health Library, http://www.hopkinsmedicine.org/healthlibrary/test_procedures/neurological/electroencephalogram_eeg_92,p07655/.

Kabir, E., Siuly and Zhang, Y., (2016) 'Epileptic Seizure Detection from EEG signals Using Logistic Model Trees', *Brain Informatics*, 3(2), 93–100.

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.

Levy DE, Caronna JJ, Singer BH et al. Predicting outcome from hypoxicischemic coma. JAMA 1985; 253: 1420–1426.

Lotte, F. (2009) Study of electroencephalographic signal processing and classification techniques towards the use of brain-computer interfaces in virtual reality applications, PhD thesis.

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.

MCDS (Ministerial Council on Drug Strategy) 2011. The National Drug Strategy 2010–2015. Canberra: Commonwealth of Australia.

Minguez C and Winblad B (2010) Biomarkers for Alzheimer's disease and other forms of dementia: clinical needs, limitations and future aspects. Experimental Gerontology 45(1): 5–14.

Musialowicz, T., and Lahtinen, P. Current Status of EEG-Based Depth-of-Consciousness Monitoring During General Anesthesia, Advances in Monitoring for Anesthesia (TM Hemmerling, Section Editor) First Online: 01 May 2014, DOI:10.1007/s40140-014-0061-x.

Neto E, Allen EA, Aurlien H, Nordby H, Eichele T. EEG Spectral Features Discriminate between Alzheimer's and Vascular Dementia. *Front Neurol*. 2015. 6:25.

Nolan MA, Redoblado MA, Lah S, et al. Memory function in childhood epilepsy syndromes. *J Paediatr Child Health*. 2004 Jan–Feb. 40(1–2):20–7.

Ordan KG. Emergency EEG and continuous EEG monitoring in acute ischemic stroke. J Clin Neurophysiol. 2004;16: 341–352.

Perel P, Arango M, Clayton T et al. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. BMJ 2008; 336: 425–429.

Sacco RL, VanGool R, Mohr JP et al. Nontraumatic coma. Glasgow coma score and coma etiology as predictors of 2-week outcome. Arch Neurol 1990; 47: 1181–1184.

Seizure, Drugs.com; https://www.drugs.com/health-guide/seizure.html.

Siuly and Y. Li, (2015), 'Discriminating the brain activities for brain–computer interface applications through the optimal allocation-based approach', *Neural Computing & Applications*, Vol. 26, Issue 4, pp. 799–811.

Siuly and Zhang, Y (2016) Medical Big Data: Neurological Diseases Diagnosis Through Medical Data Analysis, *Data Science And Engineering*, DOI:10.1007/s41019-016-0011-3 (in Press).

Siuly, E. Kabir, H. Wang and Y. Zhang, (2015) 'Exploring sampling in the detection of multi-category EEG signals', *Computational and Mathematical Methods in Medicine,* Volume 2015, Article ID 576437, 12 pages, http://dx.doi.org/10.1155/2015/576437.

Smith, S.J.M. EEG in the Diagnosis, Classification, and Management Of Patients With Epilepsy, *J Neurol Neurosurg Psychiatry* 2005;76 (Suppl II): ii2–ii7. doi:10.1136/jnnp.2005.069245.

Spence, S.J., and Schneider, M.T. The Role of Epilepsy and Epileptiform EEGs in Autism Spectrum Disorders, *Pediatr Res*. 2009 June; 65(6): 599–606.

Staudinger T, Polikar R. Analysis of complexity based EEG features for the diagnosis of Alzheimer's disease. *Conf Proc IEEE Eng Med Biol Soc*. 2011 Aug. 2011:2033–6.

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.

Supriya, Siuly and Y. Zhang (2016) 'Automatic epilepsy detection from EEG introducing a new edge weight method in the complex network', *Electronics Letters*, DOI:10.1049/el.2016.1992 (in press).

Sutter, R., and Kaplan, P.W. Electroencephalographic Patterns in Coma: When Things Slow Down, Epileptologie 2012; 29.

Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet 1974; 304: 81–84.

Tuhrim S, Dambrosia JM, Price TR et al. Prediction of intracerebral haemorrhage survival. Ann Neurol 1988; 24: 258–263.

Urbach H. Imaging of the epilepsies. *Eur Radiol*. 2005 Mar. 15(3):494–500.

Walter G. The location of cerebral tumours by electroencephalography. *Lancet*. 1936. 8:305–8.

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.

Chapter 3
Objectives and Structures of the Book

In the medical and health community, EEG signals are the most utilized signals in the clinical assessment of brain states, detection of epileptic seizures and identification of mental states for BCI systems. A reliable automatic classification and detection system would help to ensure an objective assessment thus facilitating treatments, and it would significantly improve the diagnosis of epilepsy. EEG signals could also be used for the long-term monitoring and treatment of patients. The main goal of this book is to explore the development of several methods that are capable of classifying different categories of EEG signal to help in the evaluation and treatment of brain diseases and abnormalities. This chapter outlines the book's objectives and its structures. This chapter also provides a description of the experimental databases and performance evaluation measures used in this research. Furthermore, this chapter discusses the commonly used methods of EEG signal classification.

3.1 Objectives

The work presented in this book focuses on how different EEG signals from different brain activities can be classified to analyze different brain disorders. We have developed a number of techniques for classifying EEG signals in the epileptic diagnosis and also several techniques for the identification of different categories of MI tasks in BCI applications. The main objective of this study is to develop methods for identifying different EEG signals. To investigate the performances of those techniques, we also compare our proposed algorithms with other recently reported algorithms.

Our main objective in this book is achieved through focusing on three goals:

- Introducing methods for analyzing EEG signals for the detection of epileptic seizures

S. Siuly et al., *EEG Signal Analysis and Classification*,
Health Information Science, DOI 10.1007/978-3-319-47653-7_3

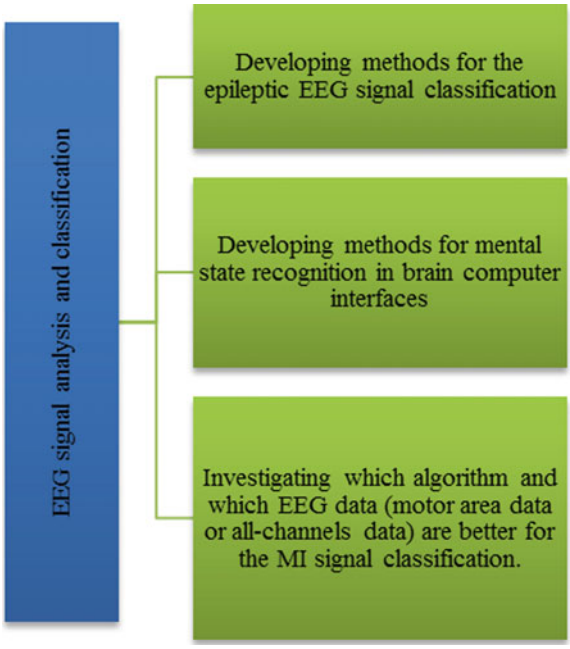


Fig. 3.1 Objectives addressed in this book

- Developing methods for the identification of mental states in brain computer interfaces (BCIs)
- Investigating which algorithm and which kind of EEG data (motor area data or all-channels data) are better for motor imagery (MI) signal classification (Fig. 3.1).

To achieve Goals 1 and 2, we will introduce methods for feature extraction, segmentation analysis and information retrieval. The methods are based on algorithms developed from statistical techniques and machine learning methods, and are specifically designed to deal with the characteristics of EEG signals. The performances will be studied by means of empirical analyses. To achieve Goal 3, we will apply our developed algorithms for EEG based MI signal classification on data from the motor cortex area and all-channel EEG data; and will then make a decision about which algorithm is most suited for which data type.

3.2 Structure of the Book

The examination of EEG signals has been recognized as the most preponderant approach to the problem of extracting knowledge of brain dynamics. EEG recordings are particularly important in the *diagnosis of epilepsy* and in *brain computer interfaces* (*BCI*). The main use of EEGs is to detect

and investigate epilepsy, a condition that causes repeated seizures. An EEG will help physicians to identify the type of epilepsy a patient has, what may be triggering seizures and how best to treat the patient. In BCI systems, EEG signals help to restore sensory and motor function in patients suffering with severe motor disabilities. Analyzing EEG signals is very important both for supporting the diagnosis of brain diseases and for contributing to a better understanding of mental state identification of motor disabled people in BCI applications. Correctly and efficiently identifying different EEG signals for diagnosing abnormalities in health is currently a major challenge for medical science. Hence, this book intends to present advanced methods for the analysis and classification of epileptic EEG signals and also for the identification of mental state based on MI EEG signals in BCI's development.

Part I provides a basic overview of EEG signals, including its concepts, generation procedure, characteristics, nature and abnormal patterns. This part also provides a discussion of different applications of EEG signals for the diagnosis of brain diseases and abnormalities. In addition, we provide the aims of this book, a description of the experimental data sets and performance evaluation measures used in this research and a short review of commonly used methods in the EEG signal classification. Part II presents our developed techniques and models for the detection of epileptic seizures through EEG signal processing. The implementation of these proposed methods in the real-time databases will be highlighted. In Part III, we introduce the methods for identifying mental states from EEG data designed for BCI systems and their applications in some benchmark datasets. Part III also reports the experimental procedures and the results of each methodology. Finally, Part IV

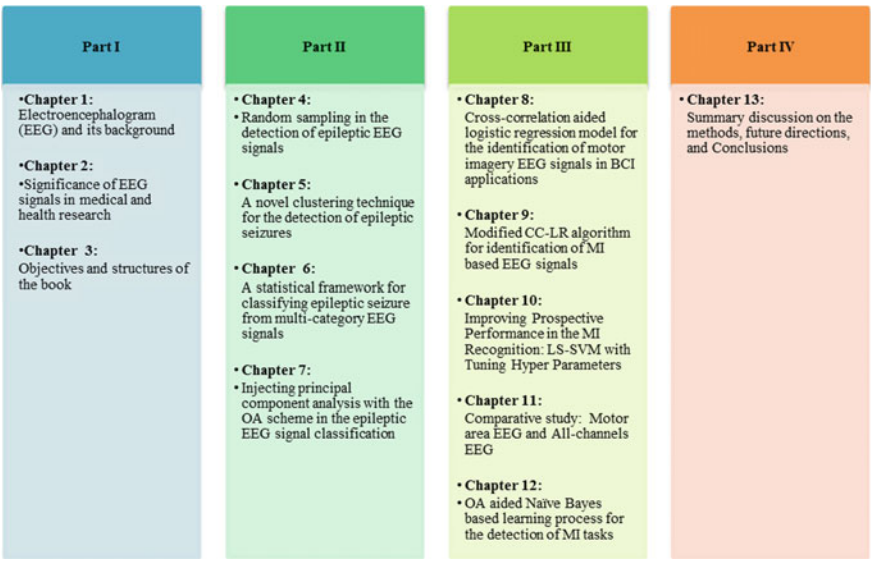


Fig. 3.2 Structure of the book's four parts

provides an overall discussion about EEG signal analysis and classification. This part gives a summary discussion on the developed methods, suggests future directions in the EEG signal analysis area and concludes with recommendations for further research. Figure 3.2 displays a brief outline of each of the four parts using chapter titles.

3.3 Materials

To achieve the objectives of the book, a set of methods is developed and implemented in various EEG signal databases. Datasets used are described in Sect. 3.3.1. Section 3.3.2 provides a brief description of the performance evaluation measures used. Several methods are commonly used in the analysis and classification of EEG signals. Section 3.4 reports these methods. Our research develops methods for *the epileptic EEG signal classification* and also for *the MI-based EEG signal recognition in BCI systems*. Section 3.4.1 reports the methods which are used in the epileptic EEG data for feature extraction and classification. The methods for the classification of MI tasks in BCI systems are described in Sect. 3.4.2.

3.3.1 Analyzed Data

3.3.1.1 The Epileptic EEG Data

The epileptic EEG data, developed by the Department of Epileptology, University of Bonn, Germany, and described in Andrzejak et al. (2001), is publicly available (EEG time series 2005). The whole database consists of five EEG data sets (denoted as Set A to Set E), each containing 100 single-channel EEG signals of 23.6 s from five separate classes. Each signal was chosen after visual inspection for artifacts such as the 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 digitized at 173.61 data points per second using 12-bit resolution. The band-pass filter settings were 0.53–40 Hz (12 dB/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 collected from the EEG records of the pre-surgical diagnosis of five epileptic patients. 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. Table 3.1 presents a summary description of the five set EEG data. 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 micro volts.

Table 3.1 Summary of the epileptic EEG data (Siuly and Li, 2014)

Subjects	Set A	Set B	Set C	Set D	Set E
	Five healthy subjects	Five healthy subjects	Five epileptic subjects	Five epileptic subjects	Five epileptic subjects
Patient's state	Awake and eyes open (normal)	Awake and eyes closed (normal)	Seizure-free (interictal)	Seizure-free (interictal)	Seizure activity (ictal)
Electrode type	Surface	Surface	Intracranial	Intracranial	Intracranial
Electrode placement	International 10–20 system	International 10–20 system	Opposite to epileptogenic zone	Within epileptogenic zone	Within epileptogenic zone
Number of channels	100	100	100	100	100
Time duration (s)	23.6	23.6	23.6	23.6	23.6

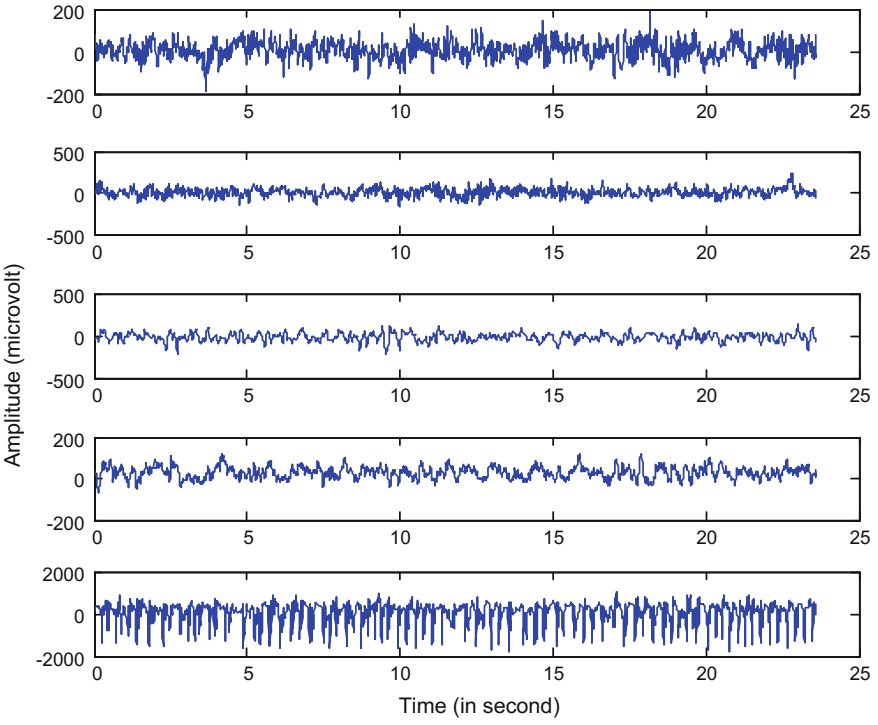


Fig. 3.3 Example of five different sets of EEG signals (from *top* to *bottom*: Set A, Set B, Set C, Set D and Set E) (Siuly et al. 2014)

3.3.1.2 Dataset IVa of BCI Competition III

Dataset IVa (BCI competition III 2005; Blankertz et al. 2006) was recorded from five healthy subjects (labelled **aa**, **al**, **av**, **aw**, **ay**) who performed right hand (class 1) and right foot (class 2) MI tasks. 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 s. 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.

3.3.1.3 Dataset IVb of BCI Competition III

Dataset IVb (BCI competition III; 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 the seven initial sessions without feedback. The EEG data consisted of two classes: left hand and right foot MI. 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. They provided a version of the data that was down-sampled at 100 Hz, which is used in this research.

3.3.1.4 Mental Imagery EEG Data of BCI Competition III

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 were 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; Chiappa and 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 sat in a normal chair, with relaxed arms resting on their legs. For a given subject, four non-feedback sessions were recorded on the same day, each lasting four minutes or so with breaks of 5–10 min between each session. The subjects performed a given task for about 15 s and then switched randomly to the next task at the operator's request (Chiappa and Millán 2005). The sampling rate of the raw EEG potential signals was 512 Hz. The signals were first spatially filtered by means of a surface

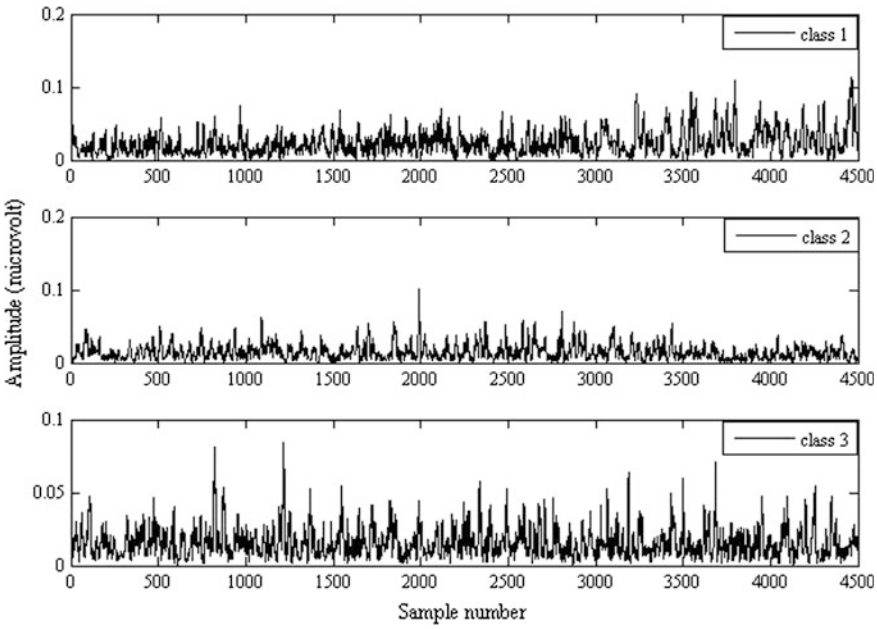


Fig. 3.4 Exemplary EEG signals for left hand movements (class 1), right hand movements (class 2) and word generations (class 3) taken from Subject 1 (Siuly et al. 2011a, b)

Laplacian. Then, every power spectral density in the band of 8–30 Hz was 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 were obtained from each of the 8 channels, producing a 96 dimensional vector.

3.3.1.5 Ripley Data

This data contains a synthetic two-class problem from Ripley data which was used in Ripley (1996). The dataset is publicly available in Ripley (1996). 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 either 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.

3.3.2 Performance Evaluation Parameters

Criteria for evaluating the performance of a methodology are an important part of its design. In pattern recognition area, there are various types of measures for the evaluation of performance. In this book, the stability of the performance of a method is assessed on the basis of standard criteria that are used in biomedical signal analysis. These include classification accuracy, sensitivity (or true positive rate (TPR) or recall), specificity, false alarm rate (FAR) and receiving operating characteristic (ROC) curve. To reduce overfitting, we also use the *k*-fold cross-validation procedure for evaluating the performance of the developed methods. These criteria allow the estimation of the behaviour of the classifiers on the extracted feature data. The definitions of these performance parameters (Siuly et al. 2011b; Guo et al. 2009; Siuly and Li 2012, 2015) are provided below:

- **Classification accuracy**: the number of correct decisions divided by the total number of cases
- **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
- **FAR**: the percentage of false-positives predicted as positive from negative class
- **ROC**: 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).
- ***k*-fold cross-validation procedure**: Cross-validation is a model validation technique for assessing the results of an analysis. It is generally used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. The cross-validation process involves partitioning data into complementary subsets, performing the analysis on one subset (called the *training set*) and validating the analysis on the other subset (called the *validation set* or *testing set*). To reduce variability, multiple rounds of cross-validation are performed using different partitions, and the validation results are averaged over the rounds.
  In the *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) (Abdulkadir 2009; Ryali et al. 2010; Siuly and Li 2012). Each time, one of the subsets is used as a testing set and the other $k - 1$ subsets are put together

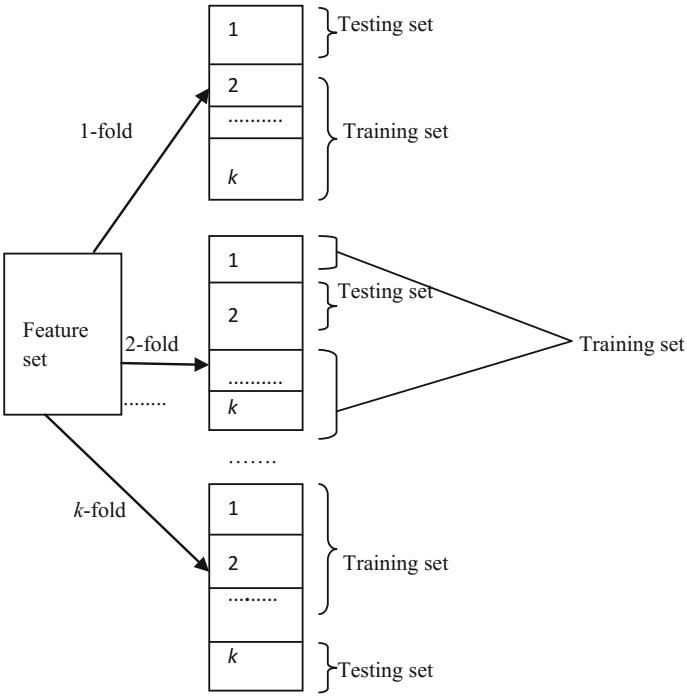


Fig. 3.5 Partitioning design of the *k*-fold cross-validation method

to form a training set. Then the average accuracy across all *k* trials is computed. Figure 3.5 presents a design of how the extracted feature vectors of this study are partitioned into *k* mutually exclusive subsets according to the *k*-fold cross-validation system. As shown in Fig. 3.5, the feature vector set is divided into *k* subsets and the procedure is repeated *k* times (the folds). Each time, one subset is used as a testing set and the remaining nine subsets are used as a training set, as illustrated in the figure. The results of each of k times on the testing set are averaged over the iterations called '*k*-fold cross-validation performance'.

One of the main reasons for using cross-validation instead of using the conventional validation (e.g. partitioning the data set into two sets of 70% for training and 30% for test) is that there is not enough data available to partition it into separate training and test sets without losing significant modelling or testing capability. In these cases, a fair way to properly estimate model prediction performance is to use cross-validation as a powerful general technique.

3.4 Commonly Used Methods for EEG Signal Classification

3.4.1 Methods for Epilepsy Diagnosis

For efficient classification, an accurate feature extraction method is very important to extract good features from original signals. In fact, if the features extracted from EEGs are not relevant and do not accurately describe the EEG signals employed, a classification algorithm using such features will have trouble identifying the classes of the features (Siuly and Zhang 2016). 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 the autoregressive (AR) model (Ubeyli 2009a), moving average (MA) model (Ubeyli 2009a), autoregressive-moving average (ARMA) model (Ubeyli 2009a) and lapunov exponents (Guler et al. 2005; Murugavel et al. 2011). 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 and Marple 1981; Proakis and 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 PSD estimation. Two common *non-parametric methods,* periodogram and the correlogram (Ubeyli 2009a), provide a 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), wavelet transform (WT) (Adeli and Dadmehr 2003; Subasi et al. 2005b; Ubeyli 2009b; Jahankhanni et al. 2006; Murugavel et al.

2011), wavelet energy function (Guo et al. 2009), wavelet entropy (Shen et al. 2013), Cohen class kernel functions (Acharjee and Shahnaj 2012) and visibility graph (Supriya et al. 2016; Zhu et al. 2014). Since Fourier methods may not be appropriate for nonstationary 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 and 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 and Manolakis 1996; Stoica and Moses 1997; Ubeyli and Guler 2003).

In the last a few years, these feature extraction methods have been combined with different types of classifiers. Examples include adaptive neuro-fuzzy inference system (Guler and Ubeyli 2005), support vector machine (SVM) (Makinac 2005; Burges 1998; Guler and Ubeyli 2007; Chandaka et al. 2009; Silver et al. 2006; Fan et al. 2006; Shen et al. 2013; Murugavel et al. 2011; Supriya et al. 2016; Siuly et al. 2015), least square support vector machine (LS-SVM) (Ubeyli 2010; Siuly et al. 2009, 2010, 2011a, b; Hanbay 2009; Siuly and Li 2014; Siuly and Li 2015) and artificial neural network (ANN) (Guler et al. 2005; Subasi 2007; Ubeyli 2008; Jahankhani et al. 2006; Subasi and Ercelebi 2005a; Guo et al. 2009; Acharjee and Shahnaj 2012), 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; Murugavel et al. 2011), mixture of experts (MEs) (Ubeyli 2009a), modified mixture of experts (MMEs) (Ubeyli 2009a), *k*-NN (Shen et al. 2013) and logistic tree model (Kabir et al. 2016). 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 identify the characteristics of data that determine classifier performance. The measures of accuracy and confusion matrix are very popular methods of evaluating 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.

3.4.2 Methods for Mental State Recognition in BCIs

In BCI applications, several methods have been studied and employed for feature extraction from MI based EEG signals. These include autoregressive (AR) (Schloglel al. 2002; Pfurtschelleret al. 1998; Burke et al. 2005; Guger et al. 2001; Blankertz et al. 2006), fast Fourier transform (FFT) (Polat et al. 2007), common spatial patterns (CSP) (Blankertz et al. 2006; Blanchard et al. 2004; Lemm et al. 2005; Lotte and Guan 2011; Lu et al. 2010; Zhang et al. 2013), regularized common spatial patterns (R-CSP) (Lu et al. 2009; Lotte and Guan 2011), spatiospectral patterns (Wu et al. 2008), wavelet coefficients (Qin et al. 2005; Ting et al. 2008), iterative spatiospectral patterns learning (ISSPL) (Wu et al. 2008), fisher ratio criterion (Song and Epps 2007), Bayesian spatiospectral filter optimization (BSSFO)-based bayesian framework (Suk and Lee 2013). Feature extraction methods based on self organizing maps (SOM) using autoregressive (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 PCA and ICA feature selection methods were used (Sanei and Chambers 2007; Wang and James 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) techniques 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).

Five different categories of classifiers: *linear classifiers*, *neural networks*, *nonlinear Bayesian classifiers*, *nearest neighbour classifiers* and *combinations of classifiers* have been studied in 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) (Blankertz et al. 2006; Yong et al. 2008; Zhang et al. 2013) 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 a 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, Ming et al. (2009) studied the performance of the 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 produced the best result with an accuracy of 91%.

There are two types of *Nonlinear Bayesian classifiers* used in BCI systems; Bayes quadratic and the 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 of the dominant class among its *k* nearest neighbours 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 variance and thus increase 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, 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), logistic regression (Siuly et al. 2013; Siuly and Li 2014), Naïve bayes (Siuly et al. 2015, 2016), least square support vector machine (Siuly and Li 2012, 2014a), and Fisher classifier and Fuzzy logic. The PSO has been incorporated in reference (Satti et al. 2009) to select a subject specific-frequency band for an efficiently tuned BCI system. The Fisher classifier was used by Bian et al. (2010) to identify 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.

References

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

Acharjee P. P. and Shahnaz C. (2012) 'Multiclass Epileptic Seizure Classification Using Time-Frequency Analysis of EEG Signals', 7th International Conference on Electrical and Computer Engineering, 20-22 December, 2012, Dhaka, Bangladesh, pp. 260–263.

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.

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

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.

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.

BCI competition III, 2005, http://www.bbci.de/competition/iii.

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.

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.

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.

Chandaka, S., Chatterjee, A. and Munshi, S. (2009) 'Cross-correlation aided support vector machine classifier for classification of EEG signals', *Expert System with Applications,* Vol. 36, pp. 1329–1336.

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.

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.

Duda, R.O., Hart, P.E. and Stork, D.G. (2001) *Pattern Classification,* 2nd edn. John Wiley & Sons, New York.

EEG time series, 2005, [Online], Viewed 30 September 2008, http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html.

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.

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.

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, 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.

Haselsteiner, E. and Pfurtscheller, G. (2000) 'Using Time-Dependant Neural Networks for EEG classification', *IEEE Transactions on Rehabilitation Engineering*, Vol. 8, pp. 457–463.

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.

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)*.

Kabir, E., Siuly, and Zhang, Y. (2016) 'Epileptic seizure detection from EEG signals using logistic model trees', *Brain Informatics*, 3(2), pp. 93–100.

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.

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.

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.

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.

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.

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.

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.

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.

Makinac, M. (2005) 'Support Vector Machine Approach for Classification of Cancerous Prostate Regions', *World Academy of Science, Engineering and Technology*, 7.

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.

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. 1026–1033.

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.

Murugavel A.S.M, Ramakrishnan S., Balasamy K. and Gopalakrishnan T. (2011) 'Lyapunov features based EEG signal classification by multi-class SVM' 2011 World Congress on Information and Communication Technologies,197–201.

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., 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.

Proakis, J. G., & Manolakis, D. G. (1996) *Digital Signal Processing Principles, Algorithms, and Applications*, New Jersey: Prentice Hall.

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.

Ripley, B.D. (1996) *Pattern recognition and neural networks. Cambridge*, Cambridge University Press.

Ripley data-online, http://www.stats.ox.ac.uk/pub/PRNN/.

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. (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.

Shen C. P., Chen C. C., Hsieh S. L., Chen W. H., Chen J. M., Chen C. M., Lai F. and Chiu M. J. (2013) 'High-performance seizure detection system using a wavelet-approximate entropy-fSVM cascade with clinical validation', Clinical EEG and Neuroscience, DOI: 10.1177/1550059413483451.

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. and Wen, P. (2011a) '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. (2011b) '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 Y. Li, (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, Vol. 20, no. 4, pp. 526–538.*
Siuly, and Li, Y. (2014) 'A novel statistical algorithm for multiclass EEG signal classification', *Engineering Applications of Artificial Intelligence*, Vol. 34, pp. 154–167.
Siuly and Y. Li, (2015), 'Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification', *Computer Methods and programs in Biomedicine, Computer Methods and programs in Biomedicine*, Vol. 119, pp. 29–42.
Siuly, S., Kabir, E., Wang, H. and Zhang, Y. (2015) 'Exploring Sampling in the Detection of Multicategory EEG Signals', *Computational and Mathematical Methods in Medicine*, 2015, pp. 1–12.
Siuly, S. and Zhang, Y. (2016) 'Medical Big Data: Neurological Diseases Diagnosis Through Medical Data Analysis', *Data Science and Engineering*, DOI:10.1007/s41019-016-0011-3, pp. 1–11.
Siuly, Wang, H. and Zhang, Y. (2016) 'Detection of motor imagery EEG signals employing Naïve Bayes based learning process', *Measurement*, 86, pp. 148–158.
Siuly, N., Li, Y. and Wen, P. (2013) 'Identification of motor imagery tasks through CC-LR algorithm in brain computer interface', *International Journal of Bioinformatics Research and Applications*, 9(2), p. 156.
Siuly, S. and Li, Y. (2014a) 'Discriminating the brain activities for brain–computer interface applications through the optimal allocation-based approach', *Neural Comput & Applic*, 26(4), pp. 799–811.
Siuly, Li, Y. and (Paul) Wen, P. (2014) 'Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain–computer interface. *Computer Methods and Programs in Biomedicine'*, 113(3), pp. 767–780.
Song L, Epps J, (2007) 'Classifying EEG for brain-computer interface: learning optimal filters for dynamical features', *Comput Intell and Neurosci,* Article ID 57180, 11 pages, doi:10.1155/2007/57180, 2007.
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.
Supriya, S., Siuly, S. and Zhang, Y. (2016) 'Automatic epilepsy detection from EEG introducing a new edge weight method in the complex network', Electronics Letters. DOI:10.1049/el.2016.1992 *(in press).*

Suk, H. and Lee, S.W. (2013) 'A Novel Bayesian framework for Discriminative Feature Extraction in Brain-Computer Interfaces', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, no. 2, 286–299.

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.

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.

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.

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.

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.

Zhang, R., Xu, P., Guo, L., Zhang, Y., Li, P. and Yao, D. (2013) 'Z-Score Linear Discriminant Analysis for EEG Based Brain-Computer Interfaces', *PLoS ONE*, Vol. 8, no. 9, e74433.

Zhu, G., Li, Y., Wen, P. (2014) 'Epileptic seizure detection in EEGs signals using a fast weighted horizontal visibility algorithm', *Computer Methods and Programs in Biomedicine, Vol.* 115, pp. 64–75.

Part II
Techniques for the Diagnosis of Epileptic Seizures from EEG Signals

Chapter 4
Random Sampling in the Detection of Epileptic EEG Signals

The detection of epileptic EEG signals is a challenging task due to bulky size and nonstationary nature of the data. From a pattern recognition point of view, one key problem is how to represent the large amount of recorded EEG signals for further analysis such as classification. It is, first, important to extract useful features from raw EEG signals, and then use the extracted features for classification. The literature reports numerous signal processing techniques being employed for the feature extraction and 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. To evaluate the performance of the proposed method, we tested it in the benchmark epileptic EEG database (EEG time series 2005; Andrzejak et al. 2001) and have reported the results in Sect. 4.3. For further assessment, we also evaluated the proposed method in the mental imagery EEG dataset: BCI Competition III (Data set V) (Chiappa and Millán 2005) and in Ripley data (1996), and these results are also provided in the experimental results section.

4.1 Why Random Sampling in Epileptic EEG Signal Processing?

The idea of using random sampling for feature extraction is completely new in epileptic seizure diagnosis. Sampling allows experts to work with a small, manageable amount of data to build and run analytical models that produce accurate findings more quickly. Sampling is particularly useful with data sets that are too large to efficiently analyze in full. To ensure reliable and valid inferences from a sample, the probability sampling technique is used to obtain unbiased results

S. Siuly et al., *EEG Signal Analysis and Classification*,
Health Information Science, DOI 10.1007/978-3-319-47653-7_4

(Suresh et al. 2011). The most commonly used probability sampling method used in medicine is SRS, which is introduced in this research work.

In the SRS approach, a representative sample is randomly selected from the population and each individual has the same probability of being chosen. An effective sample of a population represents an appropriate extraction of the useful data which provides meaningful knowledge of the important aspects of the population. Owing to randomization, it is free from conscious and unconscious bias that researchers may introduce while selecting a sample (Siuly et al. 2009, 2011). Usually EEG recordings contain a huge amount of data. Due to the high volume, velocity and complexity of EEG data, it is very difficult for experts to accumulate, manage and analyze data for diagnosis and planning. Condensation of high quantity EEG data is the grand challenge for the experts who deliver clinical recommendations. Thus, this study intends to present the SRS technique for obtaining representative samples from each category of the EEG data. Using the SRS technique, we compress a large amount data into fewer parameters, which are termed 'features'. These features represent the behaviours of the EEG signals, which are particularly significant for recognition and diagnostic purposes. In this research, all the EEG data of a class is considered as a population where a sample is considered as a representative part of the population. Thus, a sample is a set of observations from a parent population. An observation in a sample is called sample unit, and the sample size is the number of observations that are included in a sample.

In the SRS approach, the most importing thing is to determine the required size of observations for a sample to describe the characteristics of the whole dataset. Because, it is not known in the sampling process how many sample units are required to represent the whole population. Table 4.1 provides our evidence for the choice of the SRS scheme in this research work. Table 4.1 presents an example of the required sample size for the population with different sizes using a sample size calculator (*Ref.* Sample Size Calculator). This table shows that the increment of the sample size is not with the same rate as the population size. It is seen, from this table, that the sample size can reach a maximum of 9604 at 95% confidence level and 16,641 at 99% confidence level when the population size is 100,000 or as large

Table 4.1 An example of the required sample size for various sizes of population (Siuly and Li 2014)

Population size	Sample size	
	99–100% confidence interval	
	95% confidence level	99% confidence level
500	475	485
1000	906	943
5000	3288	3845
10,000	4899	6247
50,000	8057	12,486
100,000	9604	16,641
5,000,000	9604	16,641
10,000,000	9604	16,641

as possible. Thus, the SRS technique is appropriate for any large size of population data such as EEG signal data analysis and classification.

Until now, the literature has proposed several techniques for the classification of EEG signals, and diverse classification accuracies have been reported for epileptic EEG data during the last decade. Due to the large data quantity involved, traditional methods or data mining approaches may not be applicable for reliable processing. Hence, in this chapter, we propose a new approach based on the SRS technique and the least square support vector machine (LS-SVM) named SRS-LS-SVM to classify epileptic EEG signals (described in the following section). As simple random samples are a representative part of a population, it is a natural expectation that selecting samples from EEG signals may well represent original data which can help to improve the performance of the classification method. 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.

4.2 Simple Random Sampling Based Least Square Support Vector Machine

In this chapter, we develop a combined algorithm based on SRS and least square support vector machine (LS-SVM) for identifying epileptic seizure in EEG. In this process, we employ the SRS technique for feature extraction, and LS-SVM for EEG signal classification. The block diagram of the proposed method is depicted in Fig. 4.1. This figure shows the different steps of the proposed EEG signal classification method. The proposed method 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 the EEG signals. These parts are described in detail in the following three subsections.

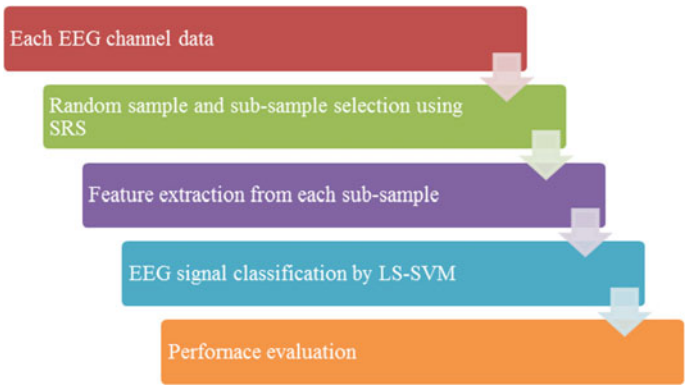


Fig. 4.1 Block diagram of the SRS-LS-SVM method for EEG signal classification

4.2.1 Random Sample and Sub-sample Selection Using SRS Technique

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, it will represent the characteristics of the population. Different types of sampling techniques are used in statistics (Cochran 1977; Islam 2007). The application of these techniques depends 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 Sect. 4.2.2. Figure 4.2 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 set, 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 obtained in the first step. Here m is the number of random sub-samples and $m \geq 2$.

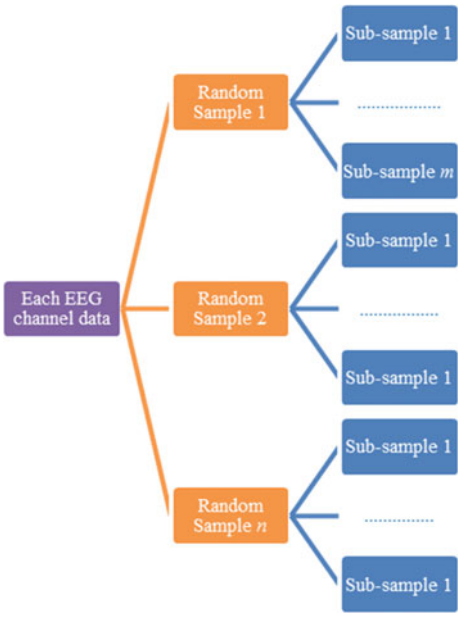


Fig. 4.2 Random sample and sub-sample selection diagram using SRS technique

For any applications, the number of sample (n) and sub-sample (m) selections are chosen based on an empirical approach. The consistency of the results will be improved if the number of samples is sufficiently large. In Steps 1 and 2, the sizes of each random sample and sub-sample are determined using the following formulas (Eqs. 4.1 and 4.2) (Islam 2007; De Veaux et al. 2008; Sample Size Calculator) for each EEG class's data.

$$n = \frac{Z^2 \times p \times (1-p)}{e^2} \tag{4.1}$$

where n = desired sample size (or sub-sample size); Z = standard normal variate (the value for Z is found in statistical tables which contain the area under the normal curve) for the desired confidence level (1.96 for 95% confidence level and 2.58 for 99% confidence level) (see Z Distribution Table); p = estimated proportion of an attribute that is present in the population; e = margin of error or the desired level of precision (e.g. e = 0.01 for 99–100% confidence interval). If the population is finite, the required sample size (or sub-sample size) is given by

$$n_{\text{new}} = \frac{n}{1 + \frac{n-1}{N}} \tag{4.2}$$

where N = population size. If the estimator p is not known, 0.50 (50%) is used as it produces the largest sample size. The larger the sample size, the more sure we can be that the answers truly reflect the population. In this research, we consider p = 0.50 so that the sample size is at a maximum and Z = 2.58 and e = 0.01 for 99% confidence level in Eq. (4.1). It is worth mentioning that, in the calculation of a random sample size, EEG channel data is considered as a population while the random sample is considered as a population for the calculation of the size of sub-sample size. In this study, we obtained the size of each random sample: n_{new} = 3287 when N = 4096 in each class and obtained the size of each sub-sample: n_{new} = 2745 when N = 3287. 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 section.

4.2.2 Feature Extraction from Different Sub-samples

The goal of the feature extraction is to pull out features (special patterns) from the original data to achieve 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 identify the key features that represent the whole dataset, depending on the characteristics of that dataset.

In this chapter, nine statistical features are extracted from each sub-sample data point 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* (IQR) 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 of the characteristics of a dataset.

Again, a database can be symmetric or skewed. In this study, some of the EEG data files used have symmetric distributions while others have skewed distributions. For a symmetric distribution, an appropriate method 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 IQR are the appropriate methods for measuring the centre and spread of the data. On the other hand, *mode* is the most frequent value, and is also a measure of locations in a series of data. Like *mean* and *median*, *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 EEG signals and 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.

4.2.3 *Least Square Support Vector Machine (LS-SVM) for Classification*

The least square support vector machine (LS-SVM) is a relatively new powerful tool in the field of biomedical, employed for classification purposes. 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 for solving 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 computational time associated with problem solving. 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

$$y(x) = \text{sign}[w^T \phi(x) + b]. \tag{4.3}$$

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)

$$\text{Min } J(w, b, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{N} e_i^2. \tag{4.4}$$

Subject to the equality constraint

$$y_i[w^T \phi(x_i) + b] = 1 - e_i, \quad i = 1, 2, \ldots, N. \tag{4.5}$$

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 Eq. (4.4) (Suykens et al. 2002; Guo et al. 2006) as

$$L(w, b, e; \alpha) = J(w, b, e) - \sum_{i=1}^{N} \alpha_i \{y_i[w^T \phi(x_i) + b] - 1 + e_i\}. \tag{4.6}$$

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:

$$\begin{aligned}
&\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i \phi(x_i), \\
&\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0, \\
&\frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i, \quad \text{for } i = 1, \ldots, N, \\
&\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \phi(x_i) + b + e_i - y_i = 0, \quad \text{for } i = 1, \ldots, N
\end{aligned}$$

Now we get

$$w = \sum_{i=1}^{N} \alpha_i \phi(x_i) = \sum_{i=1}^{N} \gamma e_i \phi(x_i) \tag{4.7}$$

Putting the value of w from Eqs. (4.7) in (4.3), the following result is obtained:

$$y(x) = \sum_{i=1}^{N} \alpha_i \phi(x_i)^T \phi(x) + b = \sum_{i=1}^{N} \alpha \langle \phi(x_i)^T, \phi(x) \rangle + b \tag{4.8}$$

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

$$\begin{pmatrix} 0 & 1_v^T \\ 1 & K + \frac{1}{\gamma} I \end{pmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}, \tag{4.9}$$

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

It is seen from Eq. (4.9), all Lagrange multipliers (support vectors) are nonzero, which means that all training data contributes to the solution. After applying of the Mercer condition, the decision function of the LS-SVM is then constructed for the classification as follows:

$$y(x) = \text{sign}\left(\sum_{i=1}^{N} y_i \alpha_i K(x, x_i) + b \right), \tag{4.10}$$

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 provided in Sect. 4.3. There are different types of kernel function. For example linear kernel, polynomial kernel, radial basis function (RBF) and multi-layer perception 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 biomedical researchers consider this function to be ideal.

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 a MATLAB environment (version 7.7, R2008b) using the LS-SVMlab toolbox, available in (LS-SVMlab toolbox (version 1.5)-online).

4.3 Experimental Results and Discussions

The proposed methodology is evaluated on the epileptic EEG dataset. To further evaluate, the method is also tested on the mental imagery tasks EEG dataset of BCI competition III (Data set V) and Ripley data (1994a). The descriptions of these datasets are provided in Chap. 3. The performance of the proposed scheme is

assessed based on different statistical measurements such as sensitivity, specificity, classification accuracy and a receiving operating characteristic (ROC) curve. The description of these parameters is provided in Sect. 3.3.2 in Chap. 3. MATLAB version 7.7.0 (2008b) is used for all implementations. The experimental results of the three datasets are discussed below.

4.3.1 Results for Epileptic EEG Datasets

The epileptic EEG data consists of five sets (Set A-Set E) having 100 single-channels of EEG signals (refer to Sect. 3.3.1.1 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 five 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*, IQR and *standard deviation*) are calculated from each sub-sample. Thus we obtain a feature vector set of size [1 × 45] from the five 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.

Figures 4.3, 4.4, and 4.5 show examples of the mean feature points of the EEG recordings from the healthy subjects with eyes open (Set A), epileptic patients during a 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 Figs. 4.3, 4.4, and 4.5, it is noted that only 4000 feature points out of 45,000 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 epileptic EEG 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 remaining 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

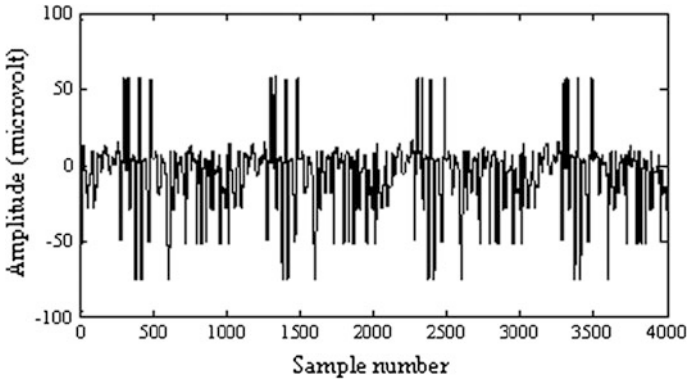


Fig. 4.3 Exemplary mean feature points obtained by the SRS from healthy subjects with eyes open (Set A)

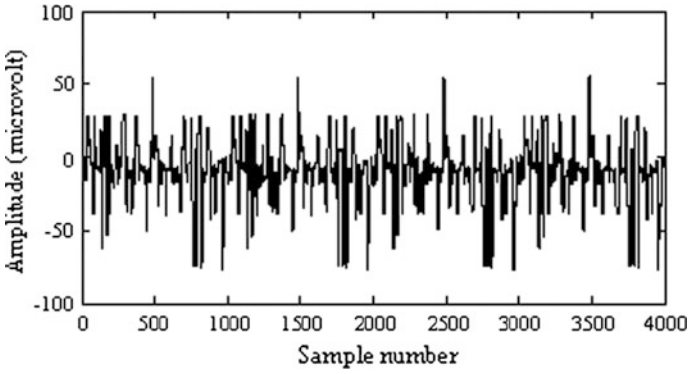


Fig. 4.4 Exemplary mean feature points obtained by the SRS from epileptic patients during seizure-free intervals within the hippocampal formation of the opposite hemisphere of the brain (Set C)

vectors with the same dimensions as the testing set. In Eq. (4.10), 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 the different kernel functions that were tested.

There are two important parameters (γ, σ^2) in the LS-SVM, which should be appropriately chosen to achieve the desired performance. The values of the two parameters significantly affect the classification performance of the LS-SVM. To achieve the best results, LS-SVM is trained with different combinations of the parameters, γ and σ^2. The proposed method is conducted with different pairs of the

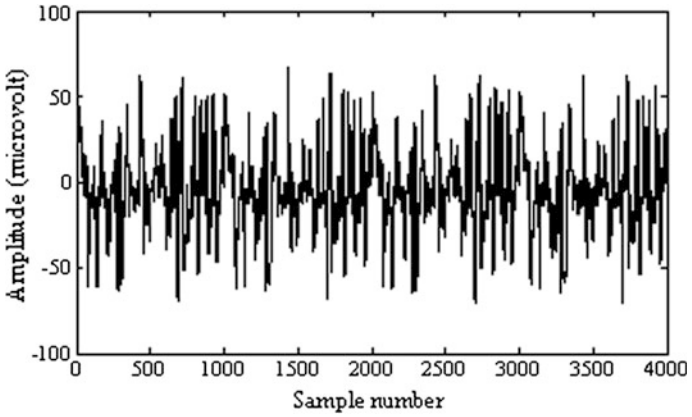


Fig. 4.5 Exemplary mean feature points obtained by the SRS from epileptic patients during seizure activity (Set E)

five EEG data sets in the epileptic EEG data and the best classification result is obtained for the pair of Set A and Set E when $\gamma = 1$ and $\sigma^2 = 1$ with a 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 epileptic EEG datasets are shown in Table 4.2. Here it is observed that the highest classification accuracy obtained is 100% for both 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 easy to classify Set A and Set E as demonstrated by the 100% 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 4.2, that 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 are more analogous to each other. On the other hand, the method obtains 94.00% classification accuracy for Set D and Set E. It is also noted, from Table 4.2, that the area values under the 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 for classifying EEG signals and produces excellent results for classifying EEG brain signals between Set A and Set E. As shown in Table 4.2, 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 study is a very promising technique for the EEG signal classification.

Table 4.2 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	–

Table 4.3 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

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 4.3.

Table 4.3 shows 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 study) 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%. In contrast, our proposed method reaches 100% classification accuracy for the same pair of data sets. The results demonstrate that our approach can more accurately classify the EEG signals of all epileptic and healthy subjects using the extracted features from the SRS technique.

In Fig. 4.6, the receiving operating characteristic (ROC) curve is drawn 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

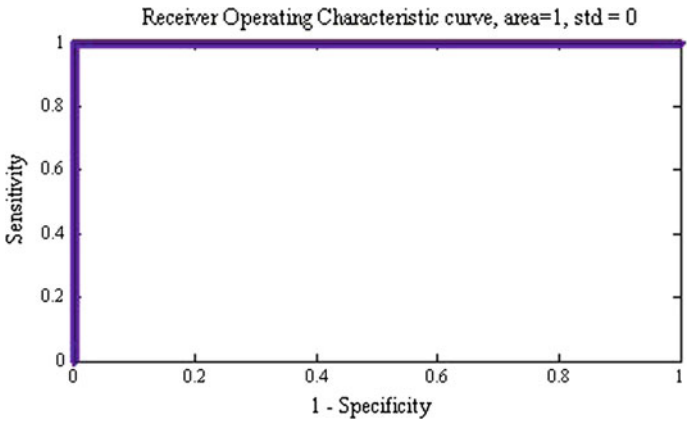


Fig. 4.6 ROC curve for healthy subjects with eyes open (Set A) and epileptic patients during seizure activity (Set E) in the EEG epileptic data

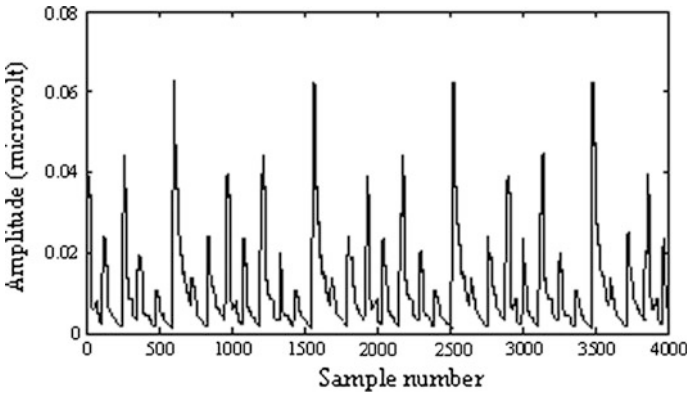


Fig. 4.7 Exemplary mean feature points obtained by the SRS from left hand movements (class 1) of Subject 1

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 Fig. 4.7, 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.

4.3.2 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 4.4. According to the data description (see Sect. 3.3.1.4 in Chap. 3), 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 sessions' recorded EEG data of three kinds of mental imagery tasks for each subject. Thus 10,528 vectors of 96 dimensions are obtained for Subject 1, 10,400 vectors for each of Subjects 2 and 3 with the same dimensions as the original data in the experiment.

In the proposed approach, 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 of 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 Sect. 4.2.2) are then calculated from each sub-sample of a class in a subject. Finally, the LS-SVM algorithm is trained on these features to classify EEG signals.

Figure 4.7 shows the extracted mean feature points of EEG recorded data for the imagination of repetitive self-paced left hand movements (class 1) of Subject 1 from the mental imagery tasks data by the proposed SRS technique. From Fig. 4.7 it can be seen that 4000 feature points out of 43,200 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 set with 960 vectors from each class as the training set. The performance is also evaluated based on this vector set. After the feature

Table 4.4 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

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 $\gamma = 10{,}000$ and $\sigma^2 = 1$. From Table 4.5, it is observed that classification accuracies for different pairs of different subjects are obtained ranging from 93 to 100%.

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 pairs' classification are close to 1. The experimental results therefore, confirm that the features obtained through the SRS technique actually represent the most important information in the recorded EEG data, and that our approach is powerful means of EEG signal classification.

4.3.3 Results for the Two-Class Synthetic Data

In this section, we discuss the experimental results of a non-EEG dataset, which is two-class synthetic data from Ripley (1996). There are two sets of data, training and testing sets (see more in Sect. 3.3.1.5 in Chap. 3). The original Ripley data structure and the feature vectors obtained by the SRS are presented in Table 4.6.

In this experiment, the proposed method is implemented on 250 given training data with two dimensions to find the 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.

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 a random sample size of 943 and a random sub-sample size of 892. The nine features are extracted from each sub-sample. As a result, we get 10

Table 4.5 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	–

Table 4.6 The original two-class synthetic data from Ripley (1996) and the extracted feature vectors obtained 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

Table 4.7 Confusion matrix for Ripley data (1996)

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

Table 4.8 Experimental results for the Ripley data (1996)

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

vectors of 45 dimensions which are used as the test vectors in the experiment. Optimal classification results were achieved when we set $\gamma = 1$ and $\sigma^2 = 1$ in the training and testing.

Classification results of the algorithm are displayed by a confusion matrix in Table 4.7. In the confusion matrix, each cell consists of the number of vectors classified for the corresponding combinations of the predicted and actual outputs. As can be seen in Table 4.7, 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 has occurred using the proposed method.

Table 4.8 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, confirming a perfect classification capability of the approach. The outcomes of this dataset also prove that the proposed method can be successfully used in any classification area. 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.

4.4 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 pairs 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). The datasets are tested 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%. To date and to the best of our knowledge, our results present the highest classification accuracy achieved for that pair of EEG data. To determine the effectiveness of the method on non-EEG data, the proposed algorithm is also applied to the synthetic two-class problem from the Ripley data set (1996). The sensitivity, specificity and classification accuracy rate were found to be a 100% for the values of this 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.

References

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.

Chandaka, S., Chatterjee, A. and Munshi, S. (2009) 'Cross-correlation aided support vector machine classifier for classification of EEG signals', *Expert System with Applications,* Vol. 36, pp. 1329–1336.

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.

De Veaux, R.D., Velleman, P.F., and Bock, D.E., (2008) *Intro Stats,* 3rd ed., Pearson Addison Wesley, Boston, 2008.

EEG time series, 2005, [Online], Viewed 30 September 2008, http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html.

Fletcher, R. (1987) *Practical methods of optimization,* Chichester and New York, John Wiley & Sons.

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.

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.

Iasemidis, L.D., Shiau, D.S., Chaovalitwongse, W., Sackellares, J.C., Pardalos, P.M., Principe, J. C., Carney, P.R., Prasad, A., Veeramani, B., and Tsakalis, K. (2003) 'Adaptive Epileptic Seizure Prediction System', *IEEE Transactions on Biomedical Engineering*, Vol. 50, n. 5, pp. 616–627.

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.

MATLABArsenal-online, http://www.informedia.cs.cmu.edu/yanrong/MATLABArsenal/MATLABArsenal.zip.

Reynolds, E.H. (2000) 'The ILAE[2]/IBE[1]/WHO[3] global campaign against epilepsy: Bringing epilepsy "out of the shadows"', *Epilepsy & Behaviour,* Vol. 1, S3–S8.

Ripley data-online (1996), http://www.stats.ox.ac.uk/pub/PRNN/.

Sample Size Calculator; http://www.surveysystem.com/sscalc.htm.

Siuly, and Li, Y. (2014) 'A novel statistical algorithm for multiclass EEG signal classification', Engineering Applications of Artificial Intelligence, Vol. 34, pp. 154–167.

Siuly, Y. Li, and P. Wen (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.

Siuly, Y. Li, and P. Wen, (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.

Suresh, K., Thomas, S. V., and Suresh, G. (2011) 'Design, data analysis and sampling techniques for clinical research', *Ann Indian Acad Neurol*., Vol. 14, no. 4. pp. 287–290.

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.

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. (2008) 'Wavelet/mixture of experts network structure for EEG signals classification' *Expert System with Applications* Vol. 34, pp. 1954–1962.

Vapnik, V. (1995) *The nature of statistical learning theory,* Springer-Verlag, New York.

WHO (World Health Organization) Report, http://www.who.int/mediacentre/factssheets/fs999/en/index.html (accessed February 2011).

Z Distribution Table, http://ci.columbia.edu/ci/premba_test/c0331/s6/z_distribution.html.

Chapter 5
A Novel Clustering Technique for the Detection of Epileptic Seizures

This chapter presents a different clustering technique for detecting epileptic seizures from EEG signals. This algorithm uses all the data points of every EEG signal. In Chap. 4, we developed and studied the SRS technique which does not use all sample points to make representative features for classification. Sometimes valuable sample points can be omitted during sampling and this 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 aided LS-SVM algorithm named CT-LS-SVM for the detection of epileptic seizures. 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.

One of the most important criteria for choosing the best method is computational efficiency. This study presents a method for two-class EEG signal classification, which reduces the computational complexity, thus reducing execution time. In this chapter, we investigate the performance of the CT-LS-SVM algorithm in the epileptic seizure detection in EEGs with respect to accuracy and execution (running) time of experiments. We also compare the proposed method with the SRS-LS-SVM algorithm. The performance of the proposed method is also compared with existing methods reported in the literature. In addition, the performance of the proposed method is evaluated in two ways; (i) dividing the feature set into two groups as the training and testing sets and (ii) using the tenfold cross validation method.

S. Siuly et al., *EEG Signal Analysis and Classification*,
Health Information Science, DOI 10.1007/978-3-319-47653-7_5

5.1 Motivation

Advanced signal classification techniques for the analysis of EEG signals are essential for developing and understanding the current biomedical research. The classification techniques mainly 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. EEG signals are known to be aperiodic and nonstationary and the magnitudes of signals changed over time. But, during the signal analysis, it is necessary to make an EEG signal stationary. Although recorded EEG signals are not stationary, usually smaller windows, or parts of those signals will exhibit stationarity. An EEG signal is stationary for a small fraction of time. Thus, we intend to introduce the concept of the clustering technique (CT) to have representative values of a specific time period and to make the signals stationary.

In this chapter, the CT approach is proposed for feature extraction from EEG data. In a CT scheme, we partition each recorded EEG channel data into several clusters based on a specific time period to properly account for possible stationarities. In this procedure, each set of EEG channel data is divided into n mutually exclusive clusters with a specific time duration. Each cluster is again partitioned into m sub-clusters over a specific time period, and then some statistical features (discussed in Sect. 5.2.1) are extracted from each sub-cluster to represent the original EEG signals. The same features are used in the SRS-LS-SVM algorithm (Siuly et al. 2011a) in Chap. 4. 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, much less time is taken to run the program. The proposed approach is simple and thus flexible for EEG signal classification. The proposed CT-LS-SVM approach is implemented on the epileptic EEG database (EEG time series 2005; Andrzejak et al. 2001). For further assessment, we also test this method on motor imagery EEG data (data set IVa, BCI Comp III) (BCI Competition III 2005; Blankertz et al. 2006).

5.2 Clustering Technique Based Scheme

In the literature, numerous techniques have been used to obtain representations and to extract features of interest for classification purposes but sometimes they cannot properly manage complex properties of EEG data. Hence, in this chapter, we introduce a new type of clustering technique (CT) for extracting representative

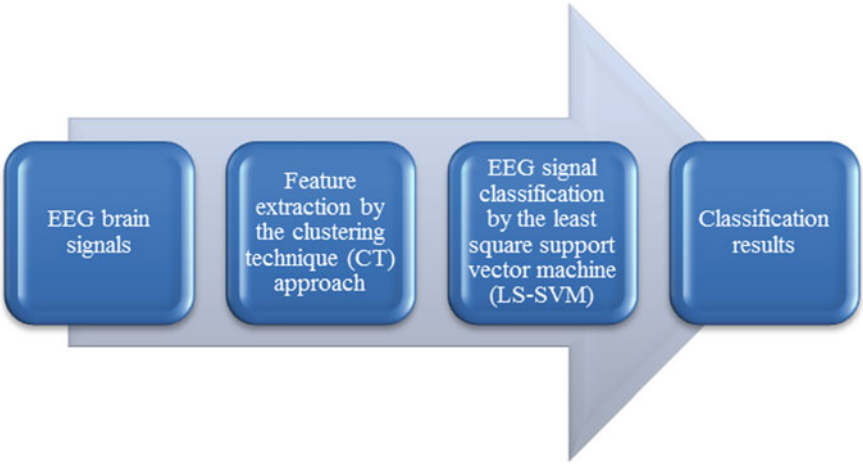


Fig. 5.1 Block diagram of the proposed methodology for EEG signal classification

information from original EEG signals. In this way, we develop a new algorithm based on the CT scheme and LS-SVM named "CT-LS-SVM" algorithm for classifying epileptic EEG signals. The block diagram of the proposed CT method based on the LS-SVM for EEG signals classification is shown in Fig. 5.1. 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 detail.

5.2.1 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 nonstationary, 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 recordings, 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 5.2 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 subsections.

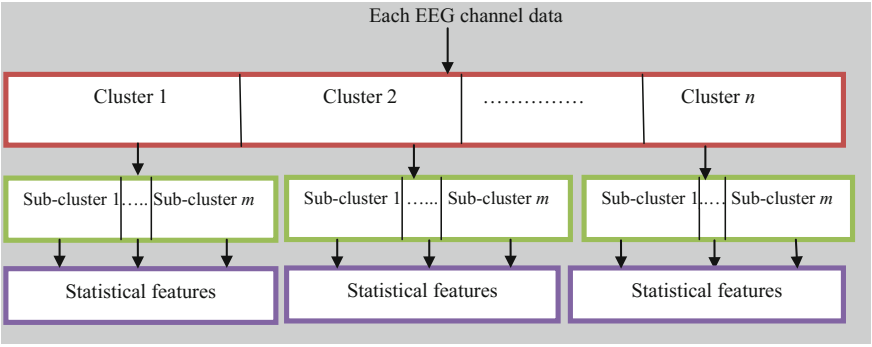


Fig. 5.2 Diagram of the proposed clustering technique

- **Stage 1: Determination of clusters**

In this technique (Siuly et al. 2011b), each EEG channel's data is considered as a population. This population is divided into n groups with a specific time duration, which are called clusters (see Fig. 5.2). Here n is the number of clusters and $n \geq 1$. For any applications, the numbers of clusters (n) are determined empirically over time.

- **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.

- **Stage 3: Statistical feature extraction**

The following nine statistical features of each sub-cluster of each EEG channel's 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 $\{X_{\text{Min}}, X_{\text{Max}}, X_{\text{Mean}}, X_{\text{Me}}, X_{\text{Mo}}, X_{\text{Q1}}, X_{\text{Q3}}, X_{\text{IQR}}, X_{\text{SD}}\}$. The same feature set is used in the previous chapter (Chap. 4) for the SRS-LS-SVM algorithm. The reasons of choosing these statistical features are explained in Chap. 4. The obtained features are employed as the input for the LS-SVM for the EEG signal classification (see the description of LS-SVM in Sect. 4.2.3).

5.3 Implementation of the Proposed CT-LS-SVM Algorithm

First, the proposed method is implemented on the Epileptic EEG data (EEG time series 2005) (see the description of this database in Sect. 3.3.1.1). 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 s. In this method, the 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 s. Then every cluster is again partitioned into 4 sub-clusters and each sub-cluster contains 64 observations of 0.3688 s.

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

For further evaluation, this proposed method is also tested on the motor imagery EEG data (see data description in Sect. 3.3.1.2). As discussed in the data description, the motor imagery EEG data used in this study consists 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 5.1 presents the number of 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

Table 5.1 The number of clusters and sub-clusters and the time period for each cluster and sub-cluster of a class for a subject for the motor imagery EEG data

Subject	Class	Number of clusters	Number of sub-clusters	Number of channels	Time period for each cluster (s)	Time period for each sub-cluster (s)
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

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 feature vector sets are divided into two groups as the training and testing vector sets, and these are used as the input of the LS-SVM algorithm. A detailed description of the LS-SVM is given in Sect. 4.2.3.

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. In the experiments, we utilize three databases into two ways. First, 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 (definitions provided in Chap. 3). Second, the tenfold cross validation method (Abdulkadir 2009; Chap. 3) is used to evaluate the accuracy of the classification method. With the tenfold cross validation method, the whole feature vector set is divided into 10 mutually exclusive subsets of an 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 nine subsets are put together to form a training set. After repeating the method 10 times, the accuracy obtained from each trial is averaged. This is named tenfold cross validation accuracy. The performance is evaluated on the testing set for all the datasets.

5.4 Experimental Results and Discussions

In this study, we investigate the potential of applying the CT algorithm for obtaining representative features from all EEG channel data and these features are used as the inputs to the LS-SVM algorithm. The RBF kernel function is employed for the LS-SVM as an optimal kernel function over the different kernel functions that were tested. The LS-SVM has two important parameters γ and σ^2, which should be appropriately chosen to achieve the desired performance. 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 and the motor imagery EEG data.

In the epileptic EEG data, the optimal classification results are obtained for Case I, Case II, Case III and Case IV (described in Sect. 5.4.1) when $\gamma = 10$ and $\sigma^2 = 4$. Case V achieves the best result when $\gamma = 1$ and $\sigma^2 = 10$ for this database. In the motor imagery EEG data (see Sect. 5.4.2), 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 two datasets are presented in the following sections.

5.4.1 Classification Results for the Epileptic EEG 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 Epileptic EEG 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 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 the two-class of EEG signals.

Five experiments are performed for the Epileptic EEG 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

Table 5.2 Performance comparison of the proposed CT-LS-SVM versus the SRS-LS-SVM method for different pairs of two-class EEG signals from the Epileptic EEG 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

Case IV: Set D versus Set E
Case V: Set A versus Set D

Table 5.2 displays the performance comparison of the proposed CT-LS-SVM method versus the SRS-LS-SVM method for the different cases in the Epileptic EEG 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 5.2, 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.

From Table 5.2, 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 5.3 shows the performance of the proposed method through the tenfold cross validation method for every case from the epileptic data. Using the tenfold cross validation method, we can achieve an overall classification performance of 94.12%. From Table 5.3, it can be observed that the classification accuracies for almost all cases are quite satisfactory, indicating the high performance of the proposed method for EEG signal classification.

Table 5.3 Classification accuracy of the proposed CT-LS-SVM method by the tenfold cross validation for the epileptic EEG-data

Different cases	Tenfold 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 that of the SRS-LS-SVM method for this dataset. Figure 5.3 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 seconds is plotted in the vertical axis. The proposed method takes 8.9 and 9.4 s for 81,940 and 163,880 data samples, respectively, but the SRS-LS-SVM method takes 11.4 and 14.2 s for the same samples, respectively. Figure 5.3 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. Thus it is noted that the proposed method is superior to the SRS method with the LS-SVM in terms of the execution time. Figure 5.3 shows the proposed algorithm results in a shorter execution time, demonstrating the flexibility and the usability of the proposed method.

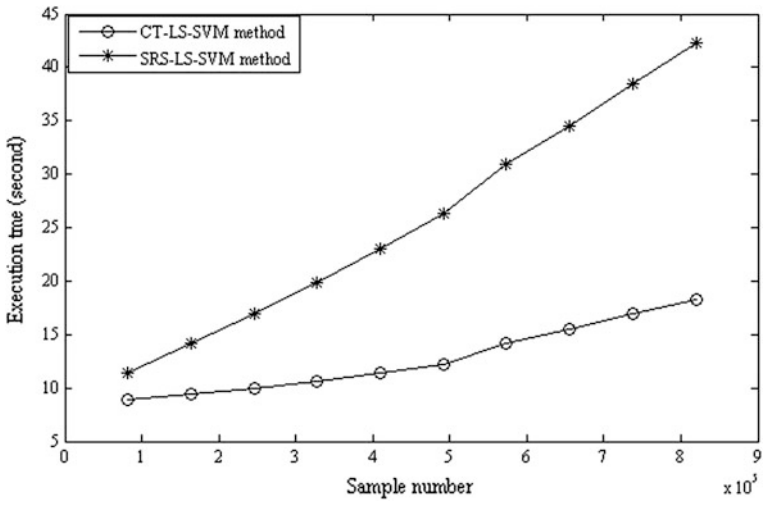


Fig. 5.3 Comparison of the execution time between the CT-LS-SVM and SRS-LS-SVM methods for the epileptic EEG data

Table 5.4 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 epileptic EEG-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. 2006)	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

For a performance comparison, 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 5.4. Compared to the results shown in Table 5.4, our proposed method produces a good classification accuracy rate (99.90%) while the SRS-LS-SVM method (Siuly et al. 2011a) reported 100% classification accuracy. The classification accuracy of the wavelet-artificial neural networks (Guo et al. 2009), wavelet-neural networks (Jahankhani et al. 2006) and expert model with a double-loop EM algorithm (Subasi 2007) were reported at 95.00, 98.00 and 94.50%, respectively. 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 diagnostic decisions. By using such a reliable tool, they can make more accurate medical diagnosing decisions.

5.4.2 Classification Results for the Motor Imagery EEG Data

This section discusses the classification results of the proposed approach for the motor imagery EEG dataset (see Chap. 3). 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

Table 5.5 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

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 5.5 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 5.5, the classification accuracy of the proposed method for almost all the subjects is higher than the previous method, SRS-LS-SVM. 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 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 tenfold cross validation accuracy rate of the proposed approach for the motor imagery data is depicted in Table 5.6. Using tenfold cross validation method, we achieve an overall performance of 88.32% for the same dataset with the same parameters of the classifier. The average classification accuracy is reached at 84.17% for the same feature set when the data are divided into two groups as the training and testing sets (as shown in Table 5.5) using general way. It is worth mentioning that the proposed method is very effective in identifying different motor imagery signals from EEG data.

Table 5.6 Classification accuracy of the CT-LS-SVM method by the tenfold cross validation for the motor imagery EEG data

Subject	Tenfold 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

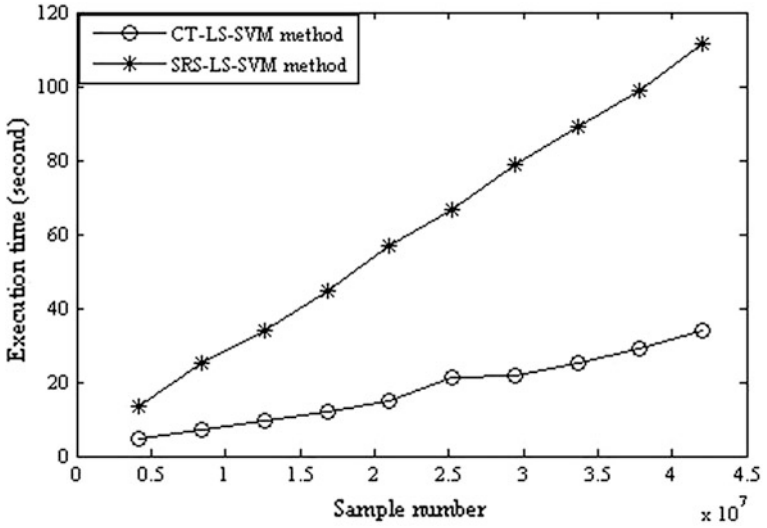


Fig. 5.4 Comparison of the execution time between the CT-LS-SVM and SRS-LS-SVM methods for the motor imagery EEG data

Figure 5.4 presents the experimental time (execution time in seconds) 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 seconds is on the *Y*-axis. Figure 5.4 shows that the proposed method and the SRS technique with LS-SVM take 4.7 and 13.3 s, respectively, when the total number of samples for two-class signals is 4,206,664. Again the program running time is 7.3 s for the proposed CT-LS-SVM algorithm when the algorithm uses 8,413,328 samples whereas the SRS-LS-SVM method takes 25.1 s for the same samples. Thus, Fig. 5.4 shows that, for all samples, the SRS-LS-SVM method takes a longer time than the proposed technique. Figure 5.3 shows the same patterns for the epileptic data. As Fig. 5.4 indicates, the proposed approach is a faster running algorithm compared to the SRS-LS-SVM.

Table 5.7 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 to the classification accuracy for the five subjects and their averages. The average classification accuracies of sparse spatial filter optimization (Yong et al. 2008), Regularized common spatial pattern with generic learning (R-CSP) (Lu et al. 2009), Composite common spatial pattern (composite CSP) (Kang et al. 2009) and Spatially regularized common spatial pattern (SRCSP) methods (Lotte and Guan 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 5.7 Comparison of classification accuracy for the motor imagery EEG data with other EEG signal classification attempts

Method	Classification accuracy (%)					
	*S*1	*S*2	*S*3	*S*4	*S*5	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 and Guan 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
Regularized 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 5.8 shows a summary of the performance of the proposed CT-LS-SVM method versus the SRS-LS-SVM for the three databases. From Table 5.8 it is observed 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. 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. 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 EEG signal classification.

Table 5.8 Summary results of the proposed CT-LS-SVM approach and the SRS-LS-SVM method applied to the epileptic EEG data and the motor imagery EEG data

Database	Average classification accuracy (%)	
	CT-LS-SVM	SRS-LS-SVM
Epileptic EEG data	94.18	95.58
Motor imagery EEG data	84.17	76.71

5.5 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 with less computation complexity. We also investigate whether the CT method is appropriate for feature extraction from EEG data. Experiments are carried out on epileptic EEG data and also motor imagery EEG data. The efficacy and superiority of the proposed CT-LS-SVM method over the SRS-LS-SVM method are validated through different measures. For the Epileptic EEG 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. In the motor imagery data, the average classification accuracy of the proposed approach is 84.17%, while the SRS-LS-SVM has 76.71%. For the CT-LS-SVM algorithm using the tenfold cross validation method, we achieve an overall classification accuracy of 94.12% for the epileptic data and 88.32% for the motor imagery data. In both datasets, the proposed CT-LS-SVM approach took much less time to compute the data compared to the SRS-LS-SVM with two datasets. One of the advantages of the proposed methodology is the computational efficiency and usability.

References

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

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.

BCI competition III, 2005, http://www.bbci.de/competition/iii.

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.

Chandaka, S., Chatterjee, A. and Munshi, S. (2009) 'Cross-correlation aided support vector machine classifier for classification of EEG signals', *Expert System with Applications,* Vol. 36, pp. 1329–1336.

EEG time series, 2005, [Online], http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html.

Guo, L., Rivero, D., Seoane, J.A. and Pazos, A. (2009) 'Classification of EEG signals using relative wavelet energy and artificial neural networks', *GCE,* pp. 12–14.

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).*

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.

Lotte, F. and Guan, C. (2010) 'Spatially regularized common spatial patterns for EEG classification', *Inria-00447435* (25 Jan 2010) version 2.

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.

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.

Siuly, Y. Li, and P. Wen, (2011a) '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, Y. Li, and P. Wen, (2011b) '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.

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.

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.

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.

Chapter 6
A Statistical Framework for Classifying Epileptic Seizure from Multi-category EEG Signals

An innovative idea for classifying epileptic seizures in multi-categories EEG signals is developed in this chapter. Due to the complex characteristics of EEG signals (e.g. poor signal-to-noise ratio, non-stationary, aperiodic), it is hard to achieve the efficient detection of epileptic seizure signs from multi-category EEG signals. Thus, designing efficient detection algorithms that ensure the proper evaluation and treatment of neurological diseases is an important goal for this study. This chapter presents an optimum allocation (OA) technique to select representative samples from every time-window considering the variation of an observation. This research investigates whether the OA scheme is suitable for extracting representative samples from the EEG signals depending on their variability within the groups in the input EEG data. In order to assess performance, the obtained samples are used as the input in the multiclass least square support vector machine (MLS-SVM) for detecting epileptic seizure from multiclass EEG signals.

6.1 Significance of the OA Scheme in the EEG Signals Analysis and Classification

This study presents a novel idea, called "optimum allocation" (OA), to determine the number of observations to be selected from every time-window of each EEG channel data considering minimum variability among the values. Generally, in a random sample section, variability is not considered within a time-window, however, it is an important factor for the provision of information about samples. If the variability within a time-window is large, the size of a sample from that time-window is also large. On the other hand, if the variability of the observations within a time-window is small, the sample size will be small in that time-window.

The OA method is an appropriate choice when a dataset (population) is heterogeneous and very large in size. An effective sample (a subset of a population)

S. Siuly et al., *EEG Signal Analysis and Classification*,
Health Information Science, DOI 10.1007/978-3-319-47653-7_6

represents an appropriate extraction of the useful data which provides a meaningful understanding of the important aspects of the population. Signal processing methods require stationary data even though the overall signal may not be stationary. If the recorded EEG signals in every category are partitioned into several groups based on a specific time period, they may help to properly account for possible stationarities. Thus, the dataset must be divided into several groups to create homogeneity within a group according to its specific characteristics and is used to select representative samples from the groups so that those samples can represent the entire data. When measuring an EEG, a large amount of data with different categories is obtained over a time period. This huge amount of data is very complex and time-consuming to analyse in practise. As EEG recordings normally include a huge amount of data and the data is heterogeneous with respect to a time period, it is reasonable to expect that dividing the whole EEG recording into sub-groups with respect to time, and then taking representative samples from each sub-group would improve the performance of a classifier. Thus, this study introduces the OA scheme for getting representative samples from each group of all channel EEG data to extract reliable information from such a large amount of data.

6.2 Optimum Allocation-Based Framework

In this work, a novel algorithm based on OA and MLS-SVM is proposed for classifying multiclass EEG signals. The method is very effective for labelling sequential data in multi-category EEG identification. Figure 6.1 exhibits the structural diagram of the proposed methodology. The entire process of this method is divided into several processing modules: sample size determination (SSD), epoch determination, OA, sample selection and classification as shown in Fig. 6.1.

First, an appropriate sample size with the desired confidence interval and confidence level for each class from the entire EEG data set is determined by survey software called "*Sample size calculator*". Second, the data of each class is segmented into different epochs, considering a specific time period. Third, the OA technique is employed to determine the best sample size for each epoch with a minimal variance. The sum of all sample sizes for all epochs will be equal to the calculated sample size that is obtained from the class of the entire EEG data (discussed in Sect. 6.2.3). Fourth, the samples are selected from the epochs

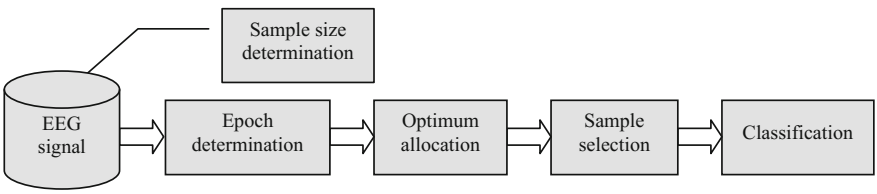


Fig. 6.1 The structural diagram of the proposed approach

considering the size that is obtained by the OA procedure. Finally, these samples obtained from all epochs of each class are used as the input to a classifier, the multiclass LS-SVM (MLS-SVM). The classification results are then acquired. A detailed description of this algorithm is provided in the following sections.

6.2.1 Sample Size Determination

Determining the sample size of a classification method is a crucial component of the study. One of the most important problems in the sample design is to determine how large a sample is needed for the estimates to be reliable enough to meet the objectives of a study. The SSD is a mathematical process for deciding the number of observations or replicates to be included in a statistical sample. The sample size is an important component of an empirical study in which the goal is to obtain results that reflect the original population as precisely as needed. Two factors are necessary to calculate the size of a sample: confidence interval and confidence level (Islam 2007; De Veaux et al. 2008). Statisticians use a confidence interval to express the degree of uncertainty associated with a sample statistic. The confidence level tells us how confident we are. It is expressed as a percentage and represents how often the true percentage of the population lies within the confidence interval. The 99% confidence level means that we are 99% confident.

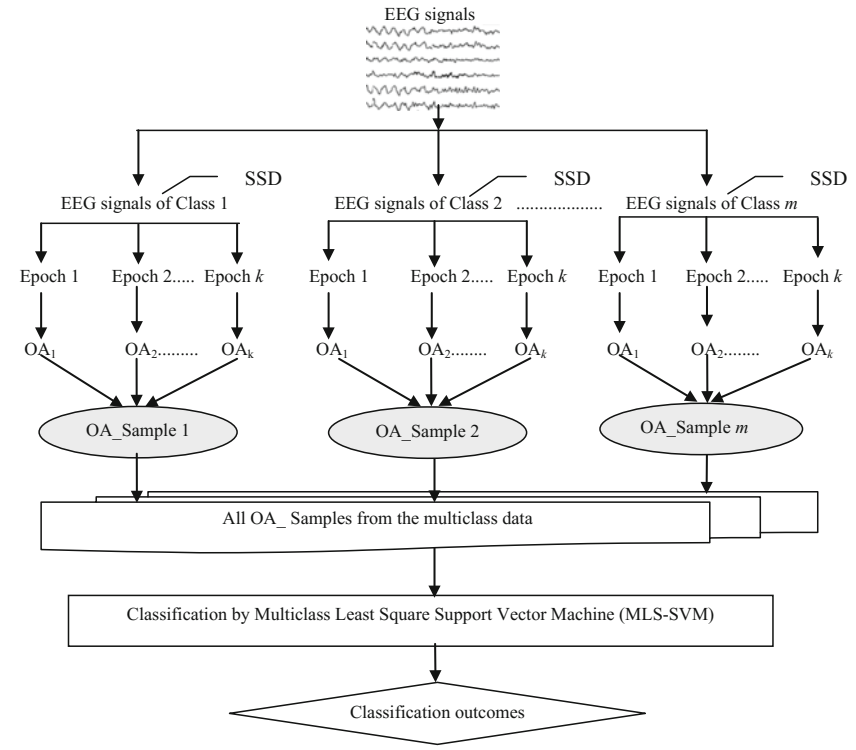
Figure 6.2 presents a design of the proposed algorithm on how a large amount of multiclass EEG data can be processed to obtain reliable information regarding the brain state. As shown in Fig. 6.2, the size of the sample (denoted by n) of each class is determined using a *sample size calculator* with the desired confidence interval and confidence level. The SSD is noted as SSD in Fig. 6.2. This is the first step of the methodology. In the sample size calculator, the following formula is used to calculate the required sample size

$$n = \frac{z^2 \times p \times (1-p)}{e^2}$$

where, n = desired sample size; Z = standard normal variate (the value for Z is found in statistical tables which contain the area under the normal curve) for the desired confidence level (1.96 for 95% confidence level and 2.58 for 99% confidence level) (see Z Distribution Table); p = estimated proportion of an attribute, that is present in the population; e = margin of error or the desired level of precision (e.g. e = 0.01 for 99–100% confidence interval). If the population is finite, the required sample size is given by

$$n_{\text{new}} = \frac{\text{n}}{1 + \frac{n-1}{N}} \tag{6.1}$$

where, N = population size.



Note: SSD= Sample Size Determination; OA= Optimum Allocation; OA_Sample= Sample with the optimum allocated size.

Fig. 6.2 Optimum allocation-based MLS-SVM algorithm for multiclass EEG signal classification

If the estimator p is not known, 0.50 (50%) is used as it produces the largest sample size. The larger the sample size, the more sure we can be that their answers truly reflect the population. In this research, we consider $p = 0.50$ so that the sample size is at a maximum and $Z = 2.58$ and $e = 0.01$ for 99% confidence level. In this study, $N = 4097$ and thus we obtain from Eq. (6.1), $n_{\text{new}} = 3288$ for each and every five classes (Set *Z*, Set *O*, Set *N*, Set *F*, Set *S*) of epileptic EEG data.

6.2.2 Epoch Determination

It is known that EEG signals are aperiodic and non-stationary and the magnitude of the signals change over time. To have representative values of a specific time period, we divide the EEG signals of a class into some mutually exclusive groups, which are called epochs in this research. Thus, each epoch consists of EEG data within a time-window. For example, an epoch of 5.9 s in a class is shown in

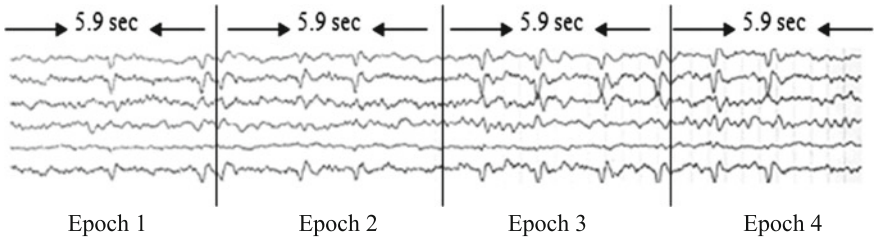


Fig. 6.3 Example of determining epoch from an EEG dataset of a class

Fig. 6.3. For the experiment design, the number of epochs (k) is determined empirically over time. It is worth mentioning that every epoch contains a number of EEG channel data. Usually, the columns of an epoch consist of EEG channel data as shown in an example in Sect. 6.2.3.

6.2.3 Optimum Allocation

OA (Siuly and Li 2014) refers to a method of sample allocation that is designed to provide the most precision. The purpose of the OA consists of determining the number of observations to be selected from different epochs with a view to minimizing their variance. The allocation of the samples to different epochs is governed by two factors: (i) the total number of observations in each epoch; (ii) the variability of observations within the epoch. Of course, the precision of the epoch (Islam 2007) largely depends on the choice of the sample size of the whole data discussed in Sect. 6.2.1.

In this section, our intention is to select a representative sample from each epoch such that the variance is the minimum in each epoch. Suppose, x_{ijl} is the value of the lth unit of the jth channel in the ith epoch in a sample. Here $i = 1, 2,\ldots, k$; $j = 1, 2, \ldots, h$; $l = 1,2,\ldots, n_i$, where n_i is the sample size of the ith epoch. X_{ijl} is the corresponding value in the population where $l = 1,2,\ldots, N_i$. To discover the variability of the mean in the epoch process, we assume that the samples are drawn independently from different epochs and the sample mean is an unbiased estimator of the population mean $\bar{X}$.

The variance of the sample mean $\bar{x}$, $V(\bar{x}) = E[\bar{x} - E(\bar{x})]^2 = E[\bar{x} - \bar{X}]^2$

$$\text{where} \quad \bar{x} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{h}\sum_{l=1}^{n_i} x_{ijl}}{n_1 h + n_2 h + \cdots + n_k h} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{h} n_i \bar{x}_{ij}}{nh}$$

$$\text{and} \quad \bar{X} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{h}\sum_{l=1}^{N_i} X_{ijl}}{n_1 h + n_2 h + \cdots + n_k h} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{h} N_i \bar{X}_{ij}}{nh}$$

$$\Rightarrow E\left[\frac{1}{h}\sum_{i=1}^{k}\sum_{j=1}^{h}\frac{N_i}{N}(\bar{x}_{ij}-\bar{X}_{ij})\right]^2$$

Assuming that the sampling fraction is the same in all epochs, e.g.

$$\begin{aligned}&\frac{n_i}{n}=\frac{N_i}{N}\ \text{where}\, n=n_1+n_2+\cdots+n_k\,\text{and}\, N=N_1+N_2+\cdots+N_K\\&\Rightarrow V(\bar{x})=\frac{1}{h^2}\left[\sum_{i=1}^{k}\sum_{j=1}^{h}\frac{N_i^2}{N^2}E(\bar{x}_{ij}-\bar{X}_{ij})\right]^2=\frac{1}{h^2}\sum_{i=1}^{k}\sum_{j=1}^{h}\frac{N_i^2}{N^2}V(\bar{x}_{ij})\end{aligned}\tag{6.2}$$

Here, $\bar{x}_{ij}$ is the mean of simple random sample in the jth channel of the ith epoch whose variance is $V(\bar{x}_{ij})=\frac{(N_i-n_i)}{N_i}\frac{s_{ij}^2}{n_i}$ by (Cocran 1977)

Putting the value of $V(\bar{x}_{ij})$ into Eq. (6.2), we obtain,

$$V(\bar{x})=\frac{1}{h^2}\sum_{i=1}^{k}\sum_{j=1}^{h}\frac{N_i^2}{N^2}\frac{(N_i-n_i)}{N_i}\frac{s_{ij}^2}{n_i}\tag{6.3}$$

where N_i is the size of the ith epoch; n_i is the required sample taken from the ith epoch;s_{ij}^2 is the standard deviation of the jth channel in the ith epoch; and n is the total sample size in the epoch process.

Now let us see how a given total sample size, n, should be allocated among different epochs so that the estimator, $\bar{x}$, will have the smallest possible variability. Formally the problem is to determine n_1, n_2, ..., n_k for minimizing, $V(\bar{x})$, subject to the constraint that the total size n equals $n=n_1+n_2+\cdots+n_k$. This is equal to minimizing the function of

$$\phi=V(\bar{x})+\lambda\left(\sum_{i=1}^{k}n_i-n\right)=\frac{1}{h^2}\sum_{i=1}^{k}\sum_{j=1}^{h}\frac{N_i^2}{N^2}\frac{(N_i-n_i)}{N_i}\frac{s_{ij}^2}{n_i}+\lambda\left(\sum_{i=1}^{k}n_i-n\right)\tag{6.4}$$

For n_i, λ is an unknown Langrange's multiplier. For the extreme case of the function, we have $\frac{\delta\phi}{\delta n_i}=0$ *and* $\frac{\delta^2\phi}{\delta n_i^2}>0$. By differentiating function ϕ with respect to n_i and equating the derivation to zero, we have,

$$\begin{aligned}&\frac{\delta\phi}{\delta n_i}=-\frac{1}{h^2}\sum_{i=1}^{k}\sum_{j=1}^{h}\frac{N_i^2}{N^2}\frac{(N_i-n_i)}{N_i}\frac{s_{ij}^2}{n_i}+\sum_{i=1}^{k}\lambda=0\\&\Rightarrow n_i=\frac{N_i}{Nh\sqrt{\lambda}}\sqrt{\sum_{j=1}^{h}s_{ij}^2}\end{aligned}\tag{6.5}$$

Summing up both sides of Eq. (6.5), we have $\sqrt{\lambda} = \frac{\sum_{i=1}^{k}\left(N_i\sqrt{\sum_{j=1}^{h} s_{ij}^2}\right)}{hNn}$ and putting the value of $\sqrt{\lambda}$ into Eq. (6.5), we get

$$n_i = \frac{N_i\sqrt{\sum_{j=1}^{h} s_{ij}^2}}{\sum_{i=1}^{k}\left(N_i\sqrt{\sum_{j=1}^{h} s_{ij}^2}\right)} \times n \tag{6.6}$$

Hence, Eq. (6.6) is used to determine the required sample size in the optimal allocation technique. Now the next important task is to decide the number of observations to be taken from each epoch as described in Sect. 6.2.4. From Fig. 6.2, it can be seen that the size of samples from each epoch is determined by the OA denoted as OA. It is important to mention that if the variability within an epoch is large, the size of the samples from that epoch is also large. This means that more observations will be needed in the samples to represent the epoch. On the other hand, if the variability of the observations within an epoch is small, the sample size will be small, meaning that fewer observations will be needed in the sample to represent the epoch.

As shown in Fig. 6.2, the OA process consists of the following steps:

- *Step 1*: Consider all the channels of the EEG data of a class
- *Step 2*: The sample size determination (SSD) is performed from that class through the "*Sample size calculator*". Assume that the selected sample size is n
- *Step 3*: The EEG data of that class is divided into k epochs considering a specific time period. Suppose the sizes of the epochs are N_1, N_2, …, N_k, respectively. Now the problem is to find out how a given total sample size, n, should be allocated among the k epochs, N_1, N_2, …, N_k, with the smallest possible variability
- *Step 4*: Now the sample sizes for each epoch is determined by the OA. Let n_1, n_2, …., n_k be the sizes of samples drawn from the epochs whose sizes are N_1, N_2,…, N_k, respectively, using Eq. (6.6). Then $n_1 + n_2 + \cdots + n_k = n$

6.2.4 Sample Selection

Considering the sizes of samples obtained by the OA using Eq. (6.1), the representative samples are selected from each epoch. The individual samples selected from each epoch in a class make a vector set denoted by OA_Sample as shown in Fig. 6.2. The vector sets of all classes construct a matrix that is used as the input to the MLS-SVM for the classification, as discussed in the next section.

6.2.5 Classification by Multiclass Least Square Support Vector Machine (MLS-SVM)

A MLS-SVM is a straightforward extension of a LS-SVM proposed by Suykens and Vandewalle (1999a). The LS-SVMs was originally designed for binary classification but now it can be effectively extended for multiclass classification. This method is very popular and considered robust in the machine learning community because it nicely deals with high dimensional data and provides good generalization properties. The detailed discussion of the LS-SVM for binary classification is available in references (Suykens et al. 2002; Vapnik 1995; Siuly et al. 2009, 2010, 2011a, b, 2013; Siuly and Li 2014). This study employs the MLS-SVM with radial basis function (RBF) kernel as a classifier to distinguish the different categories or multiclass EEG signals.

Let, $\left\{x_i, y_i^{(j)}\right\}, i = 1, 2, \ldots, N; j = 1, 2, \ldots, m$, be a training set, where N is the number of the training data and m is the number of the classes. Here $x_i = \{x_{i1}, x_{i2}, \ldots, x_{in}\} \in R^n$ is the input index; $y_i^{(j)} = (y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(m)})$ is an m dimensional vector (output index). $y_i^{(j)} \in \{1, -1\}$; $y_i^{(j)} = 1$ means that the ith input vector belongs to the jth class while $y_i^{(j)} = -1$, otherwise. Multiclass categorization problems are typically solved by reformulating the multiclass problem with m classes into a set of L binary classification problems (Allwein et al. 2000; Bishop 1995). The task of an m class classifier is to predict the class label $y_i^{(j)}$ using the input vector $x_i \in R^n$. There exist different approaches to construct the set of binary classifiers and the approaches are minimum output codes (MOC), error-correcting output codes (ECOC), One versus One (1vs1) and One versus All (1vsA).

The MOC approach (Suykens and Vandewalle 1999b) is applied to solve the multiclass problem with binary LS-SVMs, using L bits to encode up to 2^L classes. This output coding having minimal L, is called minimum output coding (MOC). The basis of the ECOC framework is to map a multiclass classification problem into performing a number of multiple binary classification processes. This approach (Dietterich and Bakiri 1995) is motivated by information theory and introduces redundancy ($m < L$) to the output coding to handle misclassifications using a binary classifier. In this approach, up to 2^{m-1-1} (where m is the number of classes) LS-SVMs are trained, each of them aims to separate a different combination of classes (Guler and Ubeyli 2007). The 1vs1 approach, also called pairwise classification, constructs $m(m-1)/2$ binary classifiers for m-class problems, where each binary classifier discriminates two opposing classes. The classifier is trained using all the data from class i as positive instances and all the data from class j as negative instances, disregarding the remaining data. To classify a new instance, x, each of the base classifiers casts a vote for one of the two classes used in its training (Bagheri et al. 2012). Then, 1vs1 method applies the majority voting scheme for labelling $\boldsymbol{x}$ to that class with the most votes. In 1vsA coding approach, each class is discriminated against the rest of the classes. This approach constructs m binary

classification processes for m-class classification problems, where each binary process discriminates a given class from the rest of the $m - 1$ classes (Rifkin and Klautau 2004). For this approach, we require $m = L$ binary classifiers, where the mth classifier is trained with positive examples belonging to class m and negative examples belonging to the other $m - 1$ classes. A new instance is classified in the class whose corresponding classifier output has the largest value (Bagheri et al. 2012).

Detailed descriptions of these four coding systems are available in the literature (Suykens and Vandewalle 1999a; Gestel et al. 2002, 2004). In this study, these four multiclass classification approaches are used to evaluate the reliability of the proposed method. The decision function of the MLS-SVM in Eq. (6.7) is derived directly from solving a set of linear equations (Suykens and Vandewalle 1999b; Sengur 2009; Youping et al. 2005; Suykens et al. 2002). The detailed description of the MLS-SVM algorithm can be found in the literature (Suykens and Vandewalle 1999b; Suykens et al. 2002; Sengur 2009; Youping et al. 2005; Gestel et al. 2002). Finally, the discrimination of the MLS-SVM is obtained as below:

$$y_j(x) = \operatorname{sign}\left(\sum_{i=1}^{N} y_i^{(j)} \alpha_{ij} K_j(x, x_i) + b_j\right), \quad j = 1, 2, \ldots, m \tag{6.7}$$

where $y_j(x)$ is the predicted class on the basis of the input index x, b_j is the bias term, α_{ij} denotes Lagrange multipliers called support values, and $K_j(x, x_i)$ is the RBF kernel defined as $K_j(x, x_i) = \exp\left(-(\|x - x_i\|)^2 / 2\sigma^2\right)$ (Suykens and Vandewalle 1999b; Vapnik 1995).

6.2.6 *Classification Outcomes*

Classification outcomes for the multi-category EEG signals are obtained in this stage. The solution of Eq. (6.7) provides the prediction results that directly assign the samples with various labels to identify which category it belongs to. Based on the outcomes, we can decide how efficiently the MLS-SVM classifier can classify multi-category EEG signals by employing the optimum allocated samples of different classes.

6.3 Implementation of the Proposed Methodology

This section describes how the proposed OA-based MLS-SVM algorithm is implemented on the benchmark epileptic EEG database. As discussed in Sect. 3.3.1.1, the epileptic EEG data has five sets (Set A–Set E), and each set contains 100 channels

data. Every channel consists of 4097 data points of 23.6 s. The different steps of the implementation are provided as below:

First, we calculate the sample size, n, for each class, by using Eq. (6.1). Here, we have $n = 3288$ for each class.
Second, we segment each class into four epochs ($k = 4$) and each epoch contains the data for 5.9 s. As every channel of a class contains 4097 data points of 23.6 s, the maximum sizes of the four epochs are $N_1 = 1024$, $N_2 = 1024$, $N_3 = 1024$, $N_4 = 1025$, respectively.
Third, in order to determine the sample size for each of the four epochs in each class, we apply the OA technique using Eq. (6.6). In this way, the calculated sample size for a class, $n = 3288$ is allocated among the four epochs with the smallest possible variability. As a result, the sum of all the sample sizes from all epochs of a class is equal to the total sample size (n) of that class. As mentioned previously, the sample size of every epoch may not be equal due to the variability of the observations within the epoch. Table 6.1 presents the sample size for each epoch in each of the five classes obtained by using the OA in Eq. (6.6). In Table 6.1, it is observed that the sizes of the samples are not equal in each epoch and the sum of the sample sizes of all the epochs is equal to the previously calculated total sample size of each class, $n = 3288$.
Fourth, the samples are selected from the epochs with the sizes obtained by the OA. The samples drawn from each epoch of a class make a vector set denoted OA_Sample in Fig. 6.2. For example, the selected samples from each of the four epochs of Class 1 create a vector denoted OA_Sample 1. Similarly, the vector sets: OA_Sample 2, OA_Sample 3, OA_Sample 4 and OA_Sample 5, are created from Class 2, 3, 4 and 5, respectively. Together, all the OA_Samples from the five categories EEG data: OA_Sample 1 (from Class 1), OA_Sample 2 (from Class 2), OA_Sample 3 (from Class 3), OA_Sample 4 (from Class 4) and OA_Sample 5

Table 6.1 The sample size for each epoch by the optimum allocation

Class	Epoch 1	Epoch 2	Epoch 3	Epoch 4	Total
Class 1 (Set A)	797 (1024)	822 (1024)	837 (1024)	832 (1025)	3288
Class 2 (Set B)	815 (1024)	840(1024)	805 (1024)	828 (1025)	3288
Class 3 (Set C)	839 (1024)	841 (1024)	780 (1024)	828 (1025)	3288
Class 4 (Set D)	828 (1024)	833 (1024)	788 (1024)	839 (1025)	3288
Class 5 (Set E)	833 (1024)	844 (1024)	815 (1024)	796 (1025)	3288

Note In each cell, the number inside the parentheses is the epoch size (e.g. (1024) in the first cell) and the number outside of the parentheses is the calculated sample size (e.g. 797 in the first cell) obtained by the optimum allocation procedure

(from Class 5), make a matrix that is employed as the input to the MLS-SVMs to identify the five categories of EEG signals.
Finally, the samples selected from each epoch are divided into two distinct sets. One set is used for training purpose, whilst the other set is used for testing the model. Here the training set is used to train the classifier and the testing set is used to evaluate the accuracy and effectiveness of the classifiers for the classification of the multiclass EEG data. In each class, the training set is constructed randomly taking 500 vectors from each epoch while the remaining vectors of each epoch are considered as the testing set. For example, in Class 1, the selected sample sizes are 797, 822, 837 and 832 for Epoch 1, 2, 3 and 4, respectively, as shown in Table 6.1. For the training set, we consider 500 observations from each epoch and the remaining 297, 322, 337 and 332 observations are taken from Epoch 1, 2, 3 and 4, respectively, for the testing set. Thus, for each of the five classes, we obtain a total of 2000 observations for the training set and 1288 observations for the testing set. Finally, for the five-class ($m = 5$) EEG data set, we acquire a total of 10,000 observations for the training set and 6440 observations for the testing set. Note that each vector is 100 dimensions here. In each classification system, this training set is fed into the MLS-SVM classifier as the input to train the classifier and the performances are assessed with the testing set.

In this research, the stability of the performance of the proposed method is assessed based on different statistical measures, such as sensitivity, specificity and total classification accuracy (TCA) and their formulas are given below (Siuly et al. 2012; Guler and Ubeyli 2007; Ubeyli 2009):

$$\bullet\ \text{Sensitivity}(i) = \frac{\text{No. of true positive decisions in class } i}{\text{No. of actual positive cases in class } i} \quad i = 1, 2, 3, 4, 5 \tag{6.8}$$

$$\bullet\ \text{Specificity}(i) = \frac{\text{No. of true negative decisions in class } i}{\text{No. of actual negative cases in class } i} \quad i = 1, 2, 3, 4, 5 \tag{6.9}$$

$$\bullet\ \text{TCA} = \frac{\text{No. of total correct decisions in all classes}}{\text{Total no. of actual cases in all classes}} \tag{6.10}$$

In this study, the calculated sensitivity for each class is called “class-specific sensitivity” and the calculated specificity from each class is called “class-specific specificity”.

6.4 Results and Discussions

Experiments are conducted to determine the capability and consistency of the proposed algorithm. This section presents the experimental outcomes of the proposed method for each of the four multiclass LS-SVM classification approaches: MOC, ECOC, 1vs1, and 1vsA. The classification by the MLS-SVM is carried out in MATLAB (version 7.14, R2012a) using the LS-SVMlab toolbox (version 1.8) (LS-SVMlab toolbox-online) and all other mathematical calculations are also performed in MATLAB with the same version. Before applying the classification algorithm, the parameters of the MLS-SVM is chosen carefully over an empirical evaluation as the classification performance depends on the parameters. In order to achieve the most reliable and consistent results, the proposed method is repeated 10 times on the same parameters for each of the four classification approaches. This section also provides a comparison between the proposed method and four other well-known existing methods.

6.4.1 Selection of the Best Possible Combinations of the Parameters for the MLS-SVM

In this chapter, the RBF kernel function is employed for the MLS-SVM as an optimal kernel function over different kernel functions that were tested. The MLS-SVM has two important parameters, γ and σ^2, which should be appropriately chosen for achieving the desired performance as these parameters play a significant role in the classification performance. The regularization parameter γ determines a trade-off between minimizing the training error and minimizing the model complexity. The parameter σ^2 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. 2009; Suykens et al. 2002), so the values must be chosen carefully. As discussed before, there are four types of output coding approaches of the LS-SVM for the multiclass problem, which are MOC, ECOC, 1vs1 and 1vsA. This study conducts several experiments for each of the four output coding approaches in different combinations of those parameters to discover the optimum values of the hyper-parameters. To find the optimal combination, we have opted for a set of values of the hyper-parameters, γ and σ^2 = 1, 10, 100, 1000 for possible combinations. We obtain sixteen combinations of γ and σ^2 as shown in Table 6.2,, which are employed in the four multiclass classification approaches. Thus, the 16 experiments for each of the 4 classification schemes are carried out on the epileptic EEG data and the outcomes of those experiments are reported in Table 6.2.

Table 6.2 presents the TCA in percentage for the five class experiment datasets in each of the four output coding approaches. They are calculated using Eq. (6.10). In order to provide a clear outlook, Fig. 6.4a–d are presented to show the patterns of the TCA for each individual value of σ^2 = 1, 10, 100, 1000 against γ = 1, 10, 100, 1000 in MOC, ECOC, 1vs1 and 1vsA, respectively. From Table 6.2 and Fig. 6.4a–d, it is observed that the classification performances are not good enough for σ^2 = 1 with any values of γ in each of the four classification approaches (MOC, ECOC, 1vs1 and 1vsA). It is also seen that, when σ^2 = 10, the TCA is dramatically increased over the TCA when σ^2 = 1. The TCA is much better for σ^2 = 100 compared to σ^2 = 10. The experimental outcomes presented in Table 6.2 show that the most favourable results are obtained when γ = 10 and σ^2 = 100 which is 100% for all four classification approaches. On the other hand, when σ^2 = 1000, the TCA decreases compared to σ^2 = 100 and σ^2 = 10. In this study, the TCA is used as a performance indicator because it measures what percentage of a classification method can be provided and how well classification is made in the process.

Considering the performances depicted in Table 6.2 and Fig. 6.4a–d, it is observed that γ = 10 and σ^2 = 100 is the best combination for the MLS-SVM method for the four output coding approaches and this combination is used as the optimal combination in this research.

Table 6.2 The total classification accuracy (in percentage) for different combinations of the parameters (γ, σ^2)

Different combinations of parameters (γ, σ^2)	Total classification accuracy (TCA) (%)			
	MOC	ECOC	1vs1	1vsA
γ = 1, σ^2 = 1	47.22	40.19	80.31	16.61
γ = 1, σ^2 = 10	99.92	99.89	99.91	99.89
γ = 1, σ^2 = 100	99.95	99.91	100	99.98
γ = 1, σ^2 = 1000	95.14	76.46	99.97	95.08
γ = 10, σ^2 = 1	54.11	50.25	81.66	36.32
γ = 10, σ^2 = 10	99.95	99.97	99.97	99.95
γ = 10, σ^2 = 100	100.0	100.0	100.0	100.0
γ = 10, σ^2 = 1000	97.73	96.71	100	98.11
γ = 100, σ^2 = 1	55.16	51.74	81.99	38.79
γ = 100, σ^2 = 10	99.97	99.97	99.97	99.97
γ = 100, σ^2 = 100	99.97	99.97	100	99.97
γ = 100, σ^2 = 1000	99.22	99.15	100	99.08
γ = 1000, σ^2 = 1	55.36	51.88	82.05	38.98
γ = 1000, σ^2 = 10	99.97	99.97	99.97	99.97
γ = 1000, σ^2 = 100	99.84	99.72	100	99.8
γ = 1000, σ^2 = 1000	98.93	98.34	100	98.51

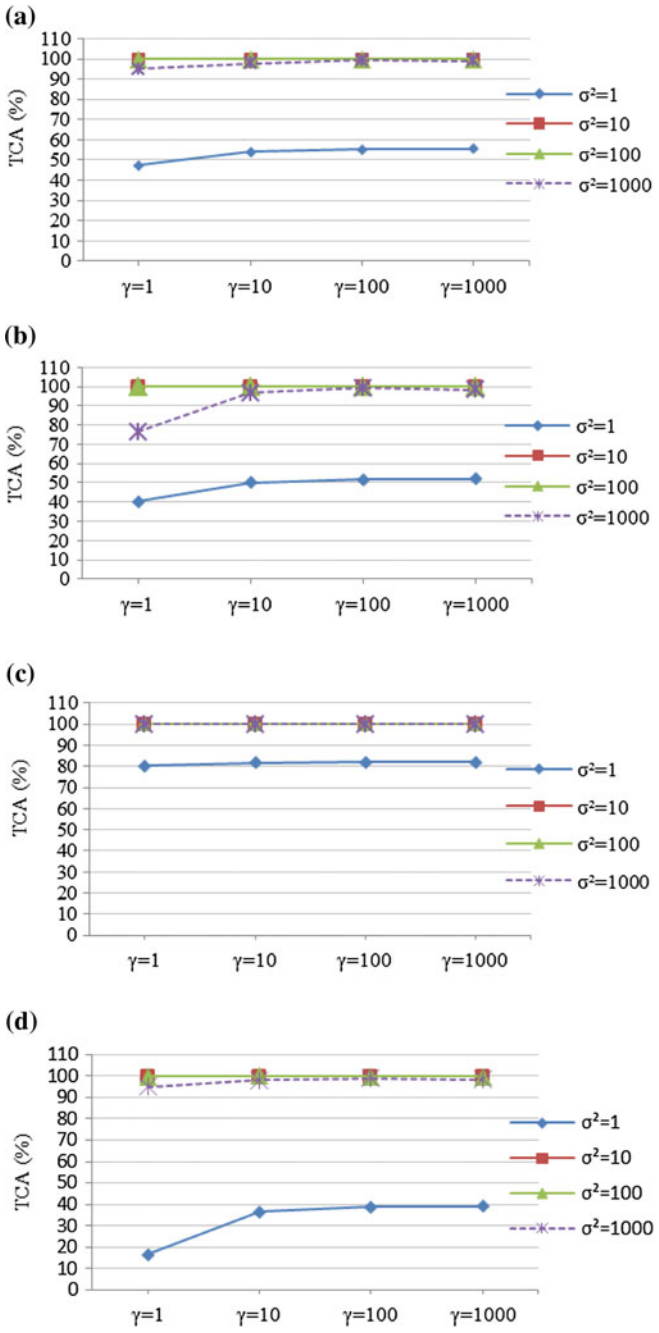


Fig. 6.4 The total classification accuracy (TCA) performance for different combinations of parameters (γ, σ^2) for four multiclass classification approaches: **a** MOC approach; **b** ECOC approach; **c** 1vs1 approach; and **d** 1vsA approach

6.4.2 Experimental Classification Outcomes

6.4.2.1 Results for the Minimum Output Codes (MOC) Approach

Table 6.3 displays the experimental results for the MOC approach for 10 repetitions using the parameter values, $\gamma = 10$ and $\sigma^2 = 100$. From Table 6.3, one can observe that, the class-specific performances in terms of sensitivity and specificity are very promising: almost 100% in each of the 10 repetitions. There are no significant differences in the performances (class-specific sensitivity and class-specific specificity) among the 10 runs, indicating the consistency of the proposed algorithm.

In order to identify the patterns of the class-specific sensitivity and class-specific specificity for each class (set), we provide an individual graphical presentation for the MOC approach. Figure 6.5 shows the patterns of the class-specific sensitivity performances for each class, and Fig. 6.6 exhibits the patterns of the class-specific specificity performances for each class. The bars in these graphs represent the standard errors. From Fig. 6.5, it is noted that the class-specific sensitivity patterns for Set A, Set B, Set C and Set D are the same but the pattern for Set E is a little bit different from the other sets. From Fig. 6.6, it is seen that Set B, Set C, Set D and Set E follow the same patterns in the class-specific specificity while Set A is slightly different.

Table 6.4 summarizes all the results for the MOC approach. In this table, the sensitivity refers to the average of all class-specific sensitivities for a class in all 10 repetitions. In the meantime, the same patterns occur for the specificity (also considered in Tables 6.6 and 6.8). As shown in Table 6.4, the sensitivity, specificity and the TCA are very high for the proposed algorithm with the MOC approach and the overall TCA is 99.99%.

6.4.2.2 Results for the Error-Correcting Output Codes (ECOC)

Table 6.5 shows the experimental results for the ECOC approach with 10 repetitions. From the table, one can see that, in each repetition the class-specific performances (sensitivity and specificity) are very similar proving the consistency of the algorithm.

Figures 6.7 and 6.8 illustrate the class-specific sensitivity and the class-specific specificity performances, respectively, for the five classes (five sets) applying the ECOC process. As shown in Fig. 6.8, one can observe that, in each repetition the patterns of the performances (sensitivity and specificity) for Sets A, B, C and D are similar, but Set E is a bit lower than the others.

The summary results of the ECOC approach is reported in Table 6.6. The results show that the sensitivity, specificity and TCA are very high: close to 100%. From these results, it is certain to conclude that the proposed method with the ECOC yields an excellent performance.

Table 6.3 Experimental classification results for the MOC approach for 10 repetitions

Repeat	EEG data sets	Class-specific performance (%)	
		Class-specific sensitivity	Class-specific specificity
1	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
2	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
3	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
4	Set A	100	99.98
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	99.92	100
5	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
6	Set A	100	99.98
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	99.92	100
7	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
8	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100

(continued)

Table 6.3 (continued)

Repeat	EEG data sets	Class-specific performance (%)	
		Class-specific sensitivity	Class-specific specificity
9	Set A	100	99.98
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	99.92	100
10	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100

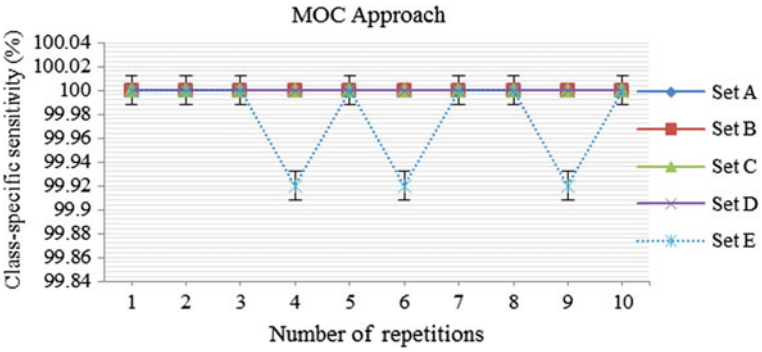


Fig. 6.5 Class-specific sensitivity patterns for each class with the MOC approach

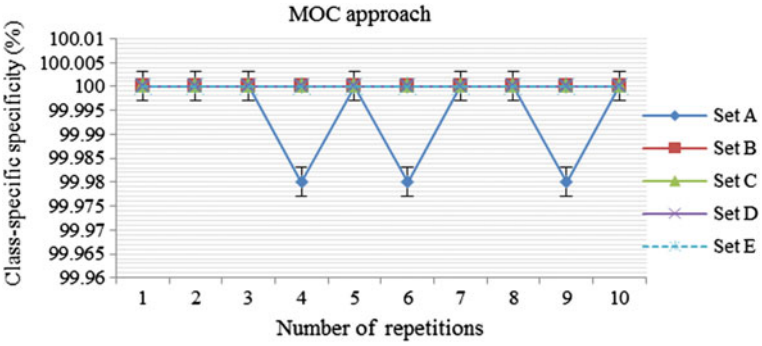


Fig. 6.6 Class-specific specificity patterns for each class with the MOC approach

Table 6.4 Performance summary for the MOC approach

Data sets	Overall performance (%)		
	Sensitivity	Specificity	Total classification accuracy (TCA)
Set A	100	99.99	99.99
Set B	100	100	
Set C	100	100	
Set D	100	100	
Set E	99.98	100	

6.4.2.3 Results for the One Versus One (1vs1) Approach

Figures 6.9 and 6.10 display the experimental results of the proposed OA-based MLS-SVM algorithm by the 1vs1 approach for 10 repetitions. The results reveal a remarkable rate of 100% for both the class-specific sensitivity and the class-specific specificity in each repetition for all five datasets. This is really an amazing outcome produced by the 1vs1 approach. Based on the results, it can be concluded that the proposed algorithm is very consistent.

As shown in Fig. 6.9, it is obvious that the class-specific sensitivity patterns are the same for each class and for each of the 10 repetitions, which is 100%. Figure 6.10 also illustrates the same patterns of the class-specific specificity for every class. In Figs. 6.9 and 6.10, the overall performance for each performance indicator (sensitivity and specificity) is 100% for the 1vs1 approach. Thus, it is clear that the overall TCA is 100% in the 1vs1 approach. It is, therefore, confident to conclude that the method with the 1vs1 approach has a high capability for the EEG classification.

6.4.2.4 Results for the One Versus All (1vsA) Approach

The experimental results of the proposed algorithm for the 1vsA are presented in Table 6.7. This table provides the repeated results for each of the 10 repetitions in terms of the class-specific sensitivity and class-specific specificity. As shown in Table 6.7, it is observed that there are no significant variations in the performances for the 10 repetitions. In most of the cases, the classification performances are 100% or close to 100%, which confirms the reliability of the proposed algorithm.

For the 1vsA approach, Figs. 6.11 and 6.12 provide a view of the whole spectrum of the class-specific sensitivity and the class-specific specificity, respectively. From Fig. 6.11, it is observed that, in each repetition, sensitivity patterns are similar for Sets A, B, C and D, but Set E is different. Figure 6.12 shows the same performance patterns for Sets B, C, D and E, but Set A is a little bit far from the others.

Table 6.5 Experimental results for the ECOC approach with 10 repetitions

Repeat	EEG data sets	Class-specific performance (%)	
		Class-specific sensitivity	Class-specific specificity
1	Set A	100	99.98
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	99.92	100
2	Set A	100	99.96
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	99.98	100
3	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
4	Set A	100	99.98
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	99.92	100
5	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
6	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
7	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
8	Set A	100	99.94
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	99.77	100

(continued)

Table 6.5 (continued)

Repeat	EEG data sets	Class-specific performance (%)	
		Class-specific sensitivity	Class-specific specificity
9	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
10	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100

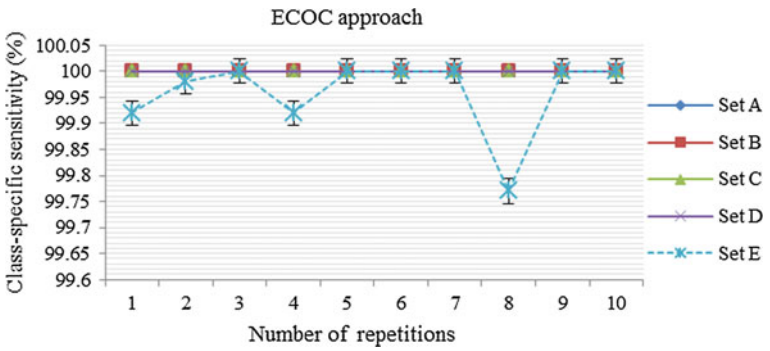


Fig. 6.7 Class-specific sensitivity patterns for each class with the ECOC approach

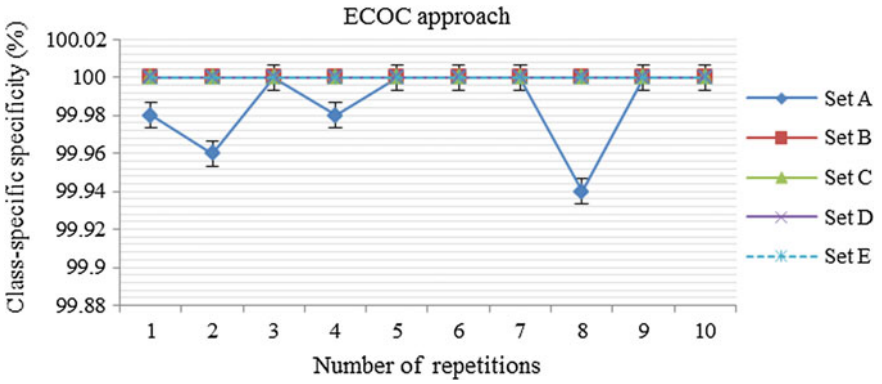


Fig. 6.8 Class-specific specificity patterns for each class with the ECOC approach

Table 6.6 Performance summary by the ECOC approach

Data sets	Overall performance (%)		
	Sensitivity	Specificity	Total classification accuracy (TCA)
Set A	100	99.986	99.99
Set B	100	100	
Set C	100	100	
Set D	100	100	
Set E	99.959	100	

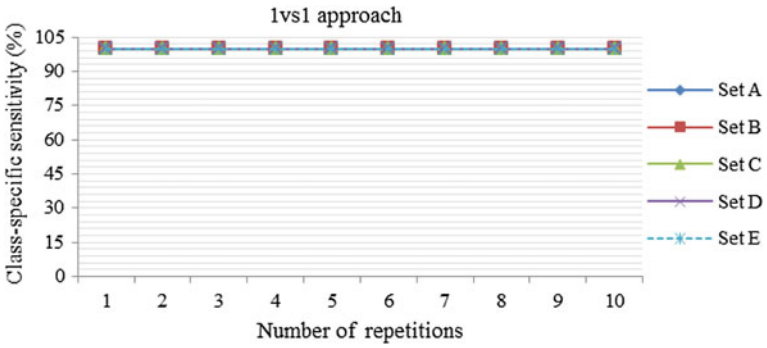


Fig. 6.9 Class-specific sensitivity patterns for each class with the 1vs1 approach

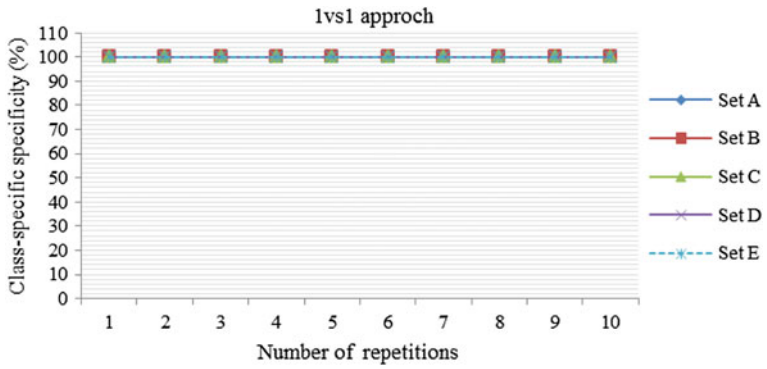


Fig. 6.10 Class-specific specificity patterns for each class with the 1vs1 approach

A performance summary of the proposed algorithm by the 1vsA approach is provided in Table 6.8. The results presented in this table demonstrate that the OA-based MLS-SVM with the 1vsA yielding a good classification performance for each class that is close to 100% in most of the cases.

Table 6.7 Experimental results for the 10 repetitions by the 1vsA approach

Repeat	EEG data sets	Class-specific performance (%)	
		Class-specific sensitivity	Class-specific specificity
1	Set A	100	99.98
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	99.92	100
2	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
3	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
4	Set A	100	99.98
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	99.92	100
5	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
6	Set A	100	99.98
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	99.92	100
7	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100
8	Set A	100	99.98
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	99.92	100

(continued)

Table 6.7 (continued)

Repeat	EEG data sets	Class-specific performance (%)	
		Class-specific sensitivity	Class-specific specificity
9	Set A	100	99.96
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	99.84	100
10	Set A	100	100
	Set B	100	100
	Set C	100	100
	Set D	100	100
	Set E	100	100

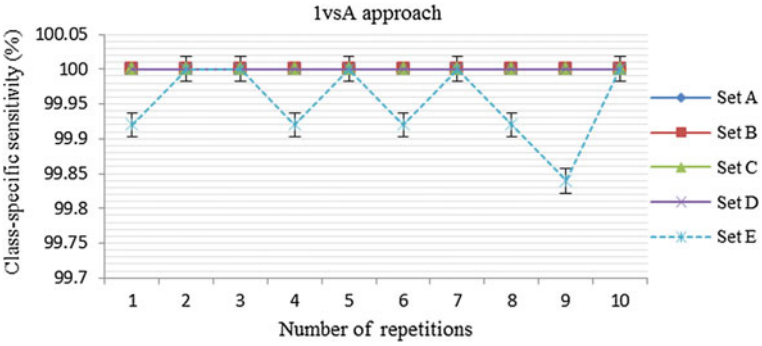


Fig. 6.11 Class-specific sensitivity patterns for each class by the 1vsA approach

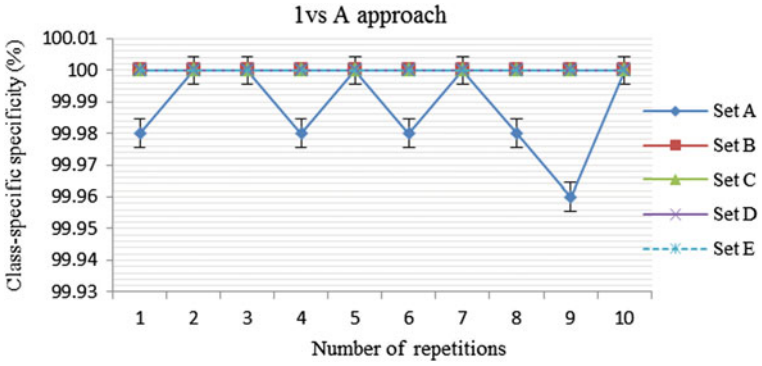


Fig. 6.12 Class-specific specificity patterns for each class by the 1vsA approach

Table 6.8 Performance summary by the 1vsA approach

Data sets	Overall performance (%)		
	Sensitivity	Specificity	Total classification accuracy (TCA)
Set A	100	99.9875	99.99
Set B	100	100	
Set C	100	100	
Set D	100	100	
Set E	99.952	100	

6.5 Comparison

The literature contains a few studies in the literature (Murugavel et al. 2011; Ubeyli 2008, 2009; Guler and Ubeyli 2007) that performed the multiclass EEG signal classification, and dealt with the same epileptic EEG data. In order to further examine the efficiency of our proposed OA-based MLS-SVM algorithm, this section provides a comparison for our proposed approach with four well-known reported algorithms. The three reference methods by Ubeyli (2008, 2009) and Guler and Ubeyli (2007) were employed with the ECOC approach and one method by Murugavel et al. (2011) was used with the 1vs1 approach for the multiclass EEG signal classification. Hence, we compare the proposed algorithm for the ECOC approach with the three reference methods in Table 6.9 and also compare our method for the 1vsA in Table 6.10. We could not present the comparison results for the MOC and 1vs1 as there are no reported research results available for the epileptic EEG dataset.

Table 6.9 presents a comparative study for the ECOC approach between our proposed method and the three reference algorithms for the epileptic EEG dataset. This table reports the overall classification performances of the five categories EEG signals in terms of sensitivity, specificity and the TCA. For each method, the highest classification performances among the four algorithms are highlighted in **bold** font. From Table 6.9, it is clear that our proposed algorithm yields the highest performance in each statistical parameter in each category compared to the three reference methods.

For example, the TCA of the proposed method is 99.99% while 99.20, 99.30 and 99.28%, reported by Ubeyli (2009, 2008) and Guler and Ubeyli (2007), respectively. Thus, our algorithm improves accuracy by at least 0.79% compared to the reference methods.

Table 6.10 shows a comparison of the performances for the 1vsA approach between the proposed algorithm and the reference method by Murugavel et al. (2011). In this table, one can see that, for each class, the proposed method produces the best performance, which is 100% while the reference method generates only 96%. Based on the performances presented in Tables 6.9, it can be concluded that our proposed method is the best in the epileptic EEG classification: achieving a promising classification result.

Table 6.9 Performance comparison by the ECOC approach

Methods	Data sets	Statistical parameters (%)		
		Sensitivity	Specificity	Total classification accuracy (TCA)
Proposed algorithm	Set A	**100**	**99.99**	**99.99**
	Set B	**100**	**100**	
	Set C	**100**	**100**	
	Set D	**100**	**100**	
	Set E	**99.96**	**100**	
Ubeyli (2009)	Set A	99.25	99.84	99.20
	Set B	99.13	99.81	
	Set C	99.25	99.72	
	Set D	99.38	99.62	
	Set E	99.00	100.00	
Ubeyli (2008)	Set A	99.38	99.81	99.30
	Set B	99.25	99.87	
	Set C	99.13	99.78	
	Set D	99.50	99.65	
	Set E	99.25	100.00	
Guler and Ubeyli (2007)	Set A	99.25	99.84	99.28
	Set B	99.38	99.84	
	Set C	99.25	99.75	
	Set D	99.38	99.65	
	Set E	99.13	100.00	

Table 6.10 Performance comparison by the 1vsA approach

Methods	Total classification accuracy (TCA) (%)
Proposed algorithm	99.99
Murugavel et al. (2011)	96.00

6.6 Concluding Remarks

This chapter presents a new classification structure to classify multiclass EEG signals using the OA and the MLS-SVM. The OA technique is employed to select representative samples as EEG signal features, considering the variability of the observations within the groups called "epoch". In this research, the MLS-SVM with RBF is used as the classifier where the extracted samples are fed to identify several categories of EEG signals. Four output coding approaches: MOC, ECOC, 1vs1 and 1vsA are applied in the MLS-SVM and their individual effectiveness is investigated. Before the classification procedure, the parameter values ($\gamma = 10$ and $\sigma^2 = 100$) of the MLS-SVM method are determined after an extensive experimental evaluation. To examine the consistency of the method, the experiments of the

proposed algorithm are repeated 10 times for each of the four output coding approaches with the selected optimal parameter values. For further performance evaluation, the proposed algorithm is compared with the four well-known existing methods: Murugavel et al. (2011), Ubeyli (2008, 2009) and Guler and Ubeyli (2007). The experimental results show that our developed algorithm is consistent in each repetition, and yields very high classification performances for each of the four output coding approaches. There are no significant differences among the performances by the MOC, ECOC, 1vs1 and 1vsA approaches. The results also demonstrate that our method is superior in comparison to the existing methods for the same epileptic EEG database. This research leads us to confirm that the OA reliably captures valuable information from the original EEG data and the MLS-SVM is very promising for the classification of multiclass EEG signals.

References

Allwein, E. L., Schapire, R. E., and Singer, Y. (2000) 'Reducing multiclass to binary: A unifying approach for margin classifiers', *Journal of Machine Learning Research,* Vol. 1, 113–141.

Bagheri M. A., Montazer G. A. and Sergio Escalera S. (2012), 'Error Correcting Output Codes for multiclass classification, Application to two image vision problems', *The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012)*, pp 508–513.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.

Cochran, W. G. (1977) *Sampling Techniques*, Wiley, New York.

Chandaka, S., Chatterjee A., and Munshi, S. (2009) 'Cross-correlation aided support vector machine classifier for classification of EEG signals', *Expert System with Applications*, Vol. 36, 1329–1336.

Dietterich, T. G., and Bakiri, G. (1995) 'Solving multiclass learning problems via error-correcting output code', *Journal of Artificial Intelligence Research,* Vol. 2, 263–286.

De Veaux, R. D., Velleman, P.F. and Bock, D.E. (2008) Intro Stats (3rd edition), Pearson Addison Wesley, Boston.

EEG time series, (epileptic EEG data) (2005, Nov.) [Online], http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html.

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, 117–126.

Gestel, T.V., Suykens, J.A.K., Lanckriet, G., Lambrechts, A., De Moor, B. and Vandewalle, J. (2002) ' Multiclass LS-SVMs: Moderated Outputs and Coding-Decoding Schemes', *Neural processing letters*, Vol. 15, 45–58.

Gestel, T.V., Suykens, J.A.K., Baesens B., Viaene S., Vanthienen, J., Dedene, G., De Moor, B. and Vandewalle, J. (2004) 'Benchmarking Least Square Support Vector Machine classifiers ', *Machine Learning'*, Vol. 54, 5–32.

Islam, M. N. (2007) *An introduction to sampling methods: theory and applications*, revised ed., Book World, Dhaka New Market and P.K. Roy road, Bangla Bazar, Dhaka-1100.

LS-SVMlab toolbox (version 1.8), http://www.esat.kuleuven.ac.be/sista/lssvmlab/.

Murugavel A.S.M, Ramakrishnan S., Balasamy K. and Gopalakrishnan T. (2011) 'Lyapunov features based EEG signal classification by multi-class SVM' *2011 World Congress on Information and Communication Technologies,* 197–201.

Rifkin R. and Klautau A. (2004) 'Parallel networks that learn to pro-nounce english text', *Journal of Machine Learning Research*, pages 101–141.

Sample size calculator, http://www.surveysystem.com/sscalc.htm.

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

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) '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. (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, Vol. 20, no. 4, 526–538.

Siuly, Li, Y. and Wen, P. (2013) 'Identification of Motor Imagery Tasks through CC-LR Algorithm in Brain Computer Interface', *International Journal of Bioinformatics Research and Applications,* Vol. 9, no. 2, pp. 156–172.

Siuly and Y. Li, (2014), 'A novel statistical framework for multiclass EEG signal classification', *Engineering Applications of Artificial Intelligence*, Vol. 34, pp. 154–167.

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. (1999a) 'Least Square Support Vector Machine classifier', *Neural Processing Letters*, Vol. 9, no. 3, 293–300.

Suykens, J. A. K. and Vandewalle, J. (1999b) 'Multiclass Least Squares Support Vector Machines', *In: Proc. International Joint Conference on Neural Networks (IJCNN'99)*, Washington DC, 1999.

Ubeyli, E. D. (2009) 'Decision support systems for time-varying biomedical signals: EEG signals classification', *Expert Systems with Applications,* Vol. 36, 2275–2284.

Ubeyli, E. D. (2008) 'Analysis of EEG signals by combining eigenvector methods and multiclass support vector machines', *Computer in Biology and Medicine,* Vol. 38, 14–22.

Vapnik, V. (1995) *The nature of statistical learning theory,* Springer-Verlag, New York.

Youping F., Yungping C., Wansheng S. And Yu L. (2005) 'Multi-classification algorithm and its realization based on least square support vector machine algorithm', *Journal of Systems Engineering and Electronics,* Vol. 16, no. 4, 901–907.

Chapter 7
Injecting Principal Component Analysis with the OA Scheme in the Epileptic EEG Signal Classification

This chapter presents a different design for reliable feature extraction for the classification of epileptic seizures from multiclass EEG signals. In this chapter, we introduce a principal component analysis (PCA) method with the optimum allocation (OA) scheme, named as OA_PCA for extracting reliable characteristics from EEG signals. As EEG data from different channels are correlated and huge in number, the OA scheme is used to discover the most favourable representatives with minimal variability from a large number of EEG data, and the PCA is applied to construct uncorrelated components and also to reduce the dimensionality of the OA samples for enhanced recognition. To discover a suitable classifier for the OA_PCA feature set, four popular machine learning classifiers, such as LS-SVM, naive Bayes classifier (NB), *k*-nearest neighbor algorithm (KNN) and linear discriminant analysis (LDA) are applied and tested. Furthermore, our approaches are compared with some recent reported research work. This proposed method can also be applicable in biomedical signal processing and also in time series data analysis for acquiring evocative features when data size is large.

7.1 Background

Finding traces of epilepsy through the visual marking of long EEG recordings by human experts is a very tedious, time-consuming and high-cost task (Sharma and Pachori 2015). Furthermore, due to visual analog analysis, different experts could give contradictory diagnosis for the same EEG segment. Therefore, current biomedical research is challenged with the task of determining how to classify time-varying epileptic EEG signals as accurately as possible (Bronzino 2000; Siuly et al. 2011). To classify signals, the first and most important task is to extract distinguishing features or characteristics from the epileptic EEG data. These features determine detection and classification accuracy. The features characterizing the original EEGs are used as the input of a classifier to differentiate normal and

S. Siuly et al., *EEG Signal Analysis and Classification*,
Health Information Science, DOI 10.1007/978-3-319-47653-7_7

epileptic EEGs. As optimal features play a very important role in the performance of a classifier, this study intends to identify a robust feature extraction method for the recognition of epileptic EEG signals.

In the last few years, the literature has discussed many methods of feature extraction and classifications in EEG signal analysis. Acharya et al. (2013) discussed various feature extraction methods and the results of the different automated epilepsy stage detection techniques in detail. In that study, the authors briefly discussed some challenges to developing a computer aided diagnostic system that automatically identifies normal and abnormal activities using a minimum number of highly discriminating features for classification. Lee et al. (2014) proposed several hybrid methods to classify normal and epileptic EEG signals using wavelet transform, phase-space reconstruction and Euclidean distance based on a neural network with weighted fuzzy membership functions. The classification of abnormal activities of brain functionality was achieved by understanding abnormal activities caused by changes in neuronal electrochemical activities by identifying EEG signal features (Oweis and Abdulhay 2011). Hilbert weighted frequency was used to help discriminate between healthy and seizure EEG patterns. Li et al. (2013) designed a method for feature extraction and pattern recognition of ictal EEGs, based upon empirical model decomposition (EMD) and SVM. First, an EEG signal was decomposed into intrinsic mode functions (IMFs) using the EMD, and then the coefficients of the variation and the fluctuation index of the IMFs were extracted as features. Pachori et al. (2014) presented a method for the classification of ictal and seizure-free EEG signals based on the EMD and the second-order difference plot (SODP). The EMD method decomposed an EEG signal into a set of symmetric and band-limited signals (the IMFs). The SODP of the IMFs provided an elliptical structure. The importance of the entropy-based features was presented by Kumar et al. (2010) for recognizing normal EEGs, and ictal as well as interictal epileptic seizures. Three nonlinear features, such as wavelet entropy, sample entropy and spectral entropy, were used to extract quantitative entropy features from the given EEGs, and these features were used as the input into two neural network models: recurrent Elman network and radial basis network for the classification. Siuly et al. (2011a) introduced a new approach based on the simple random sampling (SRS) technique with least square support vector machine (LS-SVM) to classify EEG signals. In that study, the SRS technique was applied in two stages to select random samples and sub-samples from each EEG channel data file and finally different features were evaluated from each sub-sample set to represent the original EEG signals and reduced the dimensionality of the data that were used as inputs to the LS-SVM classifier.

From the literature, it is observed that many of the feature extraction methods were not a perfect choice for non-stationary epileptic EEG data (Siuly et al. 2014a) and the reported methods were limited in their success rate and effectiveness. Some of the existing methods cannot work properly for a large EEG data set. None of the prior methods considered the variability of the observations within a window where the time variability is an important consideration for describing the characteristics of the original data. In most of the cases, the methods did not select their parameters

through experimental evaluation, even though the parameters significantly affect the classification performance.

This research intends to design a robust method for feature extraction based on the optimum allocation (OA) and principal component analysis (PCA), denoted as OA_PCA. The reason for using the OA method is to acquire representative sampling points from the huge amount of EEG data when the data is heterogeneous. It is worth mentioning that epileptic EEG data is heterogeneous in term of the time period. The idea to use the PCA is due to the fact that, in most of the EEG data, there is a large amount of redundant information unnecessary for diagnostic applications. The PCA method is practical when there is a large number of correlated variables in a dataset and it is believed that there are some redundancies in those variables. The PCA is an excellent method for making uncorrelated variables that can contribute to enhanced classification accuracy and it can also reduce the dimensions of the data for easy handling. The core objective of this research is to develop a new method for feature extraction and discover a suitable classifier for the features that can improve overall classification accuracy (OCA) with low false alarm rate (FAR) Siuly and Li (2015).

To create stationarity of the EEG signals, this study partitions EEG data into a number of time-windows (discussed in detail in Sect. 7.2). Then the OA scheme is used to search representative sampling points from each window, which can efficiently describe the signals based on the variability of their observations within the window. After that, the PCA method is employed on the OA data to produce uncorrelated variables and to reduce dimensionality. The obtained principal components are treated as features in this study, and called the OA_PCA feature set. To identify an efficient classifier for the OA_PCA feature set, this study employs four prominent classifiers, namely least square support vector machine (LS-SVM), naive bayes classifier (NB), *k*-nearest neighbor algorithm (KNN) and linear discriminant analysis (LDA). In this research, four different output coding approaches of the multi-class LS-SVM [error correcting output codes (ECOC), minimum output codes (MOC), One versus One (1vs1) and One versus All (1vsA)] are also investigated to test which one is the best for the obtained features. The parameters of the proposed classification methods are selected by extensive experimental evaluations.

7.2 Principal Component Analysis-Based Optimum Allocation Scheme

This study develops a new framework for classifying epileptic EEG signals as presented in Fig. 7.1. The research work investigates and explores whether the OA based PCA features, (named OA_PCA) are suitable for epileptic EEG signal classification, evaluates which classifier is more suitable for the feature set. From Fig. 7.1, it can be seen that the proposed methodology is divided into six major

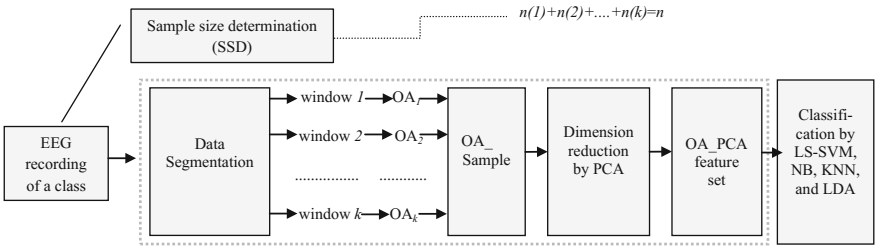


Fig. 7.1 Architecture of the proposed system for the classification of epileptic EEG signals *Note* OA_i = optimum allocation in the *i*th window; OA_Sample: selected sample by the OA together from each window; OA_PCA: feature vector set obtained by optimum allocation and principal component analysis; SSD = *n*; $n(1)$ = sample size by OA_1 from window 1; $n(2)$ = sample size by OA_2 from window 2 and $n(k)$ = sample size by OA_k from window *k*

parts: sample size determination (SSD), data segmentation, sample selection by OA (denoted as OA_Sample), dimension reduction by PCA, OA based PCA features (named as OA_PCA) and classification by the LS-SVM, NB, KNN and LDA. Brief explanations of these six parts are provided below.

7.2.1 Sample Size Determination (SSD)

The first step of the proposed method is to determine an appropriate sample size from an EEG channel data of a class. From a statistical sense, a sample is a set of observations from a population. In this study, the entire EEG data of a class (class means a specific category of EEG data, e.g. Class 1 (Set Z): healthy persons with eyes open; Class 2 (Set O): healthy persons with eyes closed etc.), is considered as a population where a sample is considered as a representative part of the population. An observation in a sample is called a sample unit, and the sample size is the number of observations that are included in a sample. In this study, the following formulas (Eqs. (7.1) and (7.2)) (*Sample size calculator*; Siuly and Li 2015) are used to calculate the desired samples for each EEG class data.

$$n = \frac{Z^2 \times p \times (1-p)}{e^2} \tag{7.1}$$

where, n = desired sample size; Z = standard normal variate (the value for Z is found in statistical tables which contain the area under the normal curve) for the desired confidence level (1.96 for 95% confidence level and 2.58 for 99% confidence level) (ref. *Z Distribution Table*); p = estimated proportion of an attribute, that is present in the population; e = margin of error or the desired level of precision (e.g. $e = 0.01$ for 99–100% confidence interval). If a population is finite, the required sample size is given by

$$n_{\text{new}} = \frac{\text{n}}{1 + \frac{n-1}{N}} \tag{7.2}$$

where, N = population size.

If estimator p is not known, 0.50 (50%) is used as it produces the largest sample size. The larger the sample size, the more sure we can be that the answers truly reflect the population. In this research, we consider $p = 0.50$ so that the sample size is maximum and $Z = 2.58$ and $e = 0.01$ for 99% confidence level in Eq. (7.1). In this study, $N = 4097$ and thus we obtain from Eq. (7.2), $n_{\text{new}} = 3288$ for each and every five classes (Set Z, Set O, Set N, Set F, Set S) of the epileptic EEG data.

7.2.2 Data Segmentation

EEG signals are nonstationary and stochastic. But, during signal analysis, an EEG signal must be made stationary. The term 'non-stationary' means that the statistical properties of a signal change over time. For example, the signal's mean, variance, kurtosis and skewness do not remain constant over the entire duration of the signal but change from one point in the signal to the next. 'Stochastic' refers to signals where the events in the signal occur in a random fashion, and self-similar, at the simplest level, means that if a portion of a signal is magnified, the magnified signal will look the same and have the same statistical properties as the original signal. Although an overall signal may not be stationary, usually smaller windows, or parts of those signals will exhibit stationarity. An EEG signal is stationary for a small amount of time. That is why we partition the recorded EEG time series signals of every category into several segments based on a specific time period to properly account for possible stationarities. In this methodology, we divide the EEG signals of a class into k-mutually exclusive groups, called 'window' considering a particular time period. For any experiments, the number (k) of windows is determined empirically over time based on the data size. Suppose the sizes of the windows are N_1, N_2,…, N_k. Hence, the number of observations in window 1, window 2 and… window k are N_1, N_2,…, N_k, respectively.

In this research, the EEG signals of a class are segmented into four ($k = 4$) windows with respect to a specific time period based on the data structure. Each window contains 5.9 s of EEG data as every channel consists of 4097 data points of 23.6 s. Thus, the sizes of the four windows (window 1, window 2, window 3 and window 4) are $N_1 = 1024$, $N_2 = 1024$, $N_3 = 1024$ and $N_4 = 1025$, respectively.

7.2.3 OA_Sample

Optimum allocation (OA) refers to a method of sample allocation that is designed to provide the most precision. A detailed discussion of the OA scheme is provided in

Sect. 6.2 in Chap. 6 and also available in reference (Siuly and Li 2014b, 2015). In this study, the OA technique is used to determine the number of observations to be selected from different windows. To find out how a given total sample size, n, should be allocated among k windows with the smallest possible variability, we calculate the best sample size for the ith window using the OA (Islam 2007; Cochran 1977) technique by Eq. (7.3). All of the selected samples by the OA from all the windows together are denoted as OA_Sample in this study. Detailed discussion of the OA technique is available in references (Siuly and Li 2014b, 2014c, 2015).

$$n(i) = \frac{N_i \sqrt{\sum_{j=1}^{p} s_{ij}^2}}{\sum_{i=1}^{k} \left(N_i \sqrt{\sum_{j=1}^{p} s_{ij}^2} \right)} \times n, \quad i = 1, 2, \ldots, k; \;\; j = 1, 2, \ldots, p \tag{7.3}$$

where $n(i)$ is the required sample size of the ith window; N_i is the data size of the ith window; s_{ij}^2 is the variance of the jth channel of the ith window; and n is the sample size of the EEG recording of a class obtained by Eq. (7.1). If $n(1)$, $n(2)$, …, $n(k)$ are the sample sizes obtained by Eq. (7.3) from the window sizes, N_1, N_2,…, N_k, respectively, the sum of all obtained sample sizes from all the windows in a class will be equal to n, i.e. $n(1 + n(2) + \cdots + n(k) = n$ as shown in Fig. 7.1.

In this work, k (the number of windows in each class) = 4 and p (the number of channels in each class) = 100. For each of the five classes, the sizes of Windows 1, 2, 3 and 4 are 1024, 1024, 1024 and 1025, respectively. The numbers of observations are determined to be selected from each window in a class using the OA by Eq. (7.3) as reported in Table 7.1. From Table 7.1, it is seen that in Set Z, the sample sizes: 797, 822, 837 and 832, are obtained by the OA from Windows 1, 2, 3 and 4, respectively. The total of these four samples is 3288 that is equal to the total sample size, n = 3288. Note that the total sample size, n of a class (e.g. Set Z or Set O, Set N or Set F or Set S) is allocated among different windows with the smallest possible variability by the OA procedure. As shown in Table 7.1, the sample sizes are not equal in every window of a class due to the variability of the observations in different windows. If the variability within a window is large, the size of the sample from that window is also large. On the other hand, if the variability of the observations within a window is small, the sample size will be small in that window. Finally, we create an OA_Sample for Set Z such as {797,822,837,832} which contains 3288 sample units. Similarly, we acquire the OA_Sample from Sets O, N, F, S, individually, which have 3288 sample units, for example {815, 840, 805, 828} for Set O; {839, 841, 780, 828} for Set N; {828, 833, 788, 839} for Set F; and {833, 844, 815, 796} for Set S. It is worth mentioning that each OA_Sample set has 100 dimensions because each class contains 100 channels of EEG data in the epileptic EEG database. Thus, the OA_Sample for each class consists of 3288 observations of 100 dimensions and all of the OA_Samples together from each of the five classes consists of 16,440 observations of 100 dimensions. The entire set (16,440 × 100) of the OA_Samples from the five classes is denoted as ***OA_Sample set*** in the experiments of this study.

Table 7.1 Calculated sample size for each window by the OA scheme

Segment	Sample sizes obtained				
	Set Z	Set O	Set N	Set F	Set S
Window 1	797 (1024)	815 (1024)	839 (1024)	828 (1024)	833 (1024)
Window 2	822 (1024)	840(1024)	841 (1024)	833 (1024)	844 (1024)
Window 3	837 (1024)	805 (1024)	780 (1024)	788 (1024)	815 (1024)
Window 4	832 (1025)	828 (1025)	828 (1025)	839 (1025)	796 (1025)
Total	3288	3288	3288	3288	3288

Note In each cell, the number inside the parentheses is the window size (e.g. (1024) in the 1st cell) and the number outside of the parentheses is the calculated sample size (e.g. 797 in the 1st cell) obtained by the OA

7.2.4 Dimension Reduction by PCA

At this step, the PCA is used to reduce the dimensionality of the ***OA_Sample set*** and to generate fewer numbers of uncorrelated variables that are utilized as features for better classification of epileptic

EEG signals. Generally, the recorded multi-channel EEG signals are huge in number, containing a large amount of redundant information and are highly correlated. Often different signals from different scalp sites do not provide the same amount of discriminative information. If there is a large amount of redundant information and a large number of correlated variables in a dataset, the PCA is a powerful technique to transform a number of correlated variables into a smaller number of uncorrelated variables called principal components. The principal components represent the most informative data carried by the original signals to provide the discriminative information about those signals. Thus, the PCA features work better in EEG signal classification. A detailed explanation of the PCA is found in references (Cao et al. 2003; Chawla et al. 2006; Duda et al. 2001; Turk and Pentl 1991). Figure 7.2 presents an example of the PCA features for the five classes of EEG signals. In this figure, we consider 100 feature points for each of the five classes. The patterns of the PCA features for each class are similar as their original signal patterns (as shown in Fig. 7.1).

In this work, we apply the PCA method to the ***OA_Sample set*** of five classes that have 16,440 data points of 100 dimensions. Here eigenvectors with eigen values greater than one are chosen. Considering eigen values greater than one, we obtain a principal component set which contains 16,440 data points of 37 dimensions. Figure 7.3 illustrates a box plot of a whole feature vector set for five class datasets. The experimental results show that the EEG channels' data are highly correlated with each other and the PCA components represent 68.22% of the variation that confirms an appropriate use of the PCA approach for this research. Thus, the obtained principal components of the ***OA_Sample set*** are used as the features denoted as 'OA_PCA'.

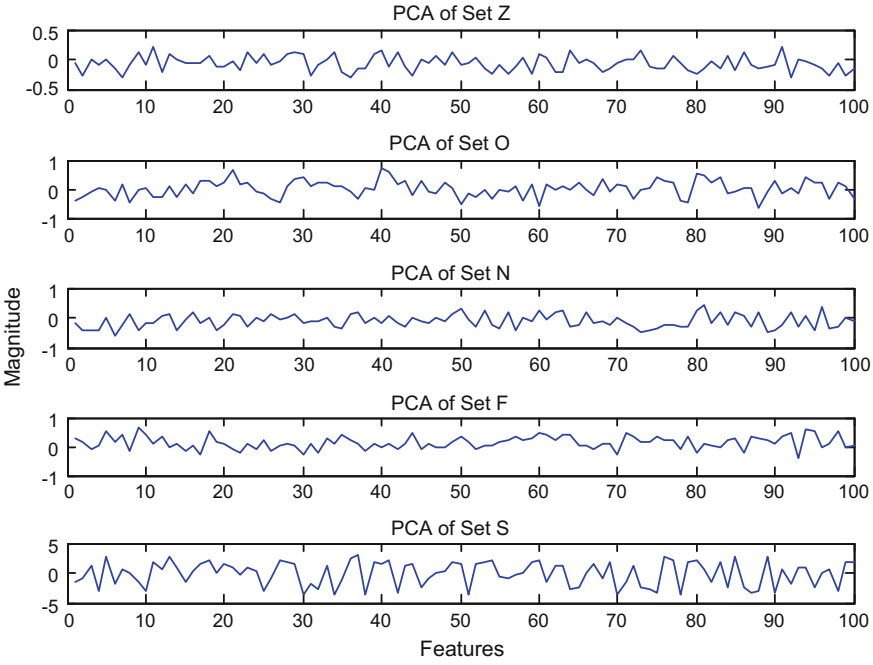


Fig. 7.2 An example of PCA features for five classes EEG data sets

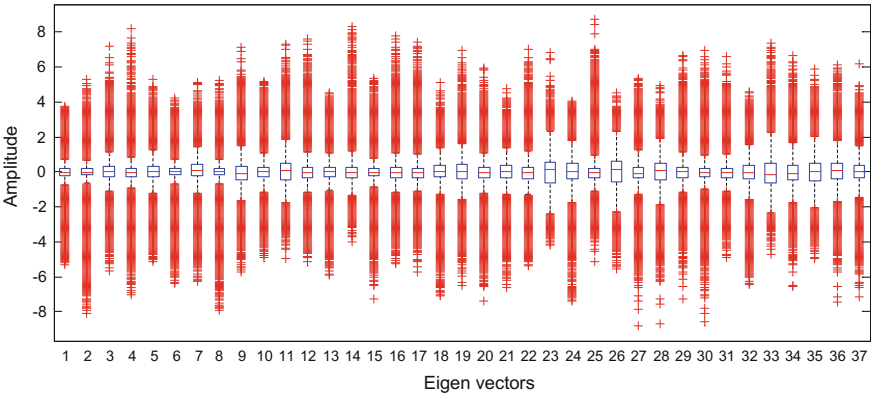


Fig. 7.3 Illustration of the entire feature vector set

7.2.5 OA_PCA Feature Set

After reducing the dimensions of the ***OA_Sample set*** by the PCA, the new generated feature vector set is denoted as 'OA_PCA'. In this study, we obtain an OA_PCA feature set that has 16,440 data points of 37 dimensions as discussed in

the last section. This feature vector set is divided into a training set and a testing set using a six fold cross-validation method, which is discussed in Sect. 7.3. In each of the six trials, the training set consists of 13,700, and the testing set consists of 2740 observations with 37 dimensions. Thus, the percentages of the training set and testing set are 83.33 and 16.67%, respectively. In this study, the training set is applied to train a classifier and the testing vectors are used to evaluate the accuracy and the effectiveness of the chosen classifiers in the classification. The testing set is used as the input to the classification algorithm. The obtained OA_PCA feature set is fed to each of the four classifiers discussed in the following section.

7.2.6 Classification by the LS-SVM, NB, KNN and LDA

After the feature extraction, the feature vector set, OA_PCA, is forwarded to each mentioned classifier. To choose the most appropriate classifier for the OA_PCA feature set, this study employs four prominent classifiers: LS-SVM, NB, KNN and LDA, for the classification of epileptic EEG signals. Brief explanations of those methods are provided below.

- **Least square support vector machine (LS-SVM)**

The LS-SVM proposed by Suykens and Vandewalle (1999) is one of the most successful classifying models. The LS-SVMs use equality constraints instead of inequality constraints and solve linear equations instead of the quadratic programming that reduces the computational cost of the LS-SVM (Suykens et al. 2002; Xing et al. 2008; Suykens and Vandewalle 1999) compared to the original SVM. A straightforward extension of LS-SVMs to multiclass problems has been proposed by Suykens and Vandewalle (2002). A common way of solving the multiclass categorization problem is to reformulate the problem into a set of binary classification problems using four output coding schemes: minimum output codes (MOC), error correcting output codes (ECOC), One versus One (1vs1) and One versus All (1vsA). The detailed description of the multiclass LS-SVM algorithms can be found in Sect. 6.2.5 of Chap. 6 and also in the literature (Suykens and Vandewalle 1999; Suykens et al. 2002; Xing et al. 2008; Youping 2005; Sengur 2009; Gestel et al. 2002). In this research, the multiclass LS-SVM with four output coding schemes: ECOC, MOC, 1v1 and 1vA is denoted as LS-SVM_ECOC, LS-SVM_MOC, LS-SVM_1v1 and LS-SVM_1vA, respectively. Then they are all applied with the radial basis function (RBF) for the classification of multiclass EEG signals, using the OA_PCA features as the input. Before the classification procedure, the hyper parameters of the LS-SVM (γ and σ^2) with RBF kernel in each of the four output coding schemes are selected through the extensive experimental evaluations in this research, as discussed in Sect. 7.4.1.

- **Naive Bayes classifier (NB)**

A NB is a probabilistic classifier based on Bayes theorem with strong (naive) independence assumptions. The NB assumes all the feature nodes are to be independent of each other for a given class, and typically the feature variables are assumed to have a Gaussian distribution if they are continuous. The classification results are determined by the posterior probability $P(y|x_1, x_2, x_3, \ldots, x_n)$ described in detail in references (Friedman 1997; Zhang 2004; Siuly et al. 2016), which can be transformed by using the chain rule and Bayes' Theorem. In this study, the class node, y, presents the five categories of the epileptic EEG signals and the feature nodes $(x_1, x_2, x_3, \ldots, x_n)$ represent the OA_PCA features. A detailed explanation of this method is available in Chap. 12 and also in references (Friedman 1997; Zhang 2004; Hope et al. 2011).

- **K-Nearest Neighbor algorithm (KNN)**

The idea behind the KNN algorithm is quite straightforward in classifying objects based on the closest training observations presented in the feature space (Duda et al. 2001; Cover and Hart 1967). Here, an object is classified by a majority vote of its neighbours, with the object being assigned to the class most common amongst its k nearest neighbours (k is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbour (Cover and Hart 1967; Han et al. 2005). The only adjustable parameter in the model is k, the number of the nearest neighbours to be included in the estimate of a class membership. The value of $P(y|x)$ is calculated simply as the ratio of members of class y among the k nearest neighbours of x. A detailed discussion of this method is available in references (Duda et al. 2001; Cover and Hart 1967; Han et al. 2005; Ripley 1996). In this study, an appropriate value of k is selected after the experimental assessments, as described in Sect. 7.4.1.

- **Linear Discriminant Analysis (LDA)**

A LDA (also known as Fisher's LDA) builds a predictive model for a group membership. The LDA uses hyper-planes to separate data representing different classes (Duda et al. 2001). The model is composed of a discriminant function based on linear combinations of predictor variables. The purpose of the LDA is to maximally separate the groups and to determine the most and parsimonious way to separate groups and to discard variables which are little related to group distinctions for online systems. This technique has a very low computational requirement. A detailed discussion of this method is available in references (Duda et al. 2001; Friedman et al. 1997). In this research, the OA_PCA features are used as the predictor variables for epileptic EEG data classification in this model.

7.3 Performance Assessment

To evaluate the performances of the proposed framework, this study applies a *k*-fold cross-validation (see description in Chap. 3) and also in references (Siuly and Li 2012; Siuly et al. 2011) method. In this study, we select $k = 6$ as the six fold cross-validation is found adequate for this dataset. As mentioned before, we obtain a total of 16,440 feature vectors of 37 dimensions where each of the five classes holds 3288 feature vectors of the same dimensions. In this research, we divide the whole feature sets into six subsets where each subset is called *fold*. Each fold consists of 2740 feature vectors of the five classes taking 548 feature vectors from each class. Figure 7.4 presents a design of how the extracted feature vectors of this study are partitioned into six mutually exclusive subsets (folds) according to the *k*-fold cross-validation system.

The procedure is repeated 6 times. Each time, one subset (fold) is used as a testing set and the remaining five subsets (folds) are used as a training set. As shown in Fig. 7.4, in each iteration, the testing set is denoted as **bold** face and other five folds together are used as the training set. Thus, the training set consists of 13,700 feature vectors of 37 dimensions and the testing set contains 2740 feature vectors of the same dimension. The classification performances from each of six iterations on the testing set are obtained for all categories of the epileptic EEG signals and summarized in Table 7.6 in Sect. 7.4.2.

In this study, the performances of the proposed methods are assessed based on different statistical measures, such as corrected percentage (CP) in each class, overall classification accuracy (OCA) and false alarm rate (FAR). Their formulas are given below (Murugavel et al. 2011; Siuly and Li 2012; Siuly et al. 2009, 2010, 2011; Siuly and Wen 2013; Bajaj and Pachori 2012, 2013):

- Corrected percentage (CP) in each class (also called sensitivity)

$$\mathrm{CP}_i = \frac{\text{Number of correctly classified cases in } i\text{th class}}{\text{Actual number of cases in } i\text{th class}} \times 100 \tag{7.4}$$

1st iteration	**Fold 1**	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6
2nd iteration	Fold-1	**Fold 2**	Fold-3	Fold-4	Fold-5	Fold-6
3rd iteration	Fold-1	Fold-2	**Fold 3**	Fold-4	Fold-5	Fold-6
4th iteration	Fold-1	Fold-2	Fold-3	**Fold 4**	Fold-5	Fold-6
5th iteration	Fold-1	Fold-2	Fold-3	Fold-4	**Fold 5**	Fold-6
6th iteration	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	**Fold 6**

Fig. 7.4 Structure of performing six fold cross validation process

- Overall classification accuracy (OCA)

$$\text{OCA} = \frac{\text{Number of total correct decisions for all classes}}{\text{Total number of cases for all classes}} \times 100 \tag{7.5}$$

- False alarm rate (FAR)

$$\text{FAR}_i = \frac{\text{Number of of normal cases that are detected as not a normal cases in } i\text{th category}}{\text{Total number of normal cases}} \times 100 \tag{7.6}$$

7.4 Experimental Set-up

In this section, we first select the parameters of the four mentioned classifiers: LS-SVM, NB, KNN and LDA, because the classification performances of a classifier depend on the values of the parameters as discussed in Sect. 7.4.1. Then we investigate the effectiveness of each classifier on the OA_PCA features with the optimally selected parameters discussed in Sect. 7.4.2. The classification by the multi-class LS-SVM is carried out in MATLAB (version 7.14, R2012a) using the LS-SVMlab toolbox (version 1.8). The classification of the NB method is also performed in MATLAB with the same version. On the other hand, the classification executions for the KNN and LDA methods are executed in SPSS package (version 21).

7.4.1 Parameter Selection

As mentioned before, this study uses four classification methods, such as LS-SVM, NB, KNN and LDA. The four output coding schemes of the multi-class LS-SVM: ECOC, MOC, 1vs1 and 1vsA, are explored in this research. As there are no specific guidelines to set the values of these parameters of the mentioned classifiers, this study presents a new way to select the parameters. In this study, the RBF kernel function is employed for the multi-class LS-SVM as an optimal kernel function over the different kernel functions that were tested. The multi-class LS-SVM with RBF has two important parameters, γ and σ^2, which should be appropriately chosen to achieve the desired performance as these parameters play a significant role in the classification performance. The NB classifier is a probabilistic classifier based on Bayes' theorem and there is no parameter in this method. The KNN model has only one parameter k which refers to the number of nearest neighbors. By varying k, the model can be made more or less flexible. In this study, the appropriate value of the

model parameter, k is selected after experimental evaluations. In the LDA method, the parameters are obtained automatically through the maximum likelihood estimation (MLE) method.

From the above discussion, it is clear that we only determine the optimal parameter values, γ and σ^2, for the LS-SVM classifier of each output coding system and an appropriate k value for the KNN classifier. As mentioned before, we denote the LS-SVM for the four output coding schemes: ECOC, MOC, 1v1 and 1vA, as LS-SVM_ECOC, LS-SVM_MOC, LS-SVM_1v1 and LS-SVM_1vA, respectively. To select appropriate combinations of (γ, σ^2) for all four LS-SVM approaches, we set up the ranges of γ and σ^2 values as γ = 1, 10, 100, 1000 and σ^2 = 1, 10, 100, 1000 for the experimental evaluations. Thus, we obtain classification outcomes for all possible combinations of γ and σ^2 and then take one combination as an optimal combination for each fold that reports the highest overall classification accuracy (OCA) rate for that fold. In each fold, the best combination of (γ, σ^2) for the LS-SVM_ECOC, LS-SVM_MOC, LS-SVM_1v1 and LS-SVM_1vA classifiers with their OCAs are presented in Tables 7.2, 7.3, 7.4 and 7.5, respectively. In every fold, the selected combinations of (γ, σ^2) for each of the LS-SVMs are highlighted in **bold** font.

Table 7.2 reports that one combination of (γ, σ^2) is obtained in Fold-1, three combinations in Fold-2, four combinations in Fold-3, two combinations in Fold-4, three combinations in Fold-5 and one combination in Fold-6. From this table, it is seen that γ = 100, σ^2 = 10 is the common combination for the LS-SVM_ECOC approach in each fold. Thus, we select γ = 100, σ^2 = 10 as an optimal combination for the LS-SVM_ECOC method. Table 7.3 also shows that γ = 100, σ^2 = 10 is the common combination that produces highest performance in every fold for the LS-SVM_MOC method. Hence γ = 100, σ^2 = 10 is considered to be the best combination for this method. In Table 7.4, it can be seen that the selected optimal parameters for the LS-SVM_1v1 classifier is γ = 10, σ^2 = 10 that yields 100% classification accuracy in each fold. Table 7.5 reports that γ = 100, σ^2 = 10 is the best pair for the hyper parameters of the LS-SVM_1vA method.

As mentioned previously, there is only one model parameter in the KNN classifier, and that is k. The choice of k is essential in building up a model which can be

Table 7.2 Selection of hyper parameters (γ, σ^2) of the LS-SVM _ECOC classifier

LS-SVM_ECOC		
Fold	Parameters (γ, σ^2)	OCA (%)
Fold-1	**γ = 100, σ^2 = 10**	99.96
Fold-2	**γ = 100, σ^2 = 10**; γ = 1000, σ^2 = 10; γ = 10, σ^2 = 10	100.00
Fold-3	γ = 1, σ^2 = 10; γ = 10, σ^2 = 10; **γ = 100, σ^2 = 10**; γ = 1000, σ^2 = 10	100.00
Fold-4	γ = 10, σ^2 = 10; **γ = 100, σ^2 = 10**	99.96
Fold-5	γ = 10, σ^2 = 10; **γ = 100, σ^2 = 10**; γ = 1000, σ^2 = 10	100.00
Fold-6	**γ = 100, σ^2 = 10**	99.89

Note Selected parameters for the LS-SVM_ECOC approach: **γ = 100, σ^2 = 10**

Table 7.3 Selection of hyper parameters (γ, σ^2) of the LS-SVM_MOC

LS-SVM_MOC		
Fold	Parameters (γ, σ^2)	OCA (%)
Fold-1	**γ = 100, σ^2 = 10**	99.96
Fold-2	γ = 10, σ^2 = 10; **γ = 100, σ^2 = 10**; γ = 1000, σ^2 = 10	100.00
Fold-3	γ = 1, σ^2 = 10; γ = 10, σ^2 = 10; **γ = 100, σ^2 = 10**; γ = 1000, σ^2 = 10	100.00
Fold-4	γ = 10, σ^2 = 10; **γ = 100, σ^2 = 10**	99.96
Fold-5	γ = 10, σ^2 = 10; **γ = 100, σ^2 = 10**; γ = 1000, σ^2 = 10	100.00
Fold-6	**γ = 100, σ^2 = 10**	99.89

Note Selected parameters for the LS-SVM_MOC approach: **γ = 100, σ^2 = 10**

Table 7.4 Selection of hyper parameters (γ, σ^2) of the LS-SVM_1v1

LS-SVM_1v1		
Fold	Parameters (γ, σ^2)	OCA (%)
Fold-1	γ = 1, σ^2 = 10; **γ = 10, σ^2 = 10**; γ = 100, σ^2 = 10; γ = 100, σ^2 = 100; γ = 1000, σ^2 = 10; γ = 1000, σ^2 = 100	100.00
Fold-2	**γ = 10, σ^2 = 10**; γ = 100, σ^2 = 10; γ = 100, σ^2 = 100; γ = 1000, σ^2 = 10; γ = 1000, σ^2 = 100;	100.00
Fold-3	γ = 1, σ^2 = 10; **γ = 10, σ^2 = 10**; γ = 100, σ^2 = 10; γ = 100, σ^2 = 100; γ = 1000, σ^2 = 10; γ = 1000, σ^2 = 100	100.00
Fold-4	γ = 1, σ^2 = 10;**γ = 10, σ^2 = 10**; γ = 100, σ^2 = 10; γ = 1000, σ^2 = 10	100.00
Fold-5	γ = 1, σ^2 = 10; **γ = 10, σ^2 = 10**; γ = 100, σ^2 = 10; γ = 1000, σ^2 = 10; γ = 1000, σ^2 = 100	100.00
Fold-6	**γ = 10, σ^2 = 10**	100.00

Note Selected parameters for the LS-SVM_1v1 approach: **γ = 10, σ^2 = 10**

Table 7.5 Selection of hyper parameters (γ, σ^2) of the LS-SVM_1vA

LS-SVM with 1vA		
Fold	Parameters (γ,σ^2)	OCA(%)
Fold-1	**γ = 100, σ^2 = 10**; γ = 1000, σ^2 = 10	99.96
Fold-2	**γ = 100, σ^2 = 10**; γ = 1000, σ^2 = 10	100.00
Fold-3	γ = 10, σ^2 = 10; **γ = 100, σ^2 = 10**; γ = 1000, σ^2 = 10	100.00
Fold-4	γ = 10, σ^2 = 10; **γ = 100, σ^2 = 10**	99.96
Fold-5	γ = 10, σ^2 = 10; **γ = 100, σ^2 = 10**; γ = 1000, σ^2 = 10	100.00
Fold-6	**γ = 100, σ^2 = 10**; γ = 1000, σ^2 = 10	99.85

Note Selected parameters for the LS-SVM_1vA approach: **γ = 100, σ^2 = 10**

regarded as one of the most important factors of the model that can strongly influence the quality of predictions. *k* should be set to a suitable value to minimize the probability of misclassification. The best choice of *k* depends upon the data. In this study, we select the appropriate *k* value in an automatic process following the

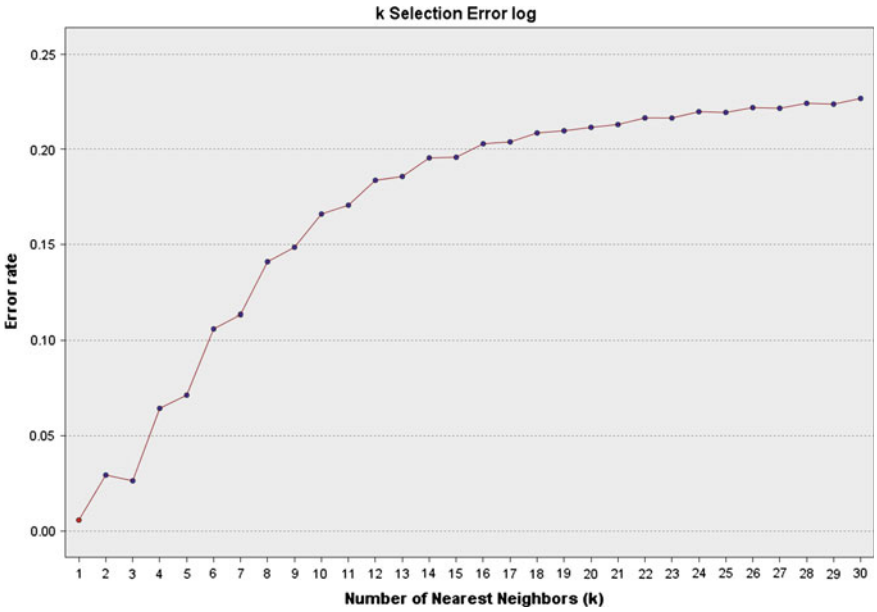


Fig. 7.5 *k* selection error log in fold-1 of the testing set

k selection error log as there is no a simple rule for selecting *k*. We consider the range of the *k* value being between 1 and 30. We pick an appropriate *k* value that results in the lowest error rate in each fold as the lowest error rate indicates having the best model. In the experimental results, we obtain the lowest error rate for $k = 1$ in every fold. Figures 7.5 and 7.6 show the two examples for Fold-1 and Fold-2 of the testing data set. In these two figures, it is seen that the number of nearest neighbours (*k*) are plotted in the *X*-axis and the error rate of the KNN model for each *k* value is presented in the *Y*-axis. The error rate increases in conjunction with an increase in the number of nearest neighbours (*k*). For example, in Fig. 7.6, the error rate is about 0.01 for $k = 1$ while it is about 0.07 for $k = 5$. Again, in Fig. 7.6, the error rate is about 0.06 for $k = 1$ when it is about 0.17 for $k = 5$. Hence the appropriate value of *k* is 1 for this epileptic dataset. Thus, the lowest error rate is obtained with this value in each fold of the dataset.

7.4.2 Results and Discussions

Table 7.6 presents the classification results for the OA_PCA feature set with the classifiers of LS-SVM_ECOC, LS-SVM_MOC, LS-SVM_1v1, LS-SVM_1vA, NB, KNN and LDA on the epileptic EEG data. In this study, the average of the corrected percentage (CP) in a category [calculated by Eq. (7.4)] of the six folds is

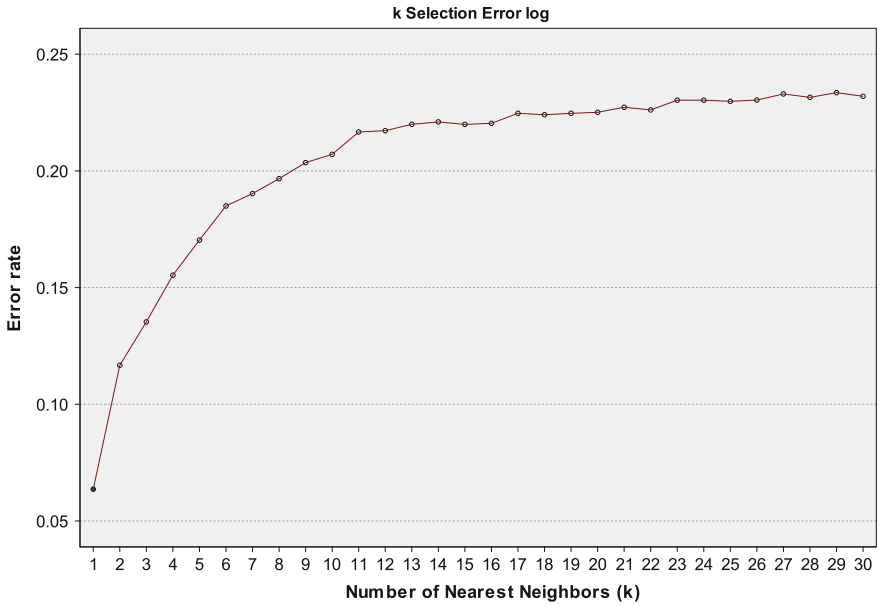


Fig. 7.6 *k* selection error login Fold-2 of the testing set

named as *CP_average*, and the variation among results from the six folds is denoted as *standard deviation* (*SD*). On the other hand, the average of the overall classification accuracy (OCA) [calculated by Eq. (7.5)] for the six folds is named *OCA_average*, and the deviation among the six folds is denoted as *standard deviation* (*SD*). As shown in Table 7.6, the highest classification performance is achieved for the LS-SVM_1v1 classifier, which is 100% for each and every category. The LS-SVM_ECOC and LS-SVM_MOC classifiers get the second position and the LS-SVM_1vA is in third position according to the CP_average and OCA_average for the OA_PCA feature set. The LDA classifier yields the lowest performance among the seven classification methods. It can be seen from Table 7.6 that the LS-SVM classifier with each of the four output coding systems produces relatively better performances for the OA_PCA feature set compared to the other three classifiers: the NB, KNN and LDA. On the other hand, the NB produces a better performance than those of the KNN and LDA. Furthermore, Table 7.6 reports that the SD for every classifier is very low, which leads to the conclusion that the mentioned classifiers are consistent for the OA_PCA features. It is also observed that the SD of the classification performance for the LS-SVM_1v1 classifier is zero, which indicates the robustness of the LS-SVM_1v1 for obtained features. The lower value of the SD represents the consistency of the proposed method. Hence, the results demonstrate that the LS-SVM is a superior classifier for epileptic EEG signal classification. The LS-SVM_1v1 is the best method.

Table 7.6 Classification results on the epileptic EEG data

Different classifiers	CP_average ± SD					OCA_average ± SD
	Set Z	Set O	Set N	Set F	Set S	
LS-SVM_ECOC	100.0 ± 0.00	99.85 ± 0.21	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00	99.97 ± 0.04
LS-SVM_MOC	100.0 ± 0.00	99.85 ± 0.21	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00	99.97 ± 0.04
LS-SVM_1v1	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00
LS-SVM_1vA	100.0 ± 0.00	99.85 ± 0.21	100.0 ± 0.00	99.96 ± 0.07	100.0 ± 0.00	99.96 ± 0.06
NB classifier	99.58 ± 0.25	99.70 ± 0.50	99.12 ± 0.47	97.98 ± 1.05	100 ± 0.00	99.24 ± 0.31
KNN classifier	100.0 ± 0.00	99.61 ± 0.27	100.0 ± 0.00	99.91 ± 0.15	94.67 ± 4.95	98.82 ± 1.01
LDA classifier	98.32 ± 1.03	98.25 ± 0.93	95.47 ± 0.46	85.40 ± 1.94	61.51 ± 1.41	87.79 ± 0.44

Note **Set Z**: Normal persons with eyes open; **Set O**: Normal persons with eyes closed; **Set N**: Epileptic patients during seizure-free from the hippocampal formation of the opposite hemisphere of the brain; **Set F**: Epileptic patients during seizure-free within the epileptogenic zone; **Set S**: Epileptic patient during seizure activity

To get a clear understanding of how the six fold cross-validation system produces the CP in each of the six folds in each class for every reported classifier, we provide Fig. 7.7a–g. These figures present the patterns of the corrected percentage (CP) in each class against each of the six folds. Figure 7.7c shows that there are no fluctuations in the performance among the six folds in each of the five classes for the LS-SVM_1v1 classifier. The results indicate that the proposed method is fairly stable. It is also seen from Fig. 7.7a–g, that the fluctuations in the performances among the different folds are very little in each class. From these figures, it is observed that the pattern of the CP in each fold for every class is very consistent, indicating the robustness of the methods. The error bars in the graphs represent the standard errors indicating the fluctuations in the performances among the folds.

Table 7.7 illustrates the false alarm rate (FAR) (sometimes called 'probability of false detection') for the seven classifiers in each of the six folds for Sets O, N, F and S. In this study, we consider Set Z as a normal case and compute the FAR using

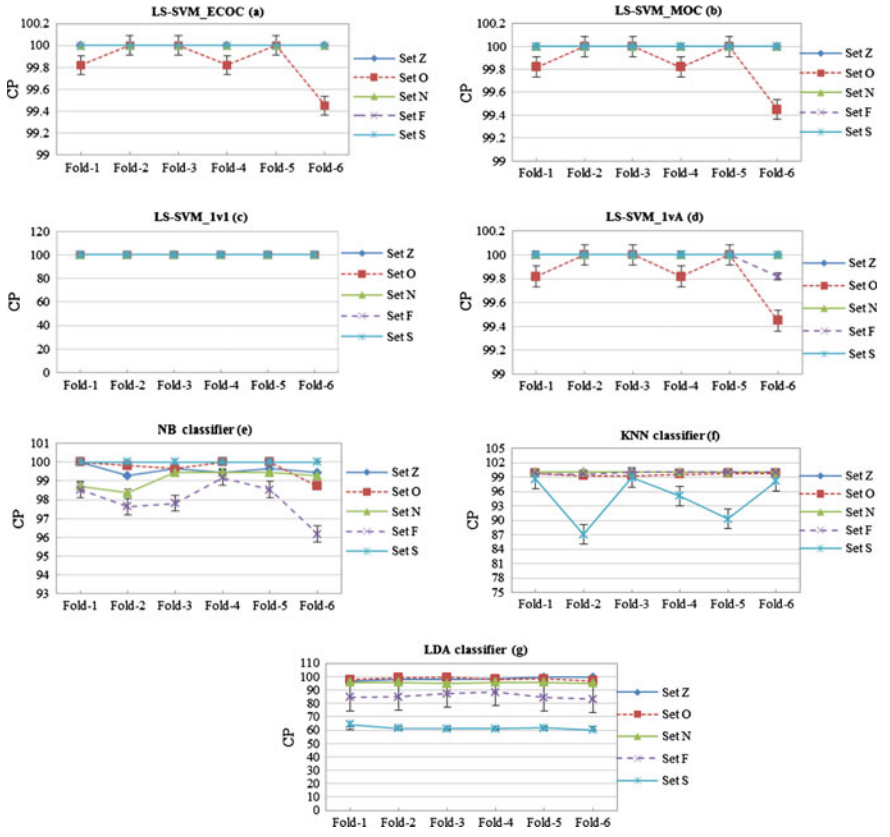


Fig. 7.7 Individual classification performances of each of the six folds in each class for the proposed classifiers: (a) LS-SVM_ECOC, (b) LS-SVM_MOC, (c) LS-SVM_1v1, (d) LS-SVM_1vA, (e) NB, (f) KNN, and (g) LDA. *Error bars* represent the standard error

Table 7.7 Obtained false alarm rate (FAR) for each of the proposed classifiers

Different classifiers	Data set	FAR			
		Set O	Set N	Set F	Set S
LS-SVM_ECOC	Fold-1	0.00	0.00	0.00	0.00
	Fold-2	0.00	0.00	0.00	0.00
	Fold-3	0.00	0.00	0.00	0.00
	Fold-4	0.00	0.00	0.00	0.00
	Fold-5	0.00	0.00	0.00	0.00
	Fold-6	0.00	0.00	0.00	0.00
LS-SVM_MOC	Fold-1	0.00	0.00	0.00	0.00
	Fold-2	0.00	0.00	0.00	0.00
	Fold-3	0.00	0.00	0.00	0.00
	Fold-4	0.00	0.00	0.00	0.00
	Fold-5	0.00	0.00	0.00	0.00
	Fold-6	0.00	0.00	0.00	0.00
LS-SVM_1v1	Fold-1	0.00	0.00	0.00	0.00
	Fold-2	0.00	0.00	0.00	0.00
	Fold-3	0.00	0.00	0.00	0.00
	Fold-4	0.00	0.00	0.00	0.00
	Fold-5	0.00	0.00	0.00	0.00
	Fold-6	0.00	0.00	0.00	0.00
LS-SVM_1vA	Fold-1	0.00	0.00	0.00	0.00
	Fold-2	0.00	0.00	0.00	0.00
	Fold-3	0.00	0.00	0.00	0.00
	Fold-4	0.00	0.00	0.00	0.00
	Fold-5	0.00	0.00	0.00	0.00
	Fold-6	0.00	0.00	0.00	0.00
NB classifier	Fold-1	0.00	0.00	0.00	0.00
	Fold-2	0.00	0.0036	0.0036	0.00
	Fold-3	0.00	0.0036	0.00	0.00
	Fold-4	0.00	0.0055	0.00	0.00
	Fold-5	0.00	0.0036	0.00	0.00
	Fold-6	0.00	0.0055	0.00	0.00
KNN classifier	Fold-1	0.00	0.00	0.00	0.00
	Fold-2	0.00	0.00	0.00	0.00
	Fold-3	0.00	0.00	0.00	0.00
	Fold-4	0.00	0.00	0.00	0.00
	Fold-5	0.00	0.00	0.00	0.00
	Fold-6	0.00	0.00	0.00	0.00
LDA classifier	Fold-1	0.00	0.011	0.020	0.00
	Fold-2	0.00	0.0055	0.015	0.0018
	Fold-3	0.0018	0.013	0.0073	0.00
	Fold-4	0.0018	0.0073	0.00	0.0073
	Fold-5	0.00	0.0036	0.0018	0.00
	Fold-6	0.00	0.00	0.0055	0.00

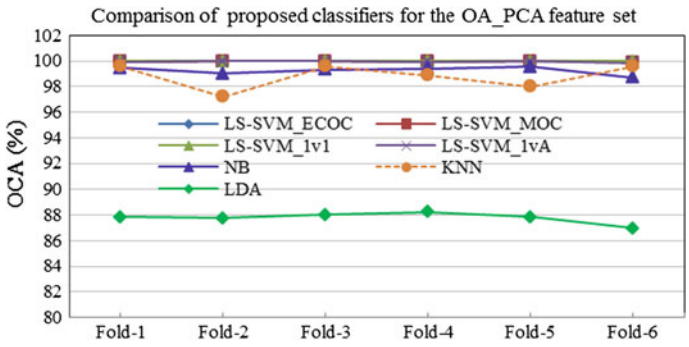


Fig. 7.8 Comparisons of the proposed classifiers for the OA_PCA feature in terms of the OCA

Eq. (7.6). As shown in Table 7.7, the FARs are zero in each of the folds in every class for the LS-SVM_ECOC, LS-SVM_MOC, LS-SVM_1v1, LS-SVM_1vA and KNN. This means that there are no incorrect responses in each fold in every class for those classifiers. For most of the cases, the FARs are also zero for the NB and LDA classifiers. For the value of the FAR, zero indicates that normal cases are detected as normal with 100% accuracy.

7.5 Comparisons

To choose the best classifier for the OA_PCA features, we compare our proposed methods in terms of their OCA against each fold and present the results in Fig. 7.8. It can be seen from the figure that the patterns of the OCA for the LS-SVMs in each of the four output coding systems are very close. That justifies that the four graph lines of the LS-SVM_ECOC, LS-SVM_MOC, LS-SVM_1v1, LS-SVM_1vA are laying in one line showing a higher performance compared to the other classifiers. This figure also demonstrates that the OCA of the LDA is a bit lower compared to the other six classifiers. From Tables 7.6 and 7.8, it is clear that the LS-SVM results are in a better performance with the OA_PCA feature set in the EEG signals classification than the other reported classifiers.

To further examine the efficiencies of the proposed classifiers with the OA_PCA features, we provide the comparisons of our approaches with some recently reported methods. In Table 7.8, we list a comparison of different methods in terms of the OCA for the epileptic EEG database. We present the results from our proposed methods and also the six reported research works discussed in Sect. 7.1. The datasets used in these experiments are also the same. The highest OCA rate among the algorithms is highlighted in **bold** font. From Table 7.8, it is obvious that the result obtained from the approach LS-SVM_1v1 is the best for this database among the recently reported approaches. The results indicate that the LS-SVM_1v1 approach improves the OCA up to 7.10% over the existing algorithms for the epileptic EEG data.

Table 7.8 Comparison with the existing methods on epileptic EEG database

Authors	Methods	OCA (%)
Siuly and Li (the proposed)	OA + PCA + LS-SVM_ECOC	99.97
	OA + PCA + LS-SVM_MOC	99.97
	OA + PCA + LS-SVM_1v1	**100.0**
	OA + PCA + LS-SVM_1vA	99.96
	OA + PCA + NB classifier	99.24
	OA + PCA + KNN classifier	98.82
	OA + PCA + LDA classifier	87.79
Shen et al. (2013)	Wavelet + ApEn + SVM	99.97
	Wavelet + ApEn + KNN	99.10
	Wavelet + ApEn + RBFNN	99.82
Acharjee and Shahnaj (2012)	twelve Cohen class kernel functions + Modular energy + Modularentropy + ANN	98.00
Murugavel et al. (2011)	WT and Lyapunov + SVM_1vA	96.00
	WT and Lyapunov + PNN	94.00
	WT and Lyapunov + RBFNN	93.00
Ubeyli (2009)	Wavelet coefficients + power spectral density + SVM_ECOC	99.20
	Wavelet coefficients + power spectral density + MME classifier	98.68
	Wavelet coefficients + power spectral density + ME	95.00
	Wavelet coefficients + power spectral density + RNN	94.85
	Wavelet coefficients + power spectral density + PNN	95.30
	Wavelet coefficients + power spectral density + CNN	93.48
	Wavelet coefficients + power spectral density + MLPNN	90.48
Ubeyli (2008)	Pisarenko, MUSIC, and minimum-norm + SVM_ECOC	99.30
	Pisarenko, MUSIC and minimum-norm + MLPNN	92.90
Guler and Ubeyli (2007)	wavelet coefficients and lyapunov exponents + SVM_ECOC	99.28
	wavelet coefficients and lyapunov exponents + PNN	98.05
	wavelet coefficients and lyapunov exponents + MLPNN	93.63

WT wavelet transform

7.6 Conclusions

This research presents an optimal feature extraction technique, named as OA_PCA, and also searches for a suitable classifier for the feature set to identify multi-category epileptic EEG signals. The proposed schemes demonstrate many advantages, such as high classification performance and very low FAR for all classifiers tested. The main conclusions of this study are summarized as follows:

- The OA_PCA system is very effective for feature extraction for the epileptic EEG data. The experimental results for the proposed classifiers, the LS-SVM_ECOC, LS-SVM_MOC, LS-SVM_1v1, LS-SVM_1vA, NB, KNN and LDA, confirm that the extracted features are very consistent in detecting epileptic EEG signals
- The parameters of the proposed methods are optimally selected through experimental evaluations indicating the reliability of the methods
- The results show that the proposed LS-SVM_1v1 classifier with the OA_PCA features achieves the best performance compared to the other classifiers with the same features
- The experimental results also indicate that our proposed approach, LS-SVM_1v1 outperforms the other six recently reported research results with the same epileptic EEG database. It demonstrates that our method is considered the best method for the epileptic EEG signal classification.

This study concludes that the LS-SVM_1v1 with the OA_PCA feature set is a promising technique for epileptic EEG signals recognition. It offers great potential for the development of epilepsy analyses. The proposed technique can assist physicians or doctors to diagnose brain function abnormalities. One limitation for performing the proposed optimum allocation based algorithm is dealing with the problem of considering the window size based on an empirical evaluation. In the future, we will extend the proposed approach to online and real-time applications in the multiclass classification problems.

References

Acharjee, P.P., and Shahnaz, C. Multiclass Epileptic Seizure Classification Using Time-Frequency Analysis of EEG Signals, 2012 7th International Conference on Electrical and Computer Engineering, 20–22 December, 2012, Dhaka, Bangladesh, pp. 260–263.

Acharya, U.R., Sree, S.V., Swapna, G., Martis, R.J., and Suri, J.S. Automated EEG analysis of epilepsy: A review, Knowledge-Based Systems 45 (2013) 147–165.

Bronzino, J.D. Principles of Electroencephalography (2nd ed.), The Biomedical Engineering Handbook: Second Edition, Boca Raton: CRC Press LLC, 2000.

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

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

Cochran, W.G. Sampling Techniques, Wiley, New York, 1977.

Cao, L. J., Chua, K. S., Chong, W. K., Lee, H. P., and Gu, Q. M. A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine, Neurocomputing 55 (2003) 321–336.

Chawla, M., Verma, H., and Vinod, K. ECG modeling and QRS detection using principal component analysis, In: Proceedings of IET international conference, paper no. 04, MEDSIP06, Glas-gow, UK, 2006.

Cover, T., Hart, P. Nearest neighbor pattern classification, IEEE Transactions in Information Theory, IT-13 (1) (1967) 21–27.
Duda, R.O., Hart, P.E., and Strok, D.G. Pattern classification (2nd ed.), John, Wiley & Sons, 2001.
Friedman, N., Geiger, D., and Goldszmidt, M. Bayesian network classifiers, Machine Learning 29 (1997) 131–163.
Gestel, TV, Suykens, J.A.K., Lanckriet, G., Lambrechts, A., De Moor, B., and Vandewalle, J. Multiclass LS-SVMs: Moderated Outputs and Coding-Decoding Schemes, Neural processing letters 15(2002) 45–58.
Guler, I., and Ubeyli, E.D. Multiclass support vector machines for EEG-signal classification, IEEE Transactions on Information Technology in Biomedicine 11(2) (2007) 117–126.
Hope, R.M., Wang, Z., Wang, Z., Ji, Q., and Gray, W.D. Workload classification across subjects using EEG, In: Proceedings of the human factors and ergonomics society 55th annual meeting - 2011, pp. 202–206.
Han, J., Kamper, M, and Pei, J. Data mining: Concepts and techniques, Morgan Kaufmann, 2005.
Islam, M.N. An introduction to sampling methods: theory and applications, revised ed., Book World, Dhaka New Market & P.K. Roy road, Bangla Bazar, Dhaka-1100, 2007.
Kumar, S.P., Sriraam, N., Benakop, P.G., Jinaga, B.C. Entropies based detection of epileptic seizures with artificial neural network classifiers, Expert Systems with Applications 37 (2010) 3284–3291.
Lee, S.H., Lim, J.S., Kim, J.K., Yang, J, and Lee, Y. Classification of normal and epileptic seizure EEG signals using wavelet transform, phase-space reconstruction, and Euclidean distance, Computer Methods and Programs in Biomedicine 116 (1) (2014) 10–25.
Li, S., Zhou, W., Yuan, Q., Geng, S., Cai, D. Feature extraction and recognition of ictal EEG using EMD and SVM, Computers in Biology and Medicine 43 (7) (2013) 807–816.
LS-SVMlab toolbox (version 1.8), http://www.esat.kuleuven.ac.be/sista/lssvmlab/.
Murugavel, A.S.M, Ramakrishnan, S., Balasamy, K., and Gopalakrishnan, T. Lyapunov features based EEG signal classification by multi-class SVM, 2011 World Congress on Information and Communication Technologies, pp. 197–201.
Oweis, R.J., and Abdulhay, E.W. Seizure classification in EEG signals utilizing Hilbert-Huang transform, Biomedical Engineering Online 10 (2011) 38–52.
Pachori, R.B., and Patidar, S. Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions, Computer Methods and Programs in Biomedicine 113 (2) (2014) 494–502.
Sharma, R and Pachori, R.B. Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions, Expert Systems with Applications42 (3) (2015) 1106–1117.
Siuly, Li, Y., Wu, J., and Yang, J. Developing a Logistic Regression Model with Cross-Correlation for Motor Imagery Signal Recognition, In: IEEE/ICME International Conference on Complex Medical Engineering (ICME 2011), 22–25 May 2011, Harbin, China, pp. 502–507.
Shen, C.P., Chen, C.C., Hsieh, S.L., Chen, W.H., Chen, J.M., Chen, C.M., Lai, F., and Chiu, M. J. High-performance seizure detection system using a wavelet-approximate entropy-fSVM cascade with clinical validation, Clinical EEG and Neuroscience DOI:10.1177/1550059413483451, 2013.
Ripley, B. Pattern recognition and neural networks, Cambridge: Cambridge university press, 1996.
Siuly, Li, Y., Wen, P. EEG signal classification based on simple random sampling technique with least square support vector machines, International journal of Biomedical Engineering and Technology 7(4) (2011a) 390–409.
Siuly, Li, Y., Wen, P. Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain computer interface, Computer Methods and programs in Biomedicine 113 (3) (2014a) 767–780.
Siuly and Li, Y. Discriminating the brain activities for brain–computer interface applications through the optimal allocation-based approach, Neural Computing & Applications DOI 10.1007/s00521-014-1753-3, 2014c.

Suykens, J.A.K., and Vandewalle, J. Least Square Support Vector Machine classifier, Neural Processing Letters 9 (3) 293–300.

Suykens, J.A.K., Gestel, T.V., Brabanter, J.D., Moor, B.D., Vandewalle, J. Least Square Support Vector Machine, World Scientific, Singapore, 2002.

Sengur, A. Multiclass least-square support vector machines for analog modulation classification, Expert System with Applications 36(2009) 6681–6685.

Siuly, H. Wang and Y. Zhang (2016), 'Detection of motor imagery EEG signal employing Naive Bayes based learning process', *Measurement* 86, 148–158.

Siuly and Y. Li, (2014b), 'A novel statistical framework for multiclass EEG signal classification', *Engineering Applications of Artificial Intelligence*, Vol. 34, pp. 154–167.

Siuly and Y. Li, (2015), 'Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification', *Computer Methods and programs in Biomedicine, Computer Methods and programs in Biomedicine*, Vol. 119, pp. 29–42.

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 20 (4) *pp.* 526–538.

Siuly, Li, Y., and Wen, P. Clustering technique-based least square support vector machine for EEG signal classification, Computer Methods and Programs in Biomedicine 104 (3) (2011b) 358–372.

Siuly, Li, Y., and Wen, P. Identification of Motor Imagery Tasks through CC-LR Algorithm in Brain Computer Interface, International Journal of Bioinformatics Research and Applications 9 (2) (2013)156–172.

Siuly, Li, Y., Wen, P. Analysis and classification of EEG signals using a hybrid clustering technique, In: The proceedings of the 2010 IEEE/ICME International Conference on Complex Medical Engineering (CME2010), 2010, pp. 34–39.

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

Turk, M., and Pentl, A. Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1991) 71–86.

Ubeyli, E.D. Decision support systems for time-varying biomedical signals: EEG signals classification, Expert Systems with Applications 36 (2009) 2275–2284.

Ubeyli, E.D. Analysis of EEG signals by combining eigenvector methods and multiclass support vector machines, Computer in Biology and Medicine 38 (2008)14–22.

Xing, Y., Wu, X, Xu, Z. Multiclass least squares auto-correlation wavelet support vector machine, ICIC Express Letters 2 (4) (2008) 345–350.

Youping, F., Yungping, C., Wansheng, S., Yu, L. Multi-classification algorithm and its realization based on least square support vector machine algorithm, Journal of Systems Engineering and Electronics 16 (4) (2005) 901–907.

Zhang, H. The optimality of naive bayes, The international flairs conference 17 (2004).

Z Distribution Table, http://ci.columbia.edu/ci/premba_test/c0331/s6/z_distribution.html.

Part III
Methods for Identifying Mental States in Brain Computer Interface Systems

Chapter 8
Cross-Correlation Aided Logistic Regression Model for the Identification of Motor Imagery EEG Signals in BCI Applications

One crucial and challenging issue in BCI systems is the identification of motor imagery (MI) task based EEG signals in the biomedical engineering research area. Although BCI techniques have been developing quickly in recent decades, there remains a number of unsolved problems such as the improvement of MI signal classification. The translation of brain activities into EEG signals in BCI systems requires a robust and accurate classification to develop a communication system for motor disabled people. In BCIs, MI tasks generate brain activities, which are generally measured by EEG signals. The aim of this chapter is to introduce a method for the extraction of discriminatory information from the MI based EEG signals for BCI applications and then identify MI tasks based on the extracted features through the classification process. This chapter proposes a new approach, the 'Cross-correlation aided logistic regression model' called "CC-LR" for the identification of MI tasks. In this methodology, the cross-correlation (CC) technique is used for extracting information from MI EEG signals and the logistic regression (LR) method is employed to estimate the class labels of MI tasks using the extracted information as the input. This study investigates the performance of the CC technique for a randomly selected reference signal and also investigates the LR for efficient classification of the cross-correlated features. The proposed method is tested on two benchmark datasets, IVa and IVb of BCI Competition III and the performance is evaluated through a cross-validation procedure.

8.1 Definition of Motor Imagery (MI)

Motor imagery (MI) can be defined as a dynamic mental state during which an individual mentally simulates a given action. This type of phenomenal experience implies that the subject feels himself performing the action (Decety 1996). In MI, subjects are instructed to imagine themselves performing a specific motor action (for example: a hand or foot movement) without an overt motor output

S. Siuly et al., *EEG Signal Analysis and Classification*,
Health Information Science, DOI 10.1007/978-3-319-47653-7_8

(Kayikcioglu and Aydemir 2010; Thomas et al. 2009). MI produces effects on the brain rhythm in the sensory-motor cortex which are the same as those associated with the real executed movement (Pfurtscheller et al. 2006). Prominent changes in brain rhythms are observed in the sensory-motor area of the brain during the planning and the execution of movements. Different mental states lead to changes in the electro-physiological signals in the brain. To use a BCI, a subject must have a sensor of some kind on the head, and must voluntarily choose to perform certain mental tasks to accomplish goals. There are various acquisition techniques for capturing MI brain activities. Among these techniques, EEG is the most studied measure of potential for non-invasive BCI designs, mainly due to its excellent temporal resolution, non-invasiveness, usability, and low set-up cost (Kayikcioglu and Aydemir 2010; Blankertz et al. 2008; Grosse-Wentrup et al. 2009). In pattern recognition, each MI task is treated as a class.

8.2 Importance of MI Identification in BCI Systems

The ability to communicate with the outside world is one of the most indispensable assets that humans possess. 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 and Aydemir 2010). Neurological diseases can disrupt the neuromuscular channels through which the brain communicates with its environment and exerts control. Therefore, it is impossible for people who are motor disabled, to live and meet their daily needs without external help. The vital object of the BCI is to provide humans with an alternative communication channel allowing the direct transmission of messages from the brain by analyzing the brain's mental activities. The BCI is a well-known emerging technology with which people are can communicate with their environment and control prosthetic or other external devices 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.

In the BCI development, the identification of MI tasks is a key issue as users (e.g. motor disabled people) produce different brain activity patterns for different MI tasks that are identified by a system and are then translated into commands (Wang et al. 2004). These commands are used as feedback for motor-disabled patients to communicate with external environments. 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 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; Siuly and Li 2012). Figure 8.1 shows how BCIs convert user *intentions* or *thoughts* into control signals to establish a direct communication channel between the human brain and output devices. This figure depicts the basic structure of BCI

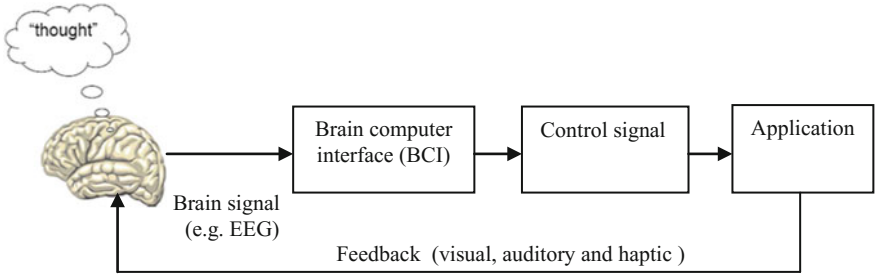


Fig. 8.1 Fundamental structure of MI EEG-based BCI systems

technologies, how a MI-based EEG signal of a *thought* passes on to a BCI system, and how the BCI system processes those signals into a control signal for a user application. Thus, 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 as MI-based BCI system translates a subject's motor intention into a command signal through real-time detection of motor imagery states.

Hence, a MI-based BCI provides a promising means of control and communication for people who are suffering from motor disabilities (Wolpaw et al. 2002). Therefore, the recognition of MI tasks is crucial for the BCI development that generates control signals.

8.3 Motivation to Use Cross-Correlation in the MI Classification

Identification of MI-based EEG signals is a difficult problem in BCI systems as EEG data are naturally nonstationary, highly noisy and contaminated with artefacts (Wu et al. 2008; Long et al. 2010). Classification techniques help to predict and identify class labels of a subject's mental state by extracting useful information from the highly multivariate recordings of the brain activities. To identify a subject's intentions from MI-based EEG data, the most important thing is to extract discriminative relevant features 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 performance and reliability of a recognition system greatly depends on the features and the classification algorithm employed. The current study proposes an innovative algorithm where a CC technique is developed for feature extraction for EEG-based BCI systems.

As discussed in Chap. 3, several feature extraction methods for MI data have been applied in BCI applications such as autoregressive (AR) (Schlogl et al. 2002; Pfurtscheller et al. 1998; Burke et al. 2005; Guger et al. 2001; Jansen et al. 1981), fast Fourier transform (FFT) (Polat and Gunes 2007), common spatial patterns (CSP) (Blanchard and Blankertz 2004; Lemm et al. 2005), spatio-spectral patterns (Wu et al. 2008) and wavelet coefficients (Qin and He 2005; Ting et al. 2008; Liao et al. 2013), phase locking value (PLV) and spectral coherence (Gysels and Celka 2004; Park et al. 2013; Lachaux et al. 1999; Quiroga 2009) and power spectral density (PSD) (Park et al. 2013). From the mentioned literature, we understand that the AR model has a superior resolution capability for the spectral estimation from which features are obtained. But the major problem of this model is that the model order is not known and has to be determined via order selection criteria.

Order selection can give inconsistent results if the data set is small. On the other hand, the FFT is a transformation that can be used in time domain and frequency domain to extract features from EEG data. However, in the Fourier algorithm, the estimation of frequencies is sensitive to noise. This method requires a stationary signal for analyzing the component of the signal where the EEG is a nonstationary signal. Furthermore, the input signal repeats periodically and the periodic length is equal to the length of the actual input. If the true signal is not periodic or if the assumed periodic length is not correct, both the amplitude and position of a frequency measurement will be inaccurate. Although the CSP is a popular method in BCI applications, it is highly non-robust to noise and outliers and often over-fits with small training sets. Moreover, despite being a spatial filter, the CSP completely ignores the spatial location of the EEG electrodes. In the spatio-spectral patterns, selecting their regularization parameters for reliable classification is not simple.

Wavelet transform (WT) is the improved version of Fourier transform that can capture transient features and localizes them in both the time and frequency domains. But, sometimes it is hard to select a suitable wavelet and determine an appropriate number of decomposition levels in the analysis of EEG signals. Detecting phase locking between two distant brainwave recordings (such as EEG, MEG, and intracranial) is not straightforward in the PLV. The PLV may be the measure of choice if one wants to focus on phase relationship. The PLV features can be influenced by the changes in electrode locations. On the other hand, the coherence may be very useful for studying interactions if one is interested in a particular frequency band. The PSD features are very sensitive to the changes in EEG electrode location, and another pitfall of the method is its potential instability.

In response to these problems, this study introduces a cross-correlation (CC) method for feature extraction from MI EEG data. There are strong grounds for using a CC technique for feature extraction in this area. EEGs record brain activities as a multi-channel time series from multiple electrodes placed on the scalp of a subject. The recorded multi-channel EEG signals are highly correlated, and the different signals from different scalp sites do not provide the same amount of discriminative information (Meng et al. 2009). The anatomical differences of the brains of the subjects could affect the correlation of signals. These signals are also typically very noisy and are not directly usable in BCI applications. The CC is a

very powerful technique to identify the relationship between the EEG signals from two different electrodes and is also able to provide discriminative information about those signals. This method can measure the degree of similarities between two signals with respect to time (Chandaka et al. 2009).

This technique produces a new signal called cross-correlogram (mathematically called cross-correlation sequence) using two signals. If the two EEG signals have the same rhythm, the pick of the cross-correlogram curve will appear in the centre. 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 CC technique works better for the feature extraction in MI EEG data.

8.4 Theoretical Background

8.4.1 Cross-Correlation Technique

In signal processing, cross-correlation (Chandaka et al. 2009; Siuly et al. 2013, 2016) 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 the identical time samples of each waveform. Cross-correlation is commonly used to search a long duration signal for a shorter, known feature, and 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 CC 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 CC as a function of lag m is defined as (Chandaka et al. 2009; Dutta et al. 2010; Siuly et al. 2013, 2014)

$$\begin{aligned} R_{xy}[m] &= \sum_{i=0}^{N-|m|-1} x[i]y[i-m]; \\ m &= -(N-1), -(N-2), \ldots 0, 1, 2, 3, \ldots, (N-2), (N-1) \end{aligned} \tag{8.1}$$

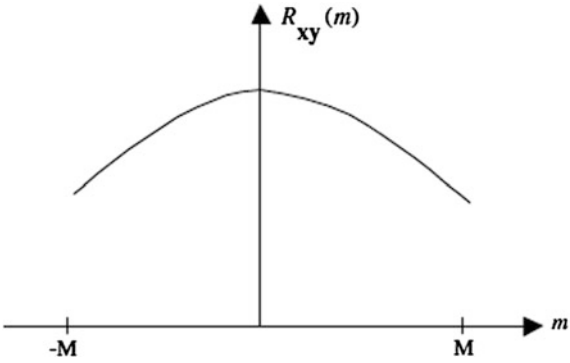


Fig. 8.2 Example of a typical cross-correlogram (Chandaka et al. 2009)

Here, index m represents time-shift parameters known as lag where $m = -(N-1), -(N-2), \ldots -1, 0, 1, \ldots, (N-2), (N-1)$, 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 CC sequence has $2M-1$ samples. If x and y are not the same length, for example, x and y have N and M numbers of samples, respectively, and if $N > M$, the resultant CC sequence has $(2N-1)$ number of samples. The shorter vector, here y, is zero-padded to the length of the longer vector, x.

Figure 8.2 shows an example of a typical cross-correlogram. The peak of the cross-correlogram represents the offset. If the two signals are identical in the same rhythm, the peak of the cross-correlogram curve will appear exactly in the centre. If they are offset from each other, the peak will occur as an offset from centre. In CC 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 synchronous signals 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 indicates less stability in the relationship.

8.4.2 Logistic Regression Model

The logistic regression (LR) model is a workhorse of statistics, and is increasingly popular in machine learning due to its similarity to the support vector machine (SVM) (Liao and Chin 2007; Xie et al. 2008; Caesarendra et al. 2010; Mrowski et al. 2009; Ryali et al. 2010). The LR fits a separating hyper plane that is a linear function of input features between two classes. The goal of the LR in Eq. (8.2) is to estimate the hyper plane that accurately predicts the 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 type.

Suppose x_1, x_2, ……, x_n are vectors of input features and y is its class label either, 0 or 1. Here x_1, x_2, ……, 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 and Ercelebi 2005; Siuly et al. 2013, 2014)

$$\mathrm{P}(y = 1|x_1, x_2, \ldots\ldots, x_n) = \pi = \frac{e^{\beta_0 + \sum_{i=1}^{n} \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^{n} \beta_i x_i}} \tag{8.2}$$

Here, π is a conditional probability of the form of P(y = 1| x_1, x_2, ……, 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, ……, x_n) can be calculated as $1 - \pi = 1 -$ P(y = 1| x_1, x_2, ……, x_n). In Eq. (8.2), β_0 is an intercept and $\beta_1, \beta_2, \ldots, \beta_n$ are the regression coefficients, related to independent variables, x_1, x_2, …, x_n. These parameters are estimated by a maximum likelihood estimation (MLE) (Hosmer and Lemeshow 1989). The above cost function results in a solution that accurately predicts class labels of a new example.

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

$$\mathrm{logit}(\pi) = \log_e\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots\cdots + \beta_n x_n = \beta_0 + \sum_{i=1}^{n} \beta_i x_i \tag{8.3}$$

In Eq. (8.3), logit(π) is a linear combination of independent variables, x_1, x_2, …, x_n, and regression coefficients, $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$. The LR applies to the MLE after transforming the dependent variables into a logit variable in order to calculate parameters, β_0, $\beta_1, \beta_2, \ldots, \beta_n$. A linear relationship is not assumed in general between the independent and dependent variables nor does it require normally distributed independent variables (Subasi and Ercelebi 2005).

8.5 Cross-Correlation Aided Logistic Regression Model

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.

8.5.1 Feature Extraction Using the CC Technique

Feature extraction plays an important role in pulling out special patterns (features) from the original data for reliable classification. In this research, we use the CC technique for feature extraction, which 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 the class of a subject as there are no specific requirements for selecting a reference signal in the CC analysis. In Eq. (8.1), $x[i]$ is considered as the reference signal and $y[i]$ is one of any other signal named as a non-reference signal. Second, 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. Third, six statistical features, *mean, median, mode, standard deviation, maximum* and *minimum,* are calculated from each CC 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 values (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 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 of the nature of the data (shows the "middle value"), especially when combined with measurements describing 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 are spread around the typical values. *Standard deviation* is the most popular measure for variability. It 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 to be the most valuable parameters for representing the distribution of the MI EEG signals and for representing brain activities as a whole.

8.5.2 MI Tasks Signal Classification by Logistic Regression (LR)

This study employs a LR to predict the probability of two categories of MI tasks from the EEG datasets. The LR is a standard method for the identification of binary outcomes. As a result, it has been applied to the MI tasks classification in this study. In this work, 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 Eq. (8.2).

8.6 Results and Discussions

This study uses two benchmark datasets, IVa and IVb from BCI Competition III, to evaluate the efficacy of the proposed approach (see description in Sects. 3.3.1.2 and 3.3.1.3 in Chap. 3). This section presents the experimental results of the proposed algorithm on these two benchmark EEG datasets, and also provides a comparison of the present method with two recently reported methods for dataset IVa. As we did not find any research reports for dataset IVb in the literature, we could not compare the experimental results with other methods. In both datasets, each subject 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 [version 7.7, (R2008b)] for all mathematical calculations and PASW (Predictive Analytics SoftWare) Statistics 18 for the LR model.

8.6.1 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. 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 8.1 presents the information about the structure of the dataset. As shown in Table 8.1, 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

Table 8.1 The information of original data for BCI Competition III, dataset IVa

Subject	Size of data with two classes (RH and RF)	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

training trials in our experiments as the proposed algorithm requires a class label at each data sequence.

Figure 8.3 shows the typical RH and RF MI signals for each of the five subjects in 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 previously, 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 CC sequences are obtained for this class. Again, in the RF class of the same

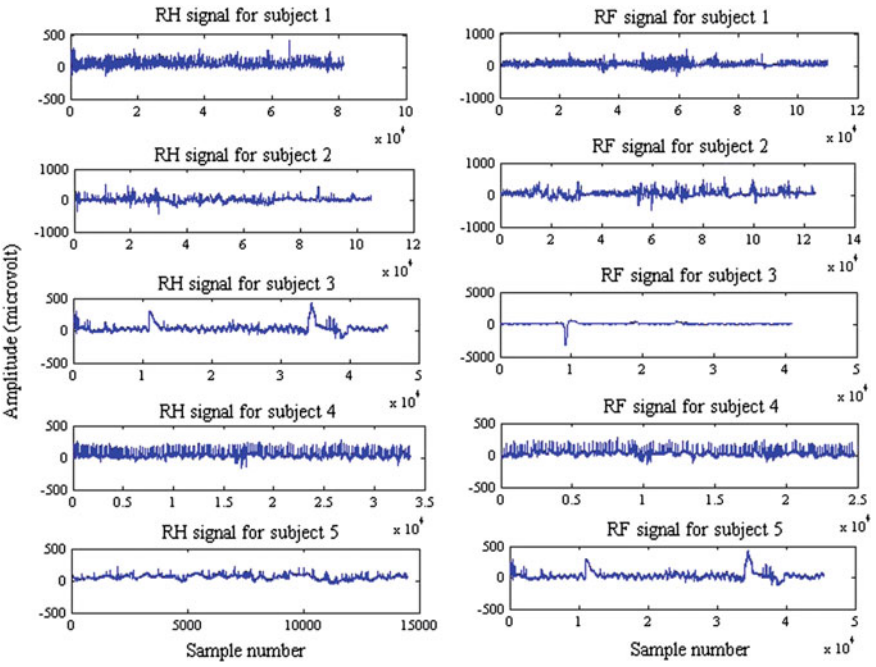


Fig. 8.3 The typical signals of the RH and the RF MI tasks for each subject in dataset IVa

subject, the reference channel data is cross-correlated with each of 118 channels' data and produces 118 CC sequences. Thus, a total of 235 CC sequences is obtained for the two-class MI data for each subject.

Figure 8.4 presents the results of CC sequences called cross-correlograms for the RH and the RF MI data of Subjects 1, 2, 3, 4 and 5, respectively. It is important to note that the cross-correlogram or CC sequences (R_{xy}) are calculated using Eq. (8.1) for each lag. From this figure, one can see that, in most of the cases, the shapes of the two curves for a subject are not exactly the same, indicating their statistical independency. This means, that there is more of a chance of achieving better separation. From each cross-correlogram of a subject in dataset IVa, the six statistical features as described in Sect. 8.5.1 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.

According to the threefold cross-validation procedure, the 235 feature vectors of six dimensions are divided into three subsets containing an equal number of

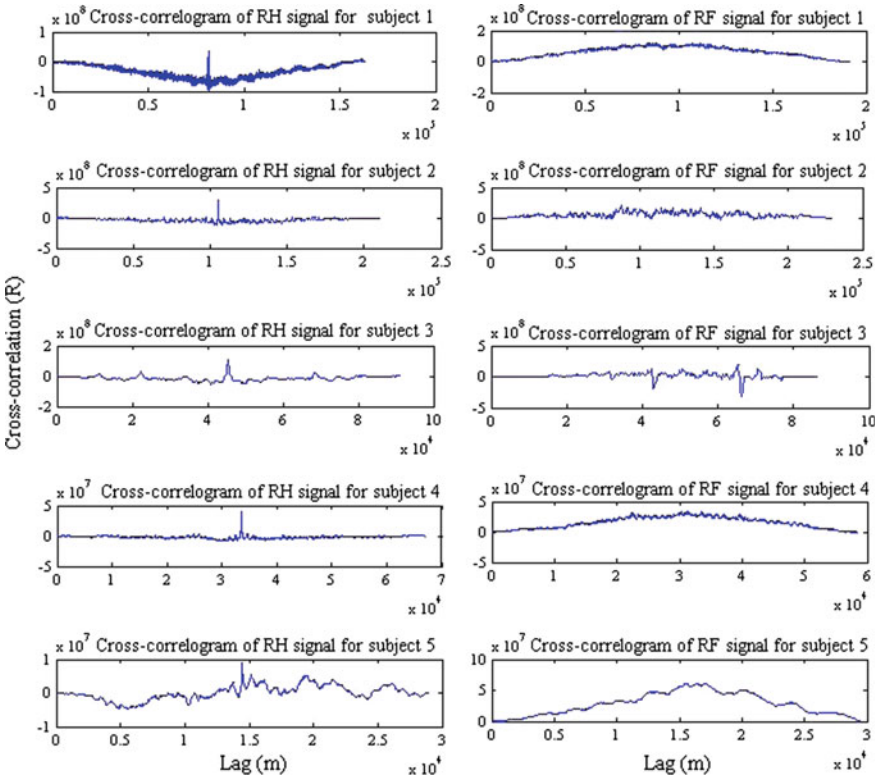


Fig. 8.4 The typical cross-correlograms for the RH and the RF MI signals of each subject in dataset IVa

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 evaluated across all three folds. This process is called the cross-validation accuracy.

We utilize the training set and the testing set to the LR model in Eq. (8.2) for estimating the probability of the dependent variable. In Eq. (8.2), 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 treated as 1 for dataset IVa. The six statistical features are considered as six independent variables in Eq. (8.2) 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 $\beta_0, \beta_1, \beta_2, \ldots, \beta_6$ are calculated using the MLE.

Table 8.2 shows the classification accuracies and standard deviations for each of the three folds. It also presents the average classification accuracy for all five subjects in dataset IVa. Using the threefold cross-validation procedure, the proposed CC-LR algorithm produces a cross-validation accuracy of 96.57, 82.9, 100, 96.6 and 82.9% for Subjects 1, 2, 3, 4 and 5, respectively. The standard deviations of the three folds of Subjects 1, 2, 3, 4 and 5, are 2.65, 5.35, 0.0, 3.21 and 0.69, respectively. One can see that there are no significant differences among the threefold accuracies of a subject, indicating the consistency of the proposed method. As shown in Table 8.2, 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 was 91.79% and 2.38, respectively.

To provide information on how the MI tasks are predicted and classified in each fold, we describe the output for the onefold of Subject 1 in dataset IVa, as an example. Table 8.3 displays a confusion matrix to show the classification results of the LR classifier for the onefold of Subject 1 to dataset IVa. From the confusion matrix, it is observed that in the onefold of Subject 1, two values of the RH class are

Table 8.2 The threefold cross-validation results by the CC-LR method on testing set for dataset IVa of BCI Competition III

Subject	Threefold cross-validation accuracy and their standard deviation (%)				
	Onefold	Twofold	Threefold	Cross-validation accuracy (average of three folds)	Standard deviation for 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

Table 8.3 Confusion matrix for the onefold of Subject 1 from dataset IVa

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

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 onefold for Subject 1.

Figure 8.5 depicts the typical scenery of a classification plot on how observed values are predicted and classified by the LR model in the onefold of Subject 1 for the dataset. In this figure, the observed group values of the two MI tasks and the predicted probability values obtained from Eq. (8.2) are plotted on the *X*-axis. The frequency values of the observed and predicted probability are plotted on the *Y*-axis. As mentioned previously, in dataset IVa, 0 is indicated as the RH class and 1 as the RF class.

From Fig. 8.5, 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 8.3 also shows similar results. The reflection of the confusion matrix is shown in Fig. 8.5.

A comparison of the proposed algorithm with two recent reported algorithms is shown in Table 8.4. The highest classification accuracy from the three algorithms is highlighted for each subject and their average. The proposed CC-LR algorithm provides better classification accuracies than the other two recently reported algorithms in three out of the five subjects. From Table 8.4, it is seen that the proposed

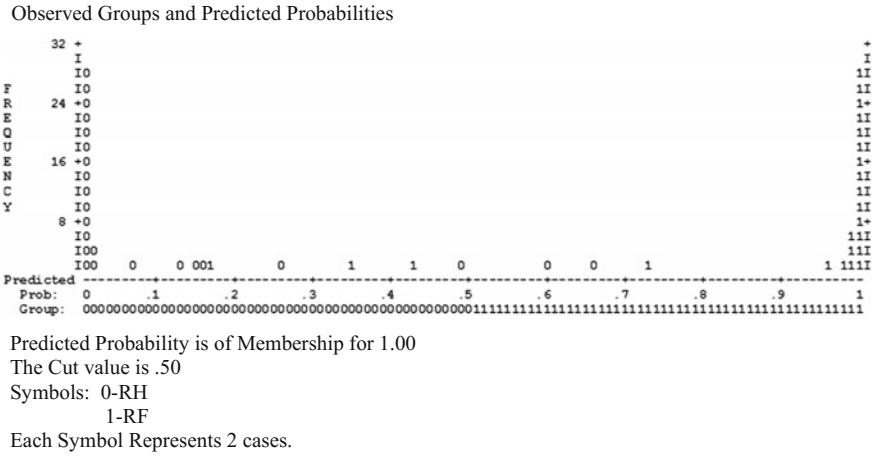


Fig. 8.5 Illustration of classification plots for the onefold of Subject 1 in dataset IVa

Table 8.4 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. 2011)
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

algorithm produces the highest classification accuracy of 96.57% for Subject 1, 100.0% for Subject 3 and 96.6% for Subject 4 compared to the R-CSP with aggregation (Lu et al. 2010) and CT-LS-SVM (Siuly et al. 2011) algorithms. We obtained 82.9% accuracy with the proposed method for Subject 2, which is slightly 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 does better than the two recently reported algorithms in the MI tasks signal classification.

For further justification, we also compare the proposed method with four other existing methods reported in the literature for the dataset shown in Fig. 8.6: 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 Fig. 8.6, compared to the four existing methods, the proposed algorithm yields the best accuracy for Subjects 1, 3 and 4.

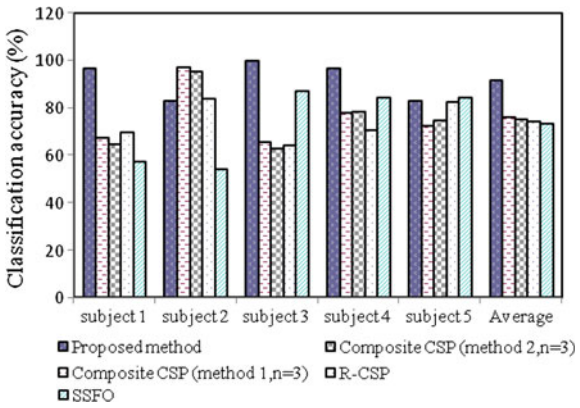


Fig. 8.6 Performance comparisons of four other existing methods (reported in the literature) with the proposed CC-LR method

The composite CSP (method 1, $n = 3$) achieved a better classification accuracy than our method for Subject 2, and a classification accuracy similar to the other methods for Subject 5. It is observed from Fig. 8.6 that the highest average classification accuracy is obtained by our proposed algorithm.

8.6.2 *Classification Results for Dataset IVb*

As previously mentioned, 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 the class labels of each observation. The original data size of the 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 the 117 channels of the RF MI class to create 117 CC sequences. The reference channel is again cross-correlated with each of 118 channels of LH MI class to produce 118 CC sequences. Therefore, a total of 235 CC sequences is obtained from the dataset.

Figure 8.7 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 the two waveforms are not the same, so there is a greater chance of achieving a better separation. The six statistical features mentioned in the previous section 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 threefold cross-validation process. The feature vector sets

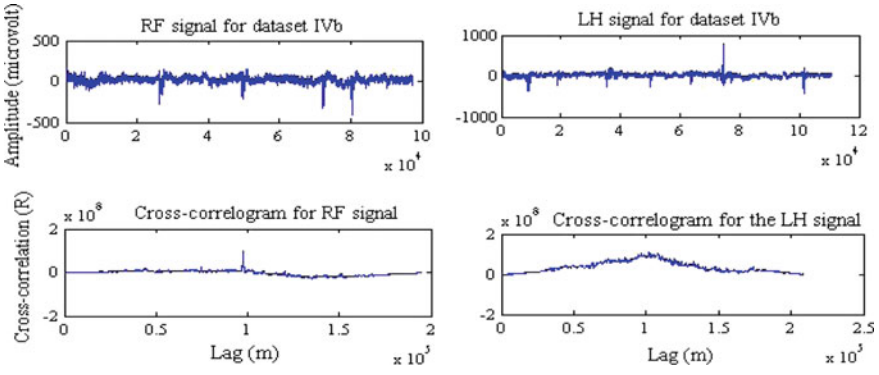


Fig. 8.7 The typical signals and cross-correlograms for the RF and the LH MI signals from dataset IVb

Table 8.5 The threefold cross-validation results by the proposed method on testing set for dataset IVb of BCI Competition III

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

Table 8.6 Confusion matrix for the onefold of dataset IVb

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

obtained are used as input variables to the LR classifier for the prediction and classification of the EEG-based MI tasks.

Table 8.5 gives the classification accuracy for each of the three folds, and the average classification accuracies and standard deviations for all three folds from dataset IVb. Table 8.5 shows that the LR model classifies the RF and LH MI data with the accuracy of 85.9, 94.9 and 100% for the onefold, the twofold and the threefold, respectively. The average cross-validation accuracy and standard deviation for the three folds is 93.6% and 7.14, respectively.

Table 8.6 provides the prediction information about the classification for the onefold of dataset IVb. From this table, we see that seven values of the RF class are

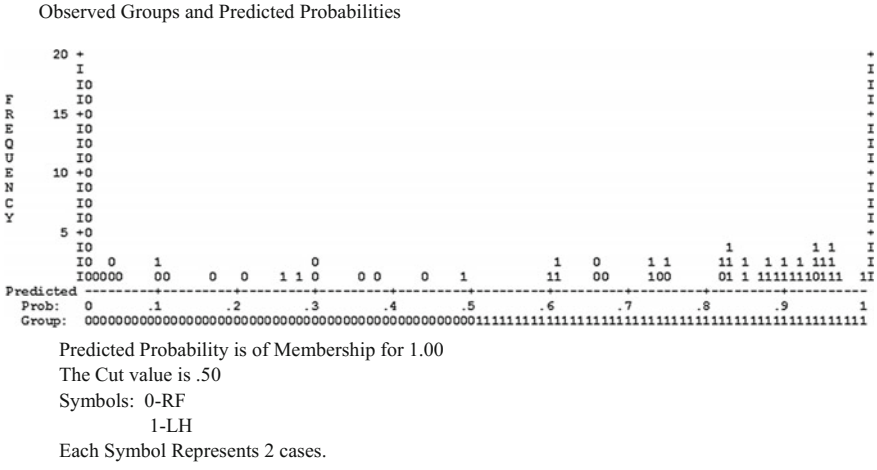


Fig. 8.8 Classification plot for the onefold for dataset IVb

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 85.9% for the onefold of this dataset. We can also see the mirror image of this confusion matrix in Fig. 8.8.

To visually describe the classification scenery of the LR classifier, Fig. 8.8 exhibits the classification plot for the onefold of dataset IVb, as an example. This figure clearly explains the confusion matrix reports presented in Table 8.6. In this figure, 0 denotes the RF class and 1 denotes the LH class.

As shown in Fig. 8.8, four values of the LH class are misclassified as the RF class, and seven values from the RF class are misclassified to the LH class in the onefold 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 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 offers a great potential for performance improvement.

8.7 Conclusions and Recommendations

This study develops an algorithm for the classification of the MI tasks in BCI applications combining the CC and logistic regression methods. The performance of the proposed algorithm is measured in terms of classification accuracy using a threefold 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. 2011) for dataset IVa. To further validate the efficacy of the proposed algorithm, it 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 MI EEG signals, randomly considering the reference signal of the 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 system. The extracted feature set was not able to produce higher performance for the competition with the existing methods. This is why, in the next chapter, we modify the CC-LR algorithm with the C3 reference channel employing the diverse features to the proposed algorithm.

References

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.

Chandaka, S., Chatterjee, A. and Munshi, S. (2009) 'Cross-correlation aided support vector machine classifier for classification of EEG signals', *Expert System with Applications,* Vol. 36, pp. 1329–1336.

Decety, J. (1996). Do executed and imagined movements share the same central structures? *Cognitive Brain Research*, 3, 87–93.

De Veaux, R. D., Velleman, P.F. and Bock, D.E. (2008) Intro Stats (3rd edition), Pearson Addison Wesley, Boston.

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.

Hosmer, D.W. and Lemeshow, S. (1989) *Applied logistic regression*, Wiley, New York.

Islam, M. N. (2004) *An introduction to statistics and probability*, 3rd ed., Mullick & brothers, Dhaka New Market, Dhaka-1205, pp. 160–161.

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.

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.

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.

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.

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.

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.

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.

Siuly, X. Yin, S. Hadjiloucas, Y. Zhang, (2016) 'Classification of THz pulse signals using two-dimensional cross-correlation feature extraction and non-linear classifiers', *Computer Methods and Programs in Biomedicine,127*, 64–82.

Siuly and Y. Li, (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, *Vol*. 20, no. 4, pp. 526–538.

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, Issue 3, pp. 358–372.

Siuly, Y. Li, and P. Wen, (2013) 'Identification of Motor Imagery Tasks through CC-LR Algorithm in Brain Computer Interface', *International Journal of Bioinformatics Research and Applications*, Vol.9, no. 2, pp. 156–172.

Siuly, Y. Li, and P. Wen, (2014) 'Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain computer interface', *Computer Methods and programs in Biomedicine*, Vol. 113, no. 3, pp. 767–780.

Subasi, A. and Ercelebi, E. (2005) 'Classification of EEG signals using neural network and logistic regression', *Computer Methods and Programs in Biomedicine*, Vol. 78, pp. 87–99.

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, pp 2730–2733.

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.

Wang, T., Deng, J. and He, B. (2004) 'Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns', *Clinical Neurophysiology,* Vol. 115, pp. 2744–2753.

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.

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.

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.

G. Pfurtscheller, C. Brunner, A. Schlogl and F. Lopes da Silva, "Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks," *Neuroimage*, vol. 31, no. 1, pp. 153–159, 2006.

T. Kayikcioglu and O. Aydemir, "A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data," *Pattern Recognition Letters*, vol. 31, pp. 1207–1215, 2010.

B. Blankertz, R. Tomioka, S. Lemm, M. Kawanable, and K.R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.

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

A. Schlogl, C. Neuper and G. Pfurtscheller, Estimating the mutual information of an EEG-based brain-computer interface, Biomed. Tech. (Berl) 47 (2002) 3–8.

G. Pfurtscheller, C. Neuper, A. Schlogl and K. Lugger, Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters, IEEE Transactions on Rehabilitation Engineering 6 (1998) 316–325.

D.P. Burke, S.P. Kelly, P. Chazal, R.B. Reilly and C. Finucane, A parametric feature extraction and classification strategy for brain-computer interfacing, IEEE Transactions on Neural Systems and Rehabilitation Engineering 13 (2005) 12–17.

C. Guger, A. Schlogl, C. Neuper, C. Walterspacher, D. Strein, T. Pfurtscheller and G. Pfurtscheller, Rapid prototyping of an EEG-based brain-computer interface (BCI), IEEE Transactions on Neural Systems and Rehabilitation Engineering 9(2001) 49–58.

B.H. Jansen,, J.R. Bourne, and J.W. Ward, Autoregressive Estimation of Short Segment Spectra for Computerized EEG Analysis, IEEE Transactions on Biomedical Engineering 28 (9) (1981) 630–637.

K. Polat and S. Gunes, Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform, Applied Mathematics and Computation 187 (2007) 1017–1026.

G. Blanchard and B. Blankertz, BCI competition 2003-Data set IIa: Spatial patterns of self-controlled brain rhythm modulations, IEEE Transactions on Biomedical Engineering 51 (2004) 1062–1066.

S. Lemm, B. Blankertz, G. Curio and K.R. Muller, Spatio-spatial filters for improved classification of single trial EEG, IEEE Transactions on Biomedical Engineering 52 (2005) 1541–1548.

W. Wu, X. Gao, B. Hong and S. Gao, Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL), IEEE Transactions on Neural Systems and Rehabilitation Engineering 55 (2008) 1733–1743.

L. Qin and B. He, A wavelet-based time-frequency analysis approach for classification of motor imagery for brain-computer interface applications, Journal of Neural Engineering 2 (2005) 65–72.

W. Ting, Y. Guo-Zheng, Y. Bang-Hua and S. Hong, EEG feature extraction based on wavelet packet decomposition for brain computer interface, Measurement 41(2008) 618–625.

K. Liao, M. Zhu and L. Ding, A new wavelet transform to sparsely represent cortical current densities for EEG/MEG inverse problems, Computer Methods and Programs in Biomedicine 111 (2013) 376–388.

E. Gysels, & P. Celka, Phase synchronization for the recognition of mental tasks in a brain-computer interface, IEEE Transactions on Neural Systems and Rehabilitation Engineering 12 (2004) 406–415.

S.A. Park, H. J. Hwang, J.H. Lim, J.H. Choi, H.K. Jung, C.H. Im, Evaluation of feature extraction methods for EEG-based brain–computer interfaces in terms of robustness to slight changes in electrode locations, Med Biol Eng Comput 51(2013) 571–579.

J.P. Lachaux, E. Rodriguez, J. Martinerie, F.J. Varela, Measuring phase synchrony in brain signals, Hum Brain Mapp 8 (1999) 194–208.

R. Q. Quiroga, Bivariable and Multivariable Analysis of EEG Signals, Book Chapter 4, pp. 109–120, 2009.

J. Meng, G. Liu, G. Huang and Xiangyang Zhu, "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, pp. 2290-2294.

G. M. Hieftje, R. I. Bystroff and Robert Lim, "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, no. 2, pp. 253–258, 1973.

S. Dutta, A. Chatterjee and S. Munshi, "An automated hierarchical gait pattern identification tool employing cross-correlation-based feature extraction and recurrent neural network based classification," *Expert systems*, vol. 26, no. 2, pp. 202–217, 2009.

Chapter 9
Modified CC-LR Algorithm for Identification of MI-Based EEG Signals

This chapter introduces a modified version of the CC-LR presented in Chap. 8. The CC-LR algorithm was proposed for the identification of MI signals where the 'Fp1' electrode signal was randomly considered as the reference signal in the CC technique. A set of features was extracted from each CC sequence to represent the MI task EEG signals. But that algorithm did not check whether the Fp1 electrode signal is suitable for providing informative measurements about motor tasks, and it did not investigate whether the considered feature set is the optimal choice or not. To alleviate these concerns, we present a modified version of the CC-LR algorithm which provides an insight into how to select a reference channel for the CC technique in EEG signals considering the anatomical structure of the human brain. We also investigate which features are superior to characterize the distribution of MI task EEG data. After these investigations, the study reaches a conclusion as to which electrode channel EEG data is appropriate for MI information and which feature set is best suited to characterize MI EEG signals.

9.1 Motivations

In our earlier study presented in Chap. 8, we developed the CC-LR algorithm for the classification of MI tasks for BCI applications but the performance was not satisfactory. This chapter develops a modified version of the CC-LR algorithm exploring an appropriately chosen reference signal and a suitable feature set that improves the performance. In the CC-LR algorithm (Siuly et al. 2013), we randomly considered the EEG signal from the electrode position Fp1 as a reference signal for the CC technique. In that algorithm, we extracted one feature set from each CC sequence to represent the MI task EEG data. However, the algorithm in (Siuly et al. 2013) could not provide a clear picture of the optimal feature set and the reference signal, as all EEG signals are not equal in providing informative measurements of motor tasks.

S. Siuly et al., *EEG Signal Analysis and Classification*,
Health Information Science, DOI 10.1007/978-3-319-47653-7_9

To overcome the problems of the CC-LR algorithm, this study proposes a modified version of the CC-LR method to classify two-class MI task EEG signals. To investigate which features are suitable for the representation of the MI signals, three statistical feature sets (described in Sect. 9.2) are extracted from each cross-correlation sequence of a subject, and are then evaluated. Finally, this study decides which feature set is best suited to characterizing 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. A popular *k*-fold cross-validation method is used to assess the performance of the proposed method in the MI task EEG signal classification. This cross-validation procedure is applied as a way of controlling the over-fitting of the data.

9.2 Modified CC-LR Methodology

This study proposes a potent method for MI task classification in the BCI development, which is shown in Fig. 9.1. This method provides an important framework for classifying the two-class MI task based EEG signals for BCI data. As we consider a pattern recognition system for the MI data, the EEG recognition procedure mainly involves the feature extraction and classification processes. As shown in Fig. 9.1, 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 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 to 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.

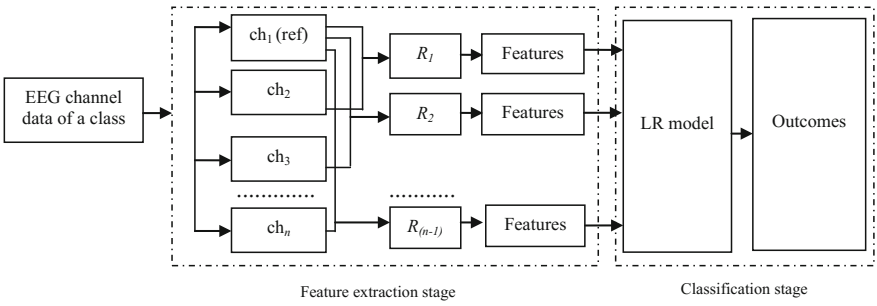


Fig. 9.1 Schematic diagram for the classification of the MI tasks based EEG signal in BCIs. Here *ch* channel, *R* cross-correlation sequence, *LR* logistic regression, *ref* reference channel

As 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 activity during the MI tasks.

Figure 9.2 gives information of the human brain structure and shows the locations of the 10-20 electrode placements. From Fig. 9.2b, it is obvious that C3 is in an important position of the motor cortex area, which is very responsive in the supply of MI information. This study considers the C3 channel as a reference channel for the CC method. As shown in Fig. 9.1, ch_1 (channel 1) is considered as the reference channel, as an example.

Step 2: In this step, the CC technique (*description provided in* Sect. 8.4.1 *in* Chap. 8) (Siuly et al. 2013, 2014, 2016) 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 Fig. 9.1, 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 chosen 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

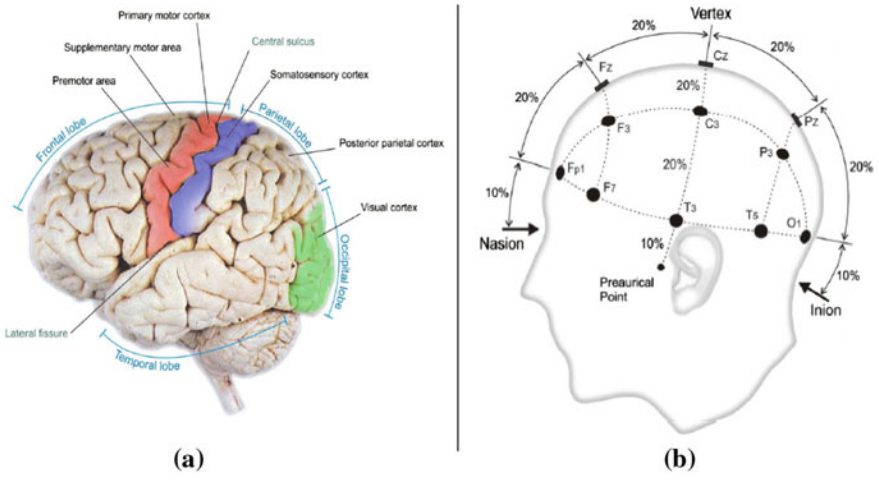


Fig. 9.2 **a** Structure of the human brain. **b** The International 10–20 electrode placement system

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 et al. 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 to 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 et al. 2008). *Kurtosis* measures 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 et al. 2008).

Step 4: As shown in Fig. 9.1, the classification stage is carried out with two phases described below:

- **Phase 1**: In this phase, the LR model (*description provided in* Sect. 8.4.2 *in* Chap. 8) 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 using *k*-fold cross-validation procedure.
- **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.

9.3 Experimental Evaluation and Discussion

This section presents an implementation procedure and the experimental results of the proposed algorithm for the two benchmark EEG datasets, IVa and IVb (described in Chap. 3), used in the BCI Competition III. These datasets are used for the evaluation of the proposed algorithm to classify different EEG signals during MI tasks. In this study, MATLAB software package (version7.7, (R2008b)) is used for the computation of the CC technique and Predictive Analytics Software (PASW) Statistics 18 is used for the LR model.

9.3.1 Implementation of the CC Technique for the Feature Extraction

Detecting features reliably is one of the most challenging tasks in BCIs. The features represent very weak signals from brain activities and are often corrupted by noise and various interfering artefacts of physiological and non-physiological origins. In this study, we develop a CC technique to extract the representative features from MI task EEG data. To reduce the dimensionality of a cross-correlation sequence, three different statistical feature sets are extracted over the resultant data (the reasons for choosing these features for are discussed in Sect. 6.2).

As mentioned in Chap. 3, dataset IVa of BCI Competition III contain MI task 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 in the 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 the experiments as the MI task EEG signals are naturally highly subject-specific according to the physical and mental tasks being performed. In each subject of dataset IVa, the 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 is obtained for the two-class MI data of each subject.

Figures 9.3, 9.4, 9.5, 9.6 and 9.7 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 Eq. (8.1) 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, indicating statistical independency. This means that there is a greater chance of achieving better separation.

From each cross-correlogram of a subject in dataset IVa, the three sets of statistical features as described in Sect. 9.2 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

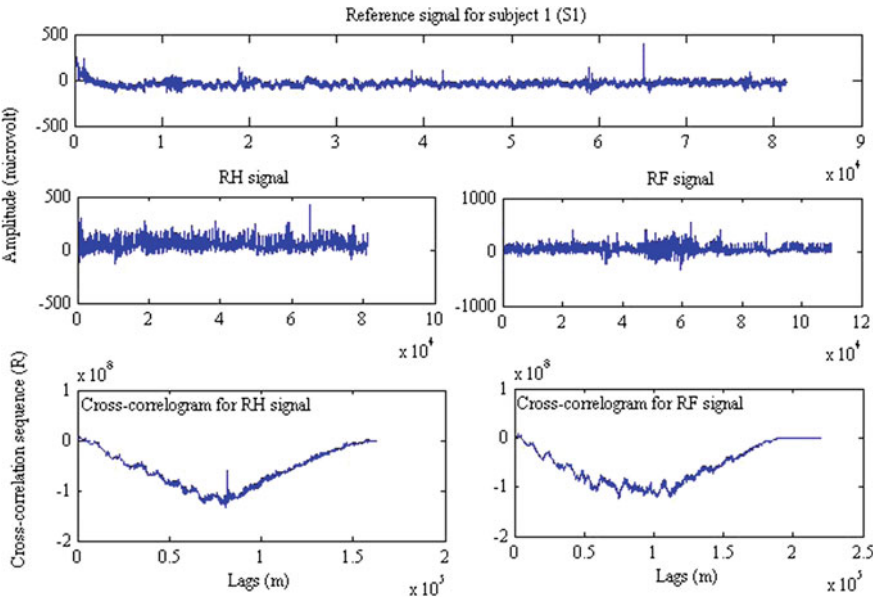


Fig. 9.3 The typical cross-correlograms for the RH and the RF MI tasks signals of S1 of dataset IVa

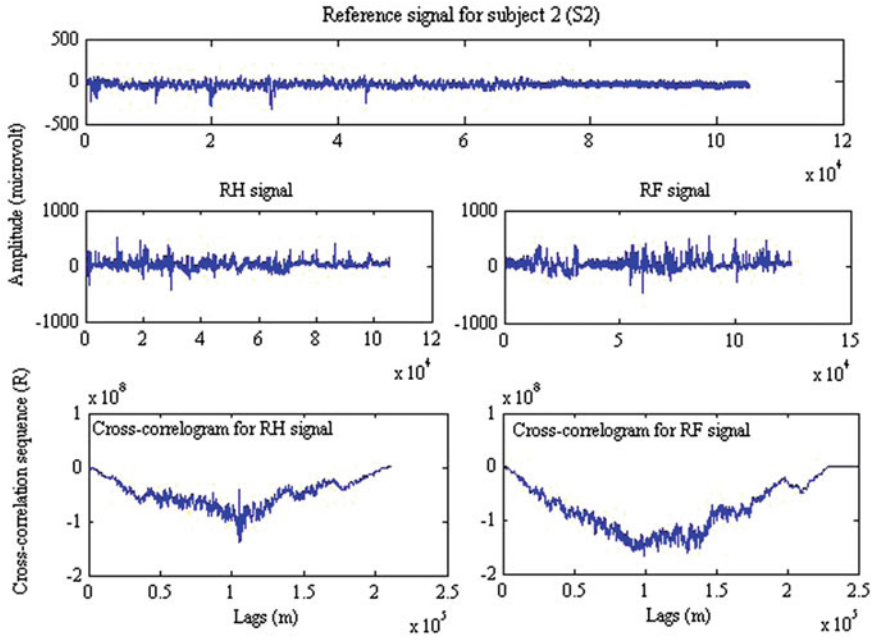


Fig. 9.4 The typical cross-correlograms for the RH and the RF MI tasks signals of S2 of dataset IVa

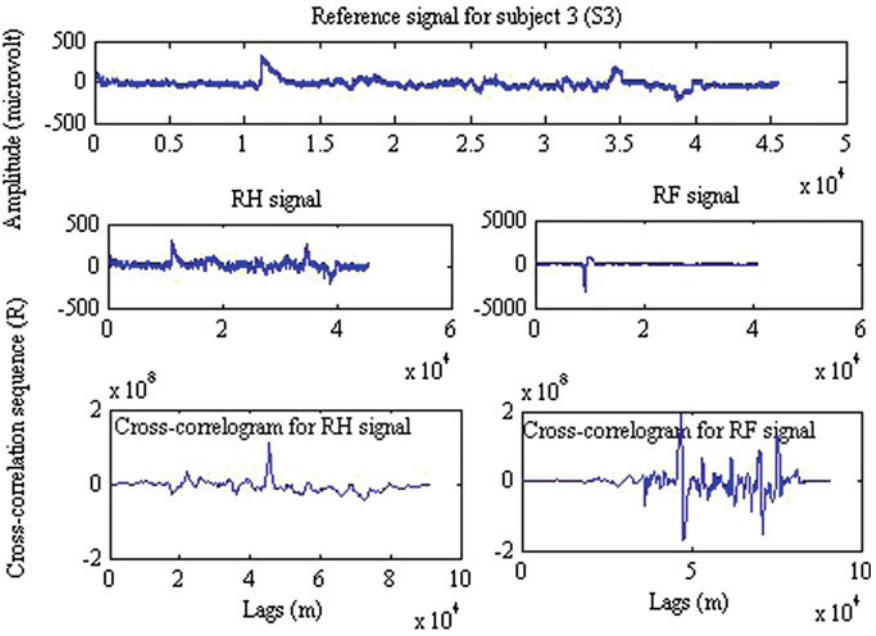


Fig. 9.5 The typical cross-correlograms for the RH and the RF MI tasks signals of S3 of dataset IVa

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 task EEG data for each subject in the dataset.

As discussed in Chap. 3, 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, the training data and testing data. We use the training data for our experiment as they include class labels with each observation but the testing data does not. The original data size of the training set is 210259×118.

For dataset IVb, the 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 9.8 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 of achieving better separation. The three statistical feature sets mentioned before (see Sect. 9.2) are calculated from each cross-correlogram.

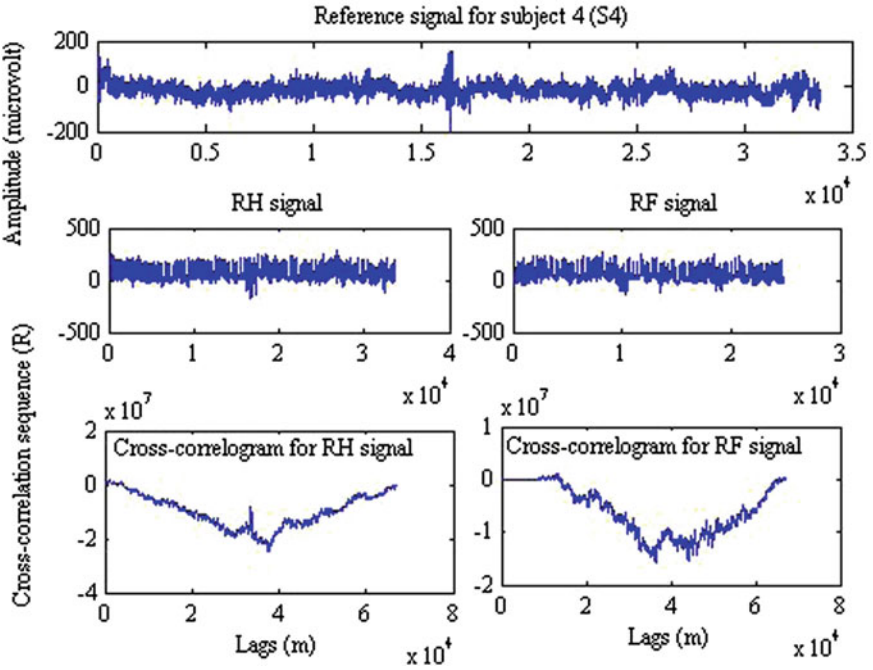


Fig. 9.6 The typical cross-correlograms for the RH and the RF MI tasks signals of S4 of dataset IVa

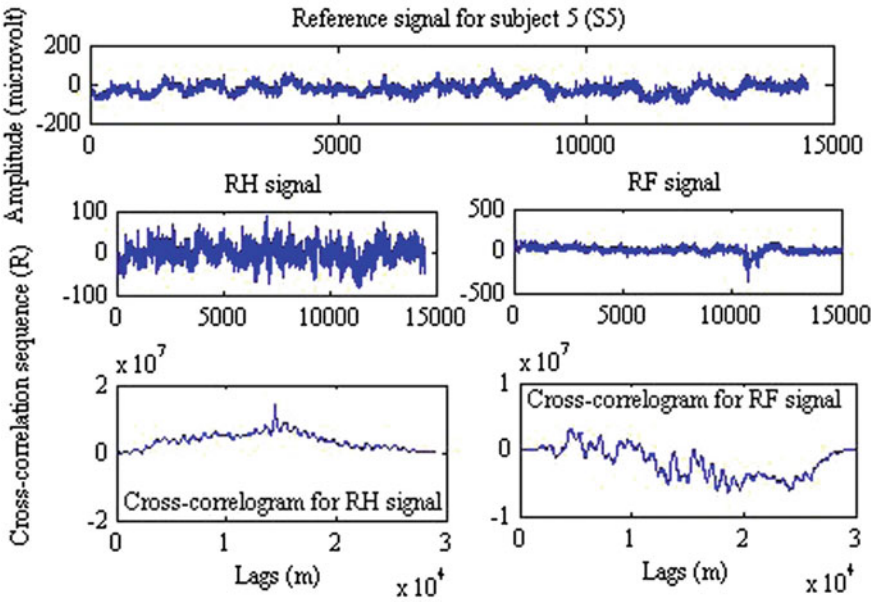


Fig. 9.7 The typical cross-correlograms for the RH and the RF MI tasks signals of S5 of dataset IVa

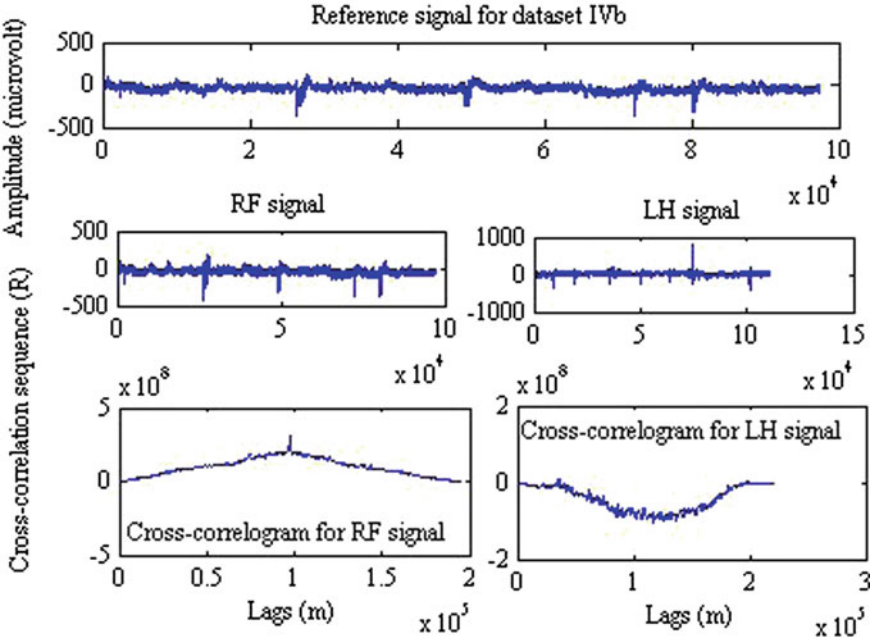


Fig. 9.8 The typical cross-correlograms for the RF and the LH MI tasks signals of dataset IVb

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 both datasets, the feature vectors of each of the three sets are segregated randomly as the training and testing sets through the *k*-fold cross-validation process (k = 3). The performance of the proposed approach is evaluated in terms of accuracy. The features are used as 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 task EEG data, and the results obtained.

9.3.2 MI Classification Results Testing Different Features

In this study, the LR model (Siuly et al. 2013, 2014) is employed to classify two-class MI EEG signals where the three feature sets (discussed in Sect. 9.2) are used separately as the input to the LR model. The expectation of this study is to find an appropriate feature set that is accurate for the two-class MI classification in the LR method. Feature vectors are extracted in such 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 threefold cross-validation method is utilized to calculate the classification accuracy on the testing dataset for each of the three feature sets.

Based on the threefold cross-validation procedure, a total of 235 feature vectors of each set from 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). For each time iteration (?), 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; named threefold 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 for the LR model in Eq. (8.2) to estimate 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 to be independent variables. In the case of both datasets, independent variables are considered for the three feature sets in Eq. (8.2) as follows.

For the two-feature set:

$$x_1 = \textit{mean}\ \text{values}; x_2 = \textit{standard deviation}\ \text{values}$$

For the four-feature set:

$x_1 = \textit{mean}$ values; $x_2 = \textit{standard deviation}$ values; $x_3 = \textit{skewness}$ values and $x_4 = \textit{kurtosis}$ values

For the six-feature set:

$x_1 = \textit{mean}$ values; $x_2 = \textit{standard deviation}$ values; $x_3 = \textit{skewness}$ values; $x_4 = \textit{kurtosis}$ values; $x_5 = \textit{maximum}$ values and $x_6 = \textit{minimum}$ values

Table 9.1 Cross-validation results with the proposed method on testing set for dataset IVa

Subject	Threefold 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 obtained classification results for the two, four and six-feature sets for each subject in dataset Iva are presented in Table 9.1. Table 9.1 provides the classification accuracy of the proposed algorithm through the threefold cross-validation accuracy. The results of each subject are reported in terms of mean ± standard deviation of the accuracy over a threefold 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 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 9.1, one can see that the average classification accuracy of the 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-feature set has been improved by 16.53% compared to the two-feature set by 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 by adding another two features, *maximum* and *minimum*. It can be concluded that more features can substantially improve the performance for all subjects. The results confirm that the six features; *mean*, *standard deviation*, *skewness*, *kurtosis*, *maximum* and *minimum* are potential characteristics for representing the original data for the MI task signal classification in

BCI applications. Table 9.1 also shows that there are no significant differences of the standard deviation in the threefold 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 threefolds) performance for all subjects. Figure 9.9a–e present the correct classification rate for the two-feature set, the four-feature set and the six-feature set for each of the threefolds and also overall

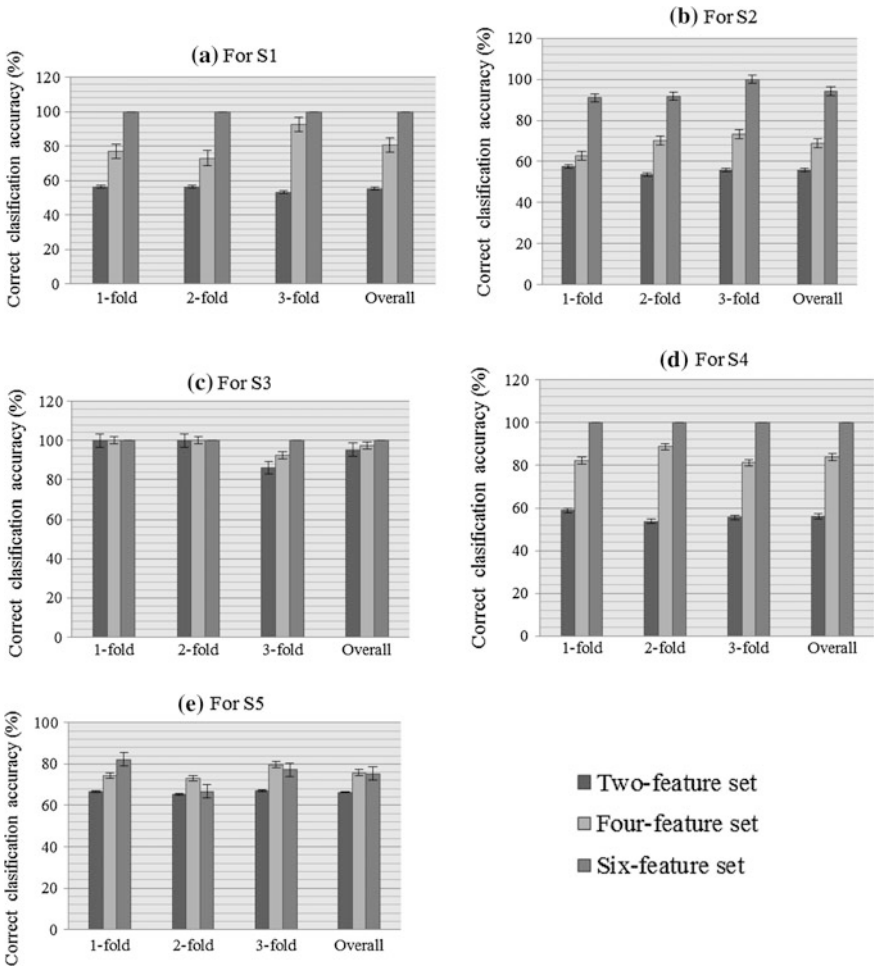


Fig. 9.9 Correct classification rate for the two-feature set, the four-feature set and the six-feature set in each of the three folds for: **a** S1 **b** S2 **c** S3 **d** S4 **e** S5 in dataset IVa (*Error bars* indicate the standard error)

Table 9.2 Cross-validation results by the proposed method on testing set for dataset IVb

Feature	Threefold 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

performance for all subjects in dataset IVa. From Fig. 9.9a–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, S3 and S4, respectively. The overall performances also

show the same results for those subjects. Figure 9.9e shows a slightly lower performance for the six-feature set in the twofold, the threefold and overall S5 than the other two-feature sets. Vertical lines on the top of the bar charts show standard error of the three folds and the overall standard error. Figure 9.9a–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 the other two features sets. Figure 9.9e illustrates a slightly higher standard error for the six-feature set in S5 than in the other two sets.

Table 9.2 gives the threefold cross-validation accuracy with their standard deviation for each of the three sets of features for dataset IVb. From this table, it can be observed that the LR model classifies the LH and RF MI task 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% on inclusion of 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% when adding another two features, *maximum* and *minimum*, into the four-feature set. This table also reports that there is no significant difference in 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.

Figure 9.10 illustrates detailed information for the classification rate for the two-feature set, the four-feature set and the six-feature set in each of the threefolds for a subject and the 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 threefolds 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; indicating the consistency of the method.

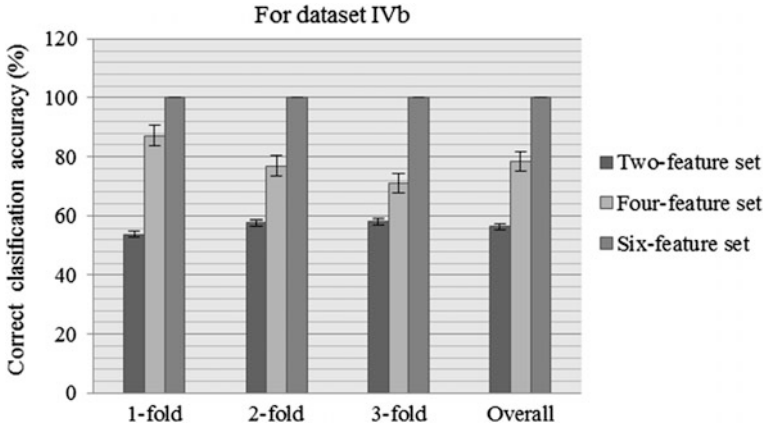


Fig. 9.10 Correct classification rate for the two-feature set, the four-feature set and the six-feature set for each of the three folds in dataset IVb (*Error bars* indicate the standard error)

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.

9.3.3 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 performance evaluation. This section presents a comparative study of 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. 2011), 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 cannot present the comparison results for dataset IVb as there are no reported research results available. The highest classification accuracy rate among the nine algorithms and their averages for each subject is highlighted in bold font.

Table 9.3 provides a comparative study of the quantitative performance achieved by employing this proposed algorithm, versus other recently reported eight well-known algorithms including the BCI Competition III Winner for dataset IVa. As can be seen in Table 9.3, the proposed algorithm yields the highest classification accuracy among the reported methods in three (e.g. S1, S3 and S4) out of the five subjects, which is 100%. For S2 and S5, the accuracy rates are 94.23 and 75.33% respectively; a little 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%. The literature summary given in Table 9.3, shows that compared to the eight other

Table 9.3 Comparison of the classification performance between our proposed algorithm and the most recently 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. 2011)	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

algorithms, the proposed method has produced the best performance for most subjects.

As shown in Table 9.3, the average classification accuracy of the proposed algorithm is 93.91% for the IVa 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.

9.4 Conclusions and Recommendations

This chapter presents a modified version of the CC-LR algorithm where the CC technique is used for feature extraction and the LR model is applied for the classification of the obtained features. This study investigates the types of features most suited to representing the distribution of MI EEG signals. The three sets of two-feature, four-feature and six-feature are tested individually as the input 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. The experimental results show that the six-feature set yields the best classification performance for both datasets, IVa and IVb. The performance of the proposed methodology is compared to eight recently reported methods including the BCI Competition III Winner algorithm. The experimental results demonstrate that our 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 task EEG data and that the C3 channel provides better classification results as a reference signal. The LR is an efficient classifier for distinguishing the features of the MI data.

References

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.

De Veaux, R. D., Velleman, P.F. and Bock, D.E. (2008) Intro Stats (3rd edition), Pearson Addison Wesley, Boston.

Islam, M. N. (2004) *An introduction to statistics and probability*, 3rd ed., Mullick & brothers, Dhaka New Market, Dhaka-1205, pp. 160–161.

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.

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.

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.

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, Issue 3, pp. 358–372.

Siuly, Y. Li, and P. Wen, (2014) 'Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain computer interface', *Computer Methods and programs in Biomedicine*, Vol. 113, no. 3, pp. 767–780.

Siuly, X. Yin, S. Hadjiloucas, Y. Zhang, (2016) 'Classification of THz pulse signals using two-dimensional cross-correlation feature extraction and non-linear classifiers', *Computer Methods and Programs in Biomedicine,127*, 64–82.

Siuly, Y. Li, and P. Wen, (2013) 'Identification of Motor Imagery Tasks through CC-LR Algorithm in Brain Computer Interface', *International Journal of Bioinformatics Research and Applications*, Vol.9, no. 2, pp. 156–172.

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.

Chapter 10
Improving Prospective Performance in MI Recognition: LS-SVM with Tuning Hyper Parameters

This chapter intends to introduce a scheme which can improve the perspective performance in MI task recognition. To achieve this purpose, we develop a CC aided LS-SVM in this chapter. The motivation for considering the LS-SVM with the CC method as the LS-SVM is a very powerful tool in pattern recognition due to its equality constraints (rather than inequality constraints) and the reduction of the computational cost and low risk of over-fitting. In the proposed approach, the LS-SVM is introduced with a two-step grid search process for tuning hyper parameters so that the optimal parameters of this method can be selected efficiently and automatically for distinguishing the MI tasks improving the classification performance. In the previous chapters, the LS-SVM classifier was used in the epileptic EEG signals classification but the hyper parameters of this classifier were selected manually (rather than using an appropriate method) even though the hyper parameters of the LS-SVM play an important role in classification performance. In addition, this chapter investigates the effects of two different reference channels, for example 'Fp1' and 'C3', on performance. To compare the effectiveness of the proposed classifier, we replace the LS-SVM with the logistic regression and kernel logistic regression (KLR) classifier, and apply it separately, for the same features extracted from the CC technique. The methodology proposed in this chapter is tested on datasets IVa, IVb and BCI Comp III (see Chap. 3 for a description of these datasets).

10.1 Motivation

The goal of this chapter is to improve the classification performance of MI task recognition in BCI applications and to investigate whether the LS-SVM with tuning hyper parameters is better than the logistic regression or KLR classifier with the CC features in EEG-based MI classification. For this purpose, this chapter proposes a different algorithm where the CC technique is developed for feature extraction apart

S. Siuly et al., *EEG Signal Analysis and Classification*,
Health Information Science, DOI 10.1007/978-3-319-47653-7_10

from the MI EEG data because the recorded multi-channel EEG signals are highly correlated and cannot supply independent information about brain activities. The different signals from different scalp sites do not provide the same amount of discriminative information (Meng et al. 2009). These signals are also usually very noisy and so not directly usable in BCI applications. The CC technique can reduce noise by means of correlation calculation. Hence the cross-correlation sequence is a nearly noise-free signal that can provide more signal information compared to the original signal (Hieftje et al. 1973). After feature extraction, the LS-SVM is employed for classifying the obtained features. In this chapter, we consider the LS-SVM in this MI classification owing to its robust and accurate and less computational cost. The LS-SVM 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 never been applied [except (Siuly et al. 2011a)] in MI task classification in BCIs. However, in the algorithm (Siuly et al. 2011a), 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 using 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 performance of the LS-SVM classifier is compared to a logistic regression classifier and a KLR classifier for the same feature vector set. 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.

10.2 Cross-Correlation Based LS-SVM Approach

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 Fig. 10.1. 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.

To evaluate the performance of the LS-SVM classifier, we test a logistic regression classifier and a KLR classifier, employed separately. They also employ the same features as those extracted from the CC method as the inputs. The block diagram of the proposed method in Fig. 10.1 depicts the procedure for MI EEG signal classification as described in the following steps.

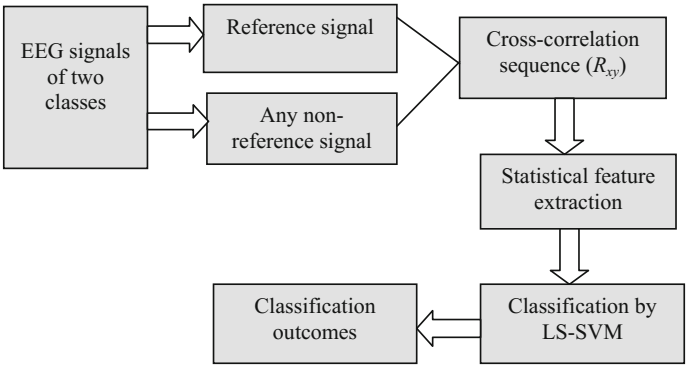


Fig. 10.1 Block diagram of the proposed CC-LS-SVM technique for MI EEG signal classification in BCI development

10.2.1 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 a signal with noise will be incoherent with anything in the reference. In this work, any signal that is not a reference signal is treated as a non-reference signal. In this study, we use two datasets—IVa and IVb of BCI Competition III (see description of both datasets in Chap. 3).

Dataset IVa consists of MI task EEG signals from the right hand class and the right foot class. Dataset IVb consists of MI task EEG signals of the left hand class and the right foot class. Both datasets contain 118 channel data in each class of a subject. For all subjects, we consider the electrode position Fp1 in the international 10/20 system as the reference signal for the CC technique. For dataset IVa, Fp1 is selected from the right hand class while the right foot class is used for dataset IVb. Figure 10.2a, b show the typical reference signals of subject **aa** for dataset IVa and dataset IVb, respectively.

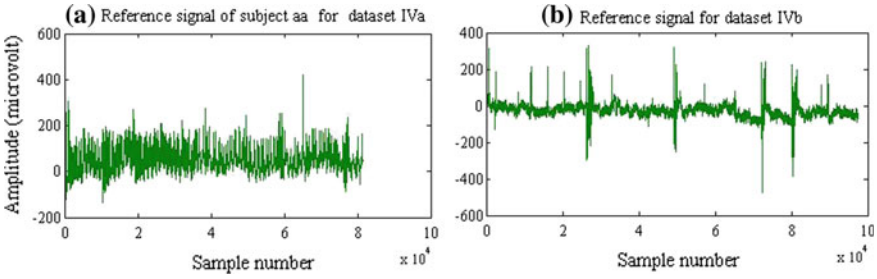


Fig. 10.2 Typical reference signals: **a** dataset IVa and **b** dataset IVb

10.2.2 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 CC technique as shown in Fig. 10.1. From Chap. 8, Eq. (8.1) of the CC method (Dutta et al. 2010; Chandaka et al. 2009; Siuly and Li 2012; Siuly et al. 2011c) 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 from 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 10.3 presents typical signals of the right hand and 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 Fig. 10.3. 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 Fig. 10.3.

Figure 10.4 shows typical signals of dataset IVb for the right foot MI and the left hand MI. This figure also presents typical results of the CC method for the right foot MI signal and the left hand MI signal. As shown in Fig. 10.4, 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.

It is known that if two curves have exactly the same shape, this means that they are highly cross-correlated with each other and CC is around 1. From Figs. 10.3 and 10.4, one can see that the shapes of the two curves are not exactly the same,

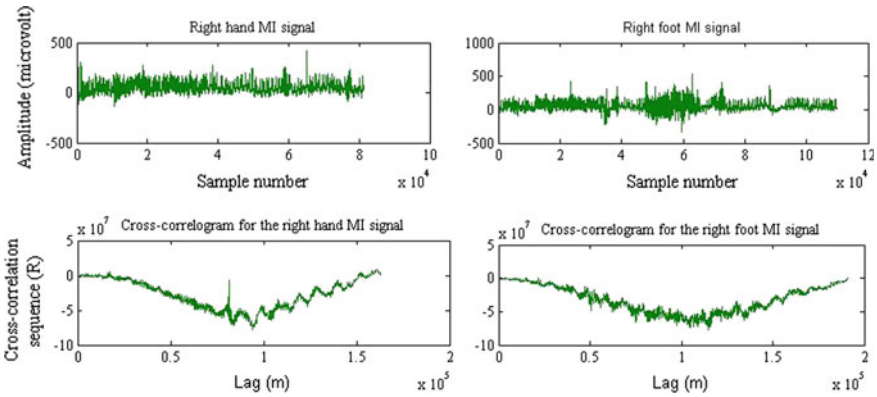


Fig. 10.3 Typical right hand and right foot MI signals and their respective cross-correlograms for subject **aa** in dataset IVa

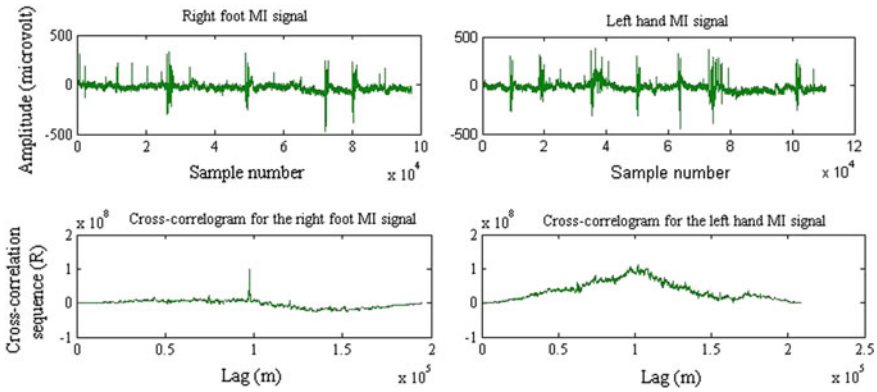


Fig. 10.4 Typical right foot and left hand MI signals and their respective cross-correlograms for dataset IVb

indicating statistical independency. This means that there is a greater chance of achieving better separation.

As mentioned in Chap. 8, if each of x and y signals has a finite number of samples N, the resulting cross-correlation sequence has $(2\,N - 1)$ samples. Hence, in each of Figs. 10.2, 10.3 and 10.4, 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 $0–2 \times 10^5$ samples. The cross-correlogram signals convey greater signal information and consist of low level noises compared to the original signal. It is worth mentioning that a cross-correlogram contains information about the frequencies 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-correlogram as discussed in the following section.

10.2.3 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 contain all important information of the original signal patterns. 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 subjects (Hieftje et al. 1973; Wren et al. 2006; Siuly and Li 2012). The reasons of choosing the feature sets are described in detail in Sect. 8.5.1 of Chap. 8.

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 tenfold cross-validation method, which is discussed in Sect. 10.2.5. These feature vectors are inputs for the LS-SVM and also for the logistic regression and the KLR classifiers in the classification stage.

10.2.4 Classification

This study employs the LS-SVM with radial basis function (RBF) kernel as a classifier to distinguish the features obtained from the CC technique. The decision function of the LS-SVM in Eq. (4.10) is derived directly from solving a set of linear equations (Thissen et al. 2004; Suykens et al. 2002; Siuly et al. 2009; Siuly and Li 2012). A detailed description of the LS-SVM algorithm is provided in Chap. 4.

In this study, the obtained a training feature vector of six dimensions used as the input in Eq. (4.10) 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 = +1 and right hand = −1, and for the dataset IVb, y_i is considered as right foot = + 1 and left hand = −1. In this study, the RBF kernel is chosen for the LS-SVM after testing of the other kernel functions. The two important parameters (γ, σ^2) of the LS-SVM are selected by a two-step grid search technique for obtaining reliable performance from the method discussed in Sect. 10.3.1. The solution of Eq. (4.10) provides the prediction results that directly assign the samples with a label +1 or −1 to identify the category to which it belongs.

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 Chap. 8, the logistic regression (Caesarendra et al. 2010; Hosmer and Lemeshow 1989; Siuly et al. 2011c) in Eq. (8.2) is applied for the classification of the MI features. A detailed description of the logistic regression is available in Chap. 8. 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 Eq. (8.2) for the both datasets. We treat the dependent variable y as right hand = 0 and right foot = 1 for dataset IVa and right foot = 0 and left hand = 1 for dataset IVb. Finally, we obtain the prediction results that directly provide the class label 0 or 1 for the samples.

We also compare the performance of the LS-SVM with a KLR classifier. KLR is a nonlinear form of logistic regression. It can be achieved via the so-called "kernel trick" which has the ability to classify data with nonlinear boundaries and can also accommodate data with very high dimensions. A detailed description of the KLR is available in (Cawley and Talbot 2008; Rahayu et al. 2009). A final solution of the KLR could be achieved using the following equation.

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x) + b \tag{10.1}$$

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 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 the iteratively re-weighted least square procedure (Rahayu et al. 2009).

In the KLR (Siuly and Li 2012), 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 KLR as an estimate of a posterior probability of the class

membership. In Sect. 10.3.3, 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 tenfold cross-validation procedure.

10.2.5 Performance Measure

The classification accuracy has been one of the main pitfalls in developed BCI systems. It directly affects the decision made in a BCI output. Thus, this study calculates classification accuracy using *k*-fold cross-validation (Abdulkadir 2009; Ryali et al. 2010, Siuly et al. 2011b) for assessing the performance of the proposed method. A detailed description of the classification accuracy and the *k*-fold cross-validation method is provided in Chap. 3.

In this study, we select $k = 10$ as it is a common choice for the *k*-fold cross-validation. As mentioned in Sect. 10.2.3, 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 3.5 presents the design for the extracted feature vectors of this study partitioned into 10 mutually exclusive subsets according to the *k*-fold cross-validation system. As shown in Fig. 3.5, 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 training set (as illustrated in Fig. 3.5). The classification accuracy, obtained for each of 10 times on the testing set, is averaged and called 'tenfold cross-validation accuracy' in this chapter.

10.3 Experiments and Results

Before classification, the hyper parameters of the LS-SVM classifier is tuned by the two-step grid search algorithm discussed in Sect. 10.3.1 as the classification performance of the LS-SVM depends on the parameters, and the values chosen for the parameters significantly affect the classification accuracy. Section 10.3.2 discusses how the variables are set up in the logistic regression classifier and the KLR classifier. The results obtained by the LS-SVM, the logistic regression and the KLR for the CC based features are compared to each other for datasets IVa and IVb (in Sect. 10.3.3). Section 10.3.4 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 KLR 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 separately employed on each subject for both datasets, as the MI EEG signals are naturally highly subject-specific depending on the required physical and mental tasks. All experimental results are presented based on the testing set in this study.

10.3.1 Tuning the Hyper Parameters of the LS-SVM Classifier

To improve the generalization performance of the LS-SVM classifier, two parameters (γ, σ^2) are chosen through an appropriate procedure. These parameters play an important role in the classification performance. The regularization parameter γ (gamma) determines the trade-off 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. 2009; Suykens and Vandewalle 1999), so the values must be chosen carefully. This study applies a two-step grid search technique to obtain the optimum values of the hyper parameters for the LS-SVM. This section describes the process for parameter selection 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 tenfolds of a subject of 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 a reliable performance of the proposed method. Figure 10.5a, 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 onefold of Subject **aa**) and dataset IVb (for the 1-fold), respectively.

The optimal range of 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

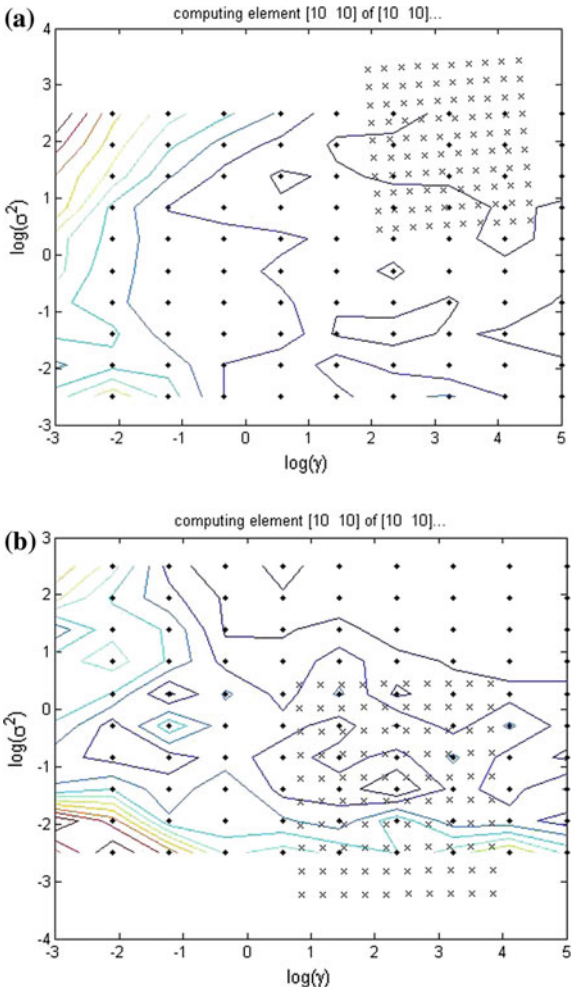


Fig. 10.5 **a** Process of the two-step grid search for optimizing the parameters γ (gamma) and σ^2 (sig2) of the LS-SVM classifier in the onefold of Subject **aa** of dataset IVa. **b** Process of the two-step grid search for optimizing the parameters γ (gamma) and σ^2 (sig2) of the LS-SVM classifier in the 1-fold of dataset IVb

error contour line. The grids denoted as "×" in the second step are 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 optimal combinations of γ and σ^2 obtained for the LS-SVM are presented in Table 10.1 for dataset IVa and in Table 10.2 for dataset IVb.

As shown in Tables 10.1 and 10.2, the optimal values of the hyper parameters for the LS-SVM are obtained in each of tenfolds 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 both datasets.

Table 10.1 Optimal values of the parameters γ and σ^2 of the LS-SVM for dataset IVa

Subject	Obtained optimal parameter values of γ and σ^2 for the LS-SVM									
	aa		**al**		**av**		**aw**		**ay**	
Parameters	γ	σ^2	γ	σ^2	γ	σ^2	γ	σ^2	γ	σ^2
Onefold	82.13	8.84	65.89	5.52	22.16	5.06	260.30	11.74	58.12	4.69
Twofold	31.59	7.04	60.05	8.32	220.64	7.88	181.25	11.19	72.15	1.53
Threefold	202.90	16.09	245.97	2.24	7.33	6.46	401.84	14.81	45.85	1.88
Fourfold	128.22	18.62	80.27	5.66	921.84	1.78	343.21	12.92	92.86	1.05
Fivefold	30.79	9.48	315.18	14.67	4.94	8.66	200.97	0.79	42.59	0.65
Sixfold	58.76	20.79	632.36	13.27	10.11	3.92	141.36	12.81	78.74	1.39
Sevenfold	46.71	7.73	261.91	11.72	719.69	5.42	193.22	8.44	26.18	0.73
Eightfold	29.75	10.85	64.29	2.21	9.23	3.99	179.95	13.72	149.09	0.84
Ninefold	208.08	32.27	49.22	1.65	10.25	4.04	75.48	7.41	246.47	1.52
Tenfold	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 10.2 Optimal values of the parameters γ and σ^2 of the LS-SVM for dataset IVb

Parameters	Obtained optimal parameter values γ and σ^2 for the LS-SVM	
	γ	σ^2
Onefold	7.0451	5.3714
Twofold	6.9404	1.5538
Threefold	47.7992	3.7401
Fourfold	107.1772	1.7566
Fivefold	10.0366	1.8417
Sixfold	320.4905	3.3605
Sevenfold	57.1816	2.7994
Eightfold	820.5462	1.3092
Ninefold	569.3277	2.1852
Tenfold	31.9349	1.8465

10.3.2 Variable Selections in the Logistic Regression and Kernel Logistic Regression Classifiers

Although the parameters of the logistic regression are obtained automatically through the 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 Eq. (8.2) 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 Eq. (8.2) 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, in logistic regression, the dependent variable y has two values, 0 and 1. 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 KLR, the model parameters in Eq. (8.2) are automatically anticipated by the iteratively re-weighted least square procedure (Rahayu et al. 2009). The feature vectors and class labels of the KLR in Eq. (10.1) are considered to be same as those in the logistic regression.

10.3.3 Performances on Both Datasets

Table 10.3 presents the classification results for the LS-SVM, the logistic regression and the KLR classifiers for the five subjects of dataset IVa. In Table 10.3, the results of each subject are reported in terms of mean ± standard deviation of the accuracy over a tenfold cross-validation method on the testing set. It is observed from Table 10.3 that the proposed LS-SVM classifier for the CC 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 the 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 KLR 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 KLR in four out of five subjects.

Table 10.3 also reports that the standard deviations for the proposed approach are much lower than those of the logistic regression and the KLR in these four subjects. The lower values of standard deviation indicate the consistency of the proposed method. As seen from Table 10.3, 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 KLR classifier. The average

Table 10.3 Classification results by the tenfold cross-validation method on testing set of dataset IVa

Subject	Tenfold 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

classification accuracy for the proposed method increases by 6.18% in comparison to the logistic regression model and 2.34% to the KLR.

In what follows, we detail how the tenfold cross-validation system produces the classification accuracy in each of the tenfolds for one subject applying the LS-SVM, the logistic regression and the KLR classifiers. Figure 10.6a–e plot the comparative results of each of the tenfolds for the five subjects for dataset IVa. The figures show the individual classification accuracies against each of the tenfolds for the logistic regression, KLR 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 tenfolds compared to the logistic regression and the KLR. An increasing tendency of prediction accuracy in every fold of all subjects for the LS-SVM is shown in these figures.

From Fig. 10.6a–e, the fluctuations of the performance of the proposed method are smaller among the tenfolds for each subject compared to the logistic regression model and the KLR model, indicating that the proposed method is fairly stable.

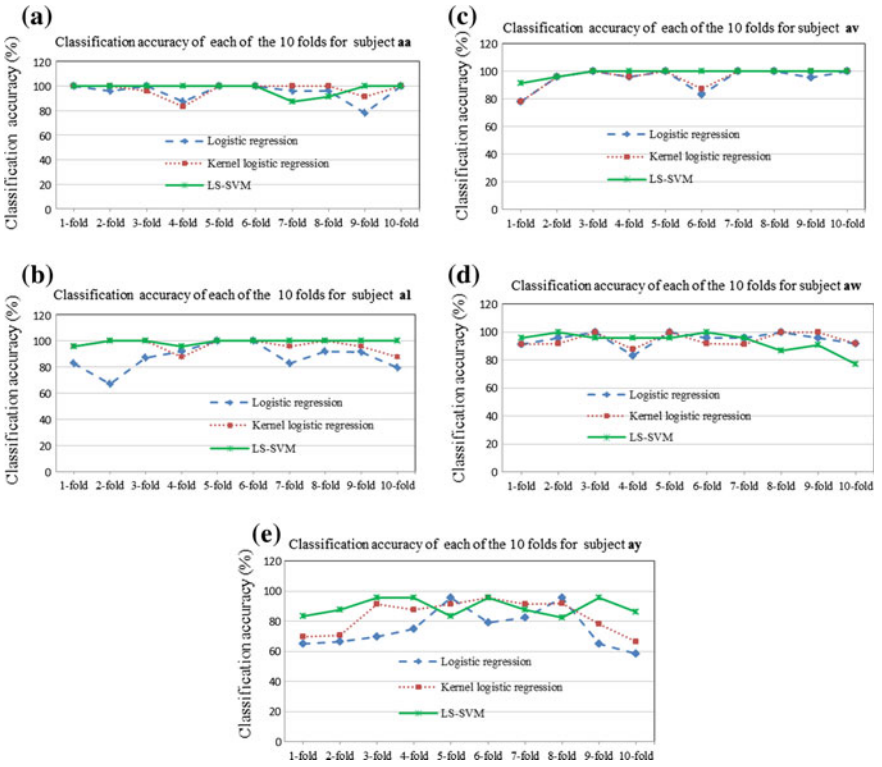


Fig. 10.6 Comparisons of the individual classification accuracies for the logistic regression, kernel logistic regression and the LS-SVM for each of the tenfolds: **a** subject **aa**, **b** subject **al**, **c** subject **av**, **d** subject **aw**, **e** subject **ay** in dataset IVa

Table 10.4 Classification results by the tenfold cross-validation method on testing set of dataset IVb

Methods	Tenfold 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

In Table 10.4, we provide the classification accuracy for the proposed LS-SVM, the logistic regression and the KLR models using the tenfold cross-validation procedure for dataset IVb. As shown in Table 10.4, 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 KLR. The results show a 2.58% improvement in the proposed LS-SVM compared to the logistic regression and 3.02% over the KLR for the same inputs.

The standard deviation value is also smaller in the LS-SVM compared to the logistic regression and the KLR, reflecting the consistency of the LS-SVM. The results in terms of the tenfold cross-validation accuracy on both datasets displayed in Tables 10.3 and 10.4, demonstrate that the proposed LS-SVM classifier is superior compared to the logistic regression and the KLR methods for the same features.

From Fig. 10.7, it is observed that in most of the tenfolds, the proposed LS-SVM generates higher accuracies and the performance variations among the tenfolds are smaller compared to those of the logistic regression and the KLR. These indicate that the proposed method is more reliable for MI signal classification. From Figs. 10.6a–e and 10.7, it is clear that the proposed algorithm achieves a better classification performance, both individually and overall, compared to the logistic regression and the KLR.

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 selected from the right hand class from dataset IVa, while it is selected from the right foot class for dataset IVb. Using the tenfold 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**,

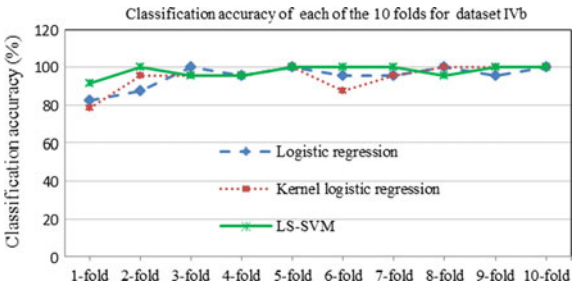


Fig. 10.7 Comparisons of the individual classification accuracies for the logistic regression, kernel logistic regression and the LS-SVM for each of the tenfolds in dataset IVb

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 10.5). 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 tenfold 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 87.82% for the reference signal C3 and 89.54% for the reference signal Fp1.

For the same dataset, the KLR produces a 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 Fp1 as (provided in Table 10.5). Thus, the KLR 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 Fp1, as shown in Table 10.6. 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 KLR is able to generate the classification performance of 95.36% where as 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 when changing the reference signal. This proves the robustness of the method.

In addition to investigating 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}} to the existing feature set, we get a feature set of nine features which are {mean, median, mode, standard deviation, maximum and minimum, IQR, P_{25} and P_{75}}. Tables 10.5 and 10.6 present the classification results for three classifiers (the LS-SVM, the logistic regression and the KLR) for the nine feature set, comparing it with the results of the existing six feature set for two reference signals, Fp1 and C3, for datasets IVa and IVb, respectively. As shown in Table 10.5, the proposed LS-SVM based algorithm with nine features achieves a 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. 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 feature set. From Table 10.5, it is also seen that the

Table 10.5 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

Table 10.6 Classification results of the three classifiers for nine features and 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

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 10.5 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 10.5, 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 10.5 that the KLR 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 as these values are 97.89 and 97.88%, respectively, for the six features as shown in Table 10.6. 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 KLR 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.

From this discussion, 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 for representing a distribution. As the datasets used in this study are almost symmetric and there are no obvious outliers, we do not obtain significantly better performance when the three features {IQR, P_{25} and P_{75}} are added to the six features.

10.3.4 Performance Comparisons with the Existing Techniques

To further examine the efficiency of the proposed algorithm, this section provides the comparisons of our approach with eight other recently reported techniques. Table 10.7 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 for each subject and their averages is highlighted in bold font for each subject.

Table 10.7 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. 2011a)	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 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
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* CC technique; *CT* clustering technique; *R-CSP* regularized common spatial pattern; *CSP* common spatial pattern

From Table 10.7, 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, with the highest classification rates being 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**. Further looking at the performance comparisons in Table 10.7, 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. 2011a) comes second (88.32%), R-CSP with aggregation (Lu et al. 2010) is third (83.9%) and so on.

The sparse spatial filter optimization (Yong et al. 2008) is the last (73.50%). These results indicate that the proposed method achieves 7.40–22.22% improvement over all the eight existing algorithms for BCI Competition III's dataset IVa.

10.4 Conclusions

In this chapter, we present the CC-LS-SVM algorithm for improving the classification accuracy of MI-based EEG signals in BCI systems. The proposed scheme utilizes a cross-correlogram-based feature extraction procedure for 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 the logistic regression and KLR models to compare 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 for 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 tenfold 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 two-class MI EEG signal classification. The feasibility of the approach has been verified with BCI Competition III datasets IVa and IVb
2. The CC 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 KLR, confirm that the extracted features reliably capture the valuable information from the original MI signal patterns
3. To further investigate the reliability of the obtained features, we add another three features {IQR, P_{25} and P_{75}}, to the current six features and then separately employ the LS-SVM, the logistic regression and the KLR algorithms as the inputs. The results show that the performance of the nine features is not much improved in comparison to 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 manual selection
5. The results show that the proposed LS-SVM classifier achieves a better performance compared to the logistic regression and the KLR 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's dataset IVa, by at least 7.40%. It demonstrates that our method is the best performing 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 potential for the development of MI-based BCI analyses which assist clinical diagnoses and rehabilitation tasks.

References

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

BCI competition III, 2005, http://www.bbci.de/competition/iii.

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.

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.

Chandaka, S., Chatterjee, A. and Munshi, S. (2009) 'Cross-correlation aided support vector machine classifier for classification of EEG signals', *Expert System with Applications,* Vol. 36, pp. 1329–1336.

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.

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.

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.

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.

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.

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.

MATLABArsenal-online, http://www.informedia.cs.cmu.edu/yanrong/MATLABArsenal/MATLABArsenal.zip.

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.

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.

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.

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. (2011a) '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. (2011b) '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, Y., Li, J. Wu, and J. Yang (2011c) '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.

Siuly and Y. Li, (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, *Vol*. 20, no. 4, pp. 526–538.

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.

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.

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.

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.

Chapter 11
Comparative Study: Motor Area EEG and All-Channels EEG

This chapter reports a comparative study between motor area EEG and all-channels EEG for the three algorithms which proposed in Chap. 10. In this chapter, we intend to investigate two particular issues: first, which of the three algorithms is the best for MI signal classification, and second, which EEG data, 'motor area data or all-channels data' is better for providing more information about MI signal classification. In Chap. 10, we introduced the three methods; the CC-based LS-SVM (CC-LS-SVM), the CC-based logistic regression (CC-LR) and the CC-based kernel logistic regression (CC-KLR) for MI signal classification in BCI applications. We implement these three algorithms on the motor area EEG and the all-channels EEG to investigate how well they perform, and also test which area EEG is better for MI EEG data classification. These three algorithms are also compared with the other existing methods. Wang and James (2007) introduced the concept of selecting EEG channels for MI tasks over the motor cortex area. In this study, we follow their work for our consideration of motor area EEG data.

11.1 Motivations

As EEG signals present brain activities as multichannel time series from multiple electrodes placed on the scalp of a subject, the different channels convey different information about the brain. The different signals from different scalp sites do not provide the same amount of discriminative information. Thus, this study aims to explore the performance of the EEG channels from the motor cortex area and from the all-channels EEG data. In the human brain, the motor cortex is a very important area that controls the voluntary muscle movements discussed in detail in Sect. 9.2. In this chapter, the EEG channels of the motor cortex area are considered according to the suggestions of Wang and James (2007). The major aim of this study is to investigate which area (motor area or the whole brain) is better for acquiring MI information for classification and to discover which algorithm performs better for

S. Siuly et al., *EEG Signal Analysis and Classification*,
Health Information Science, DOI 10.1007/978-3-319-47653-7_11

MI classification. Hence, this study addresses two questions: (i) Which algorithm is the best for MI classification? (ii) Which EEG dataset is better for MI signal classification? Is it the motor area data or is it the all-channels data? To answer these two questions, this chapter uses our three algorithms, CC-LS-SVM, CC-LR and CC-KLR, that were proposed in Chap. 10. We implement these three algorithms considering 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 2005) in almost the same way they were used in Chap. 10. There is only one difference which is, in Chap. 10, we used the tenfold cross-validation method to evaluate the performances. But in this chapter, we consider the threefolds cross-validation procedure to reduce computation time and the number of experiments.

11.2 Cross-Correlation-Based Machine Learning Methods

11.2.1 CC-LS-SVM Algorithm

The CC-LS-SVM algorithm (see detailed description in Chap. 10) is a hybrid approach where the cross-correlation (CC) technique is used for feature extraction, and the LS-SVM is applied for the classification of the extracted features. A brief description of this algorithm is provided below:

1. The C3 electrode position is considered as a reference channel
2. The C3 channel is cross-correlated with the data of the remaining channels and the cross-correlation sequences are obtained using the reference channel and any one of other channels. The detailed description of the CC technique is available in Chap. 8 and also in Ref. (Siuly and Li 2012, 2014a, b; Dutta et al. 2010; Chandaka et al. 2009)
3. The six statistical features, mean, median, mode, standard deviation, maximum and minimum, are extracted from each cross-correlation sequence to characterize the distributions of EEG signals, which reduce the dimension of the cross-correlation sequence
4. The extracted features are segmented as a training set and testing set using a threefolds cross-validation process
5. A two-step grid search technique (Siuly and Li 2012; Xie et al. 2009; Thissen et al. 2004)is implemented separately to each of the threefolds of a threefolds cross-validation method to select the optimum values of the hyperparameters (γ, σ^2) for the LS-SVM
6. After selecting the optimal values of the hyperparameters, the training vector set is used to train the LS-SVM classifier with a radial basis function (RBF) kernel and the testing vector set is applied as the inputs to evaluate the classification accuracy and effectiveness of the classifier with the selected parameters. Detail

of the LS-SVM algorithm is provided in Chap. 4 and is also available in reference (Suykens et al. 2002; Siuly et al. 2009).
7. The outputs of the LS-SVM algorithm provide the prediction results that directly assign the samples with a label +1 or −1 to identify the category to which it belongs.

11.2.2 CC-LR Algorithm

The CC-LR algorithm combines two techniques, CC and logistic regression (LR) for classifying MI tasks in BCI applications (presented in Chap. 10). This algorithm performs in two stages: feature extraction and feature classification. The CC approach is employed to extract the features from the original MI data and the LR is used to distinguish the features. The main steps of the CC-LR algorithm are given below:

1. This algorithm follows Steps 1–4 of the CC-LS-SVM algorithm to extract features by using the CC technique
2. Then we employ the training and testing feature sets (separately) to the LR classifier as the inputs. The performance of the LR classifier is assessed based on the outcomes of the testing set. A detailed description of the LR method is available in Chap. 8 and also in references (Siuly et al. 2013; Caesarendra et al. 2010; Hosmer and Lemeshow 1989; Subasi and Ercelebi 2005)
3. The parameters of the LR model are estimated separately by maximum likelihood estimation (MLE) (Subasi and Ercelebi 2005) for each of the threefolds
4. The classification results are obtained at this stage. Based on the outcomes, we can decide how many values the algorithm correctly predicts for each of the two classes.

11.2.3 CC-KLR Algorithm

In the CC-KLR (see detailed description in Chap. 10) algorithm, we employed the CC and kernel logistic regression (KLR) together for the classification of MI tasks from EEG signals. In this method, we utilize the feature vectors and their class labels with the same process as the logistic regression for the inputs. The brief description of the CC-LR algorithm is provided below:

1. Steps 1–4 of the CC-LS-SVM algorithm are also followed to extract features in this method
2. Then the training and testing feature sets are used individually, for the KLR classifier as inputs where an evaluation is performed on the testing set

3. The parameters of the KLR model are separately estimated by the MLE for each of the threefolds
4. Finally, the outputs of the kernel logistic regression are obtained as an estimate of a posterior probability of the class membership.

11.3 Implementation

To undertake an experimental evaluation, the three proposed methodologies (described in Chap. 10), CC-LS-SVM, CC-LR and CC-KLR algorithms, are implemented on two publicly available datasets, IVa and IVb of BCI Competition III (Blankertz et al. 2006; BCI competition III-online), for experimental. We described datasets IVa and IVb of BCI Competition III in Chap. 3. All of the EEG data of these two sets were collected during MI tasks.

In this study, we intend to implement our three methods on the electrodes of the motor cortex area of the brain, and on the all-channel electrodes for comparison of the two. 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 dimension data. We are interested to see the performances of the three algorithms on the two areas (motor area and all-channels data) and also to decide which algorithm is better for each given areas of the brain. We know that only a particular part of the brain is activated in response to a MI task, and this area is called the motor cortex. The motor cortex is an important brain area mostly involved in the control and execution of voluntary motor functions, and is typically associated with 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 the international 10–20 system which includes the channels of electrodes 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 and James (2007) also considered the same electrodes for their research and their experimental results suggested that these electrodes are the best channels for obtaining MI information.

As previously described, both datasets are originally recorded from 118 electrodes. Figure 11.1 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 and Praamstra 2001). Wang and James (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.

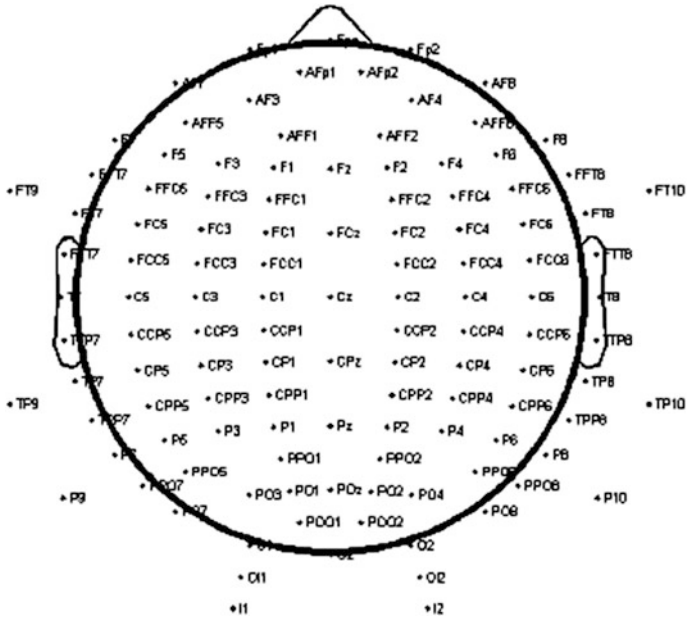


Fig. 11.1 Locations of electrodes for datasets IVa and IVb in BCI Competition III [118 electrodes are shown labelled according to the extended international 10–20 system described in (Oostenveld and Praamstra 2001)]

In this study, we first consider the electrode position C3 of the RH class as a reference channel from each subject of the 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 MI information about brain activities during the MI tasks in the international 10–20 system (Sander et al. 2010). For each subject, the C3 channel is used as a reference channel for both the motor imagery EEG data and the all-channels EEG data.

Second, 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, a total of 35 cross-correlation sequences are obtained from the two classes of each subject. Next, the six mentioned 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 the previously mentioned six statistical features from each cross-correlation sequence to generate a feature vector set of 235×6 size.

Third, we divide the feature vector set randomly as the training set and the testing set using the threefolds cross-validation method (Siuly et al. 2011a; Abdulkadir 2009) in both the motor cortex set and the all-channels data, separately. In the threefolds cross-validation procedure, a feature vector set is partitioned into three mutually exclusive subsets of approximately equal size and the method is repeated three 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 three 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 (Siuly and Li 2012; Siuly et al. 2014a, b), 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 of the threefolds for to obtain a 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 (Siuly and Li 2012; Siuly et al. 2014a, b), we separately employ the training and testing sets as the inputs 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 (Siuly and Li 2012; Siuly et al. 2014a, b), we utilize the feature vectors and their class labels as the same process of the logistic regression for the inputs. In Sect. 11.4, the classification results of these three classifiers are presented for datasets IVa and IVb. The following section discusses the performance of those three algorithms through a threefolds cross-validation procedure.

11.4 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 report a comparative study with existing methods. As accuracy is a major concern in BCI systems, this study uses 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, 2011a, b). It is worth mentioning 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. Classification by the LS-SVM is carried out in MATLAB using the LS-SVMlab toolbox (LS-SVMlab toolbox (version 1.5)-online), 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).

11.4.1 Results for Dataset IVa

The complete experimental results for dataset IVa are summarized in Table 11.1. 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 threefolds cross-validation method on the testing set. In the motor area, the CC-LS-SVM algorithm yields a classification accuracy of 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. The standard deviation value of a subject describes the variation of the classification accuracies among the threefolds. If the variation of the accuracies among the threefolds is less, it indicates robustness of the method. For the motor area data, we can see that the standard deviation among the threefolds in each subject is relatively small in the CC-LS-SVM algorithm, indicating 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 is reliable in the all-channel data. The results reveal that the CC-LS-SVM algorithm performs better on both the motor area and all-channels

Table 11.1 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

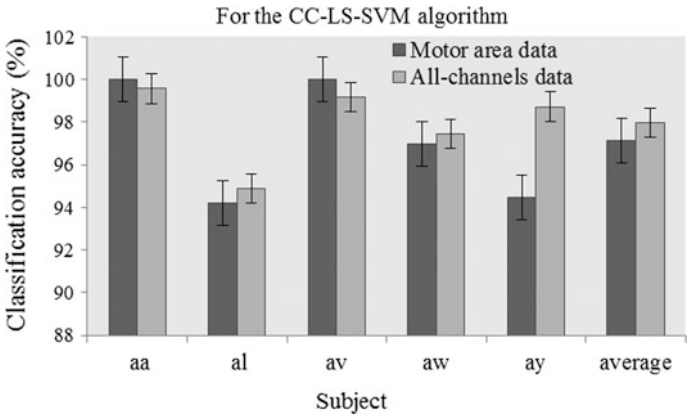


Fig. 11.2 Comparison of the performance between the motor area EEG and the all-channels EEG data for the CC-LS-SVM algorithm (The *vertical lines* show the standard errors of the test accuracies)

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 it is for the motor area data.

Figure 11.2 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 11.2 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 11.3 displays the comparison of the classification accuracy between the motor area EEG and the all-channels EEG data for the CC-LR algorithm. From the Fig., it can be observed that, compared to the motor area data, the classification accuracy rates for the all-channels data are substantially higher for Subjects, ***aa***, ***al*** and ***av*** and are the same for Subject ***aw***. 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 11.4 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 slightly 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

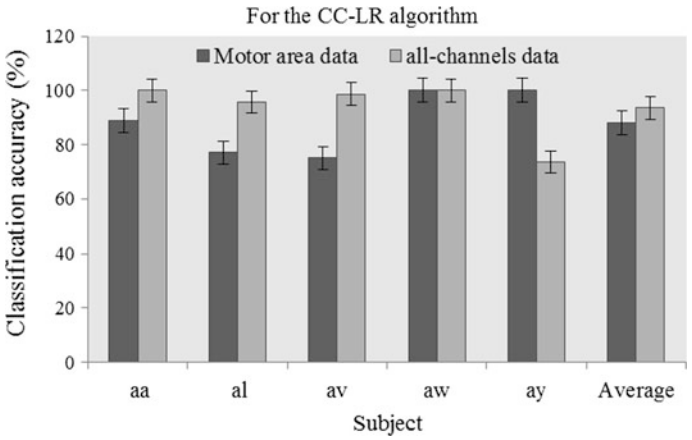


Fig. 11.3 Comparison of the performance between the motor area EEG and the all-channels EEG data for the CC-LR algorithm (The *vertical lines* show the standard error of the test accuracies)

to the motor channel data. The average performance of the all-channels is better than the motor area data for this algorithm. Figures 11.2, 11.3 and 11.4 show that the EEG data recorded from the all-channels gives the best result for both algorithms when compared to the data recorded from the motor cortex area.

Table 11.2 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 and James 2007) and SVM on dynamical system (DS) features (Song et al. 2007). From Table 11.2, 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

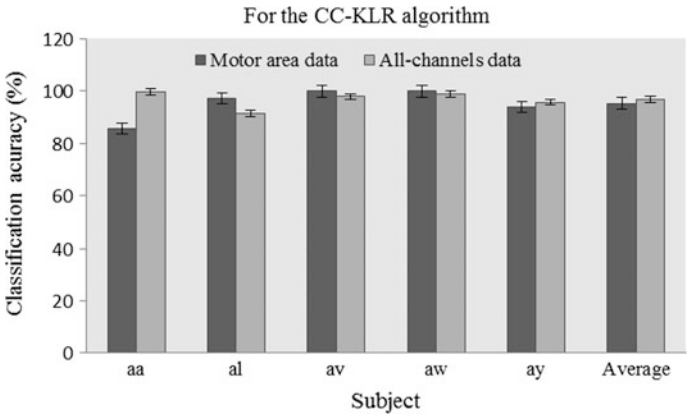


Fig. 11.4 Comparison of the performance between the motor area EEG and the all-channels EEG data for the CC-KLR algorithm (The *vertical lines* show the standard error of the test accuracies)

Table 11.2 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 and James 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

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***. In Table 11.2, 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 improvements of 1.77% to 13.06% for the motor area data over the four algorithms for dataset IVa.

Table 11.3 lists a comparison study for the all-channels data of our three algorithms with BCI III Winner (Blankertz et al. 2006) and iterative spatiospectral patterns learning (ISSPL) (Wu et al. 2008) for dataset IVa. The CC-LS-SVM algorithm produced an excellent result for Subjects ***av*** and ***ay***, while the CC-LR algorithm achieved the best results for Subjects ***aa*** and ***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 Subjects ***al*** and ***aw***. Both the BCI III Winner and ISSPL methods achieved a 100% accuracy for Subject ***al***. Obviously,

Table 11.3 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

the average classification accuracy of the CC-LS-SVM method is excellent for the all-channels data. Table 11.3 shows 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 Tables 11.2 and 11.3 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 previously existing methods. Based on these results, it can be concluded that the LS-SVM method has more potential than the existing methods for MI task 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 MI classification.

11.4.2 Results for Dataset IVb

Table 11.4 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 Fig. 11.5. 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 an accuracy of 98.72%, where the CC-LR method produces 96.83% and the CC-KLR method produces 97.04%.

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 threefolds 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 these three methods in the all-channels data.

Figure 11.5 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 Fig. 11.5, 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

Table 11.4 Experimental results of the three algorithms reported in terms of the threefolds 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

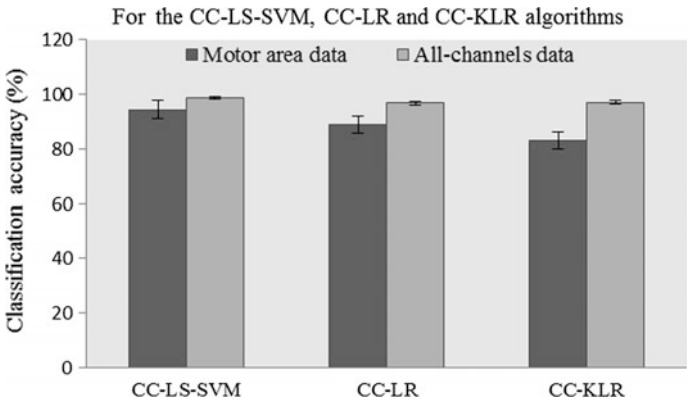


Fig. 11.5 The comparison of the performance for the CC-LS-SVM, CC-LR and CC-KLR algorithms between the motor area data and the all-channels data (The *vertical lines* show the standard errors of the test accuracies)

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.

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.

11.5 Conclusions and Contributions

In this chapter, we have employed the CC-LS-SVM, CC-LR, and CC-KLR algorithms to compare the performances 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 mixes the CC technique and the KLR classifier for MI task classification. To investigate the effectiveness of these three algorithms, we implemented them individually on the EEG data recorded from the motor cortex area and also on 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 EEG signal classification. The results also suggest that the CC-LS-SVM algorithm performs better than the reported 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 MI EEG signal classification and the all-channels EEG can provide better information than the motor area EEG for MI classification.

References

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

BCI competition III, 2005, http://www.bbci.de/competition/iii.

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.

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.

Chandaka, S., Chatterjee, A. and Munshi, S. (2009) 'Cross-correlation aided support vector machine classifier for classification of EEG signals', *Expert System with Applications,* Vol. 36, pp. 1329–1336.

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.

Hosmer, D.W. and Lemeshow, S. (1989) *Applied logistic regression*, Wiley, New York.

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.

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.

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. and Wen, P. (2011a) '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. (2011b) '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 Y. Li, (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, *Vol*. 20, no. 4, pp. 526–538.

Siuly, Y. Li, and P. Wen, (2013) 'Identification of Motor Imagery Tasks through CC-LR Algorithm in Brain Computer Interface', *International Journal of Bioinformatics Research and Applications*, Vol.9, no. 2, pp. 156–172.

Siuly, Y. Li, and P. Wen, (2014) 'Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain computer interface', *Computer Methods and programs in Biomedicine*, Vol. 113, no. 3, pp. 767–780.

Siuly, Y. Li, and P. Wen, (2014) 'Comparisons between Motor Area EEG and all-Channels EEG for Two Algorithms in Motor Imagery Task Classification', *Biomedical Engineering: Applications, Basis and Communications (BME),* Vol. 26, no. 3, pp. 1450040 (10 pages).

Song L, Epps J, (2007) 'Classifying EEG for brain-computer interface: learning optimal filters for dynamical features', *Comput Intell and Neurosci,* Article ID 57180, 11 pages, doi:10.1155/2007/57180, 2007.

Subasi, A. and Ercelebi, E. (2005) 'Classification of EEG signals using neural network and logistic regression', *Computer Methods and Programs in Biomedicine*, Vol. 78, pp. 87–99.

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.

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.

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.

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, 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.

Chapter 12
Optimum Allocation Aided Naïve Bayes Based Learning Process for the Detection of MI Tasks

This chapter presents a reliable and robust analysis system that can automatically detect motor imagery (MI) based EEG signals for the development of brain–computer interface (BCI) systems. The detection of MI tasks provides an important basis for designing a means of communication between brain and computer in the creation of devices for people with motor disabilities. In this chapter, we present a synthesis approach based on an optimum allocation scheme and the Naive Bayes (NB) algorithm for detecting mental states based on EEG signals where the optimum allocation scheme is introduced to discover the most effective representatives with minimal variability from a large number of MI based EEG data. The NB classifier is employed on the extracted features for discriminating MI signals. The feasibility and effectiveness of the proposed method is demonstrated by analyzing the results on two public benchmark datasets.

12.1 Background

In BCI development, users produce EEG signals of different brain activity for different MI tasks that will be identified by a system and are then translated into commands. These commands will be used as feedback for motor disabled patients to communicate with the external environments. If the MI tasks are reliably distinguished through detecting typical patterns in EEG data, a motor disabled people could communicate with a device by composing sequences of these mental states. Thus, a MI-based BCI provides a promising control and communication means for people suffering from motor disabilities. Therefore, the detection of MI tasks is essential for BCI development to generate control signals. In most current MI based BCIs, the detection algorithms are carried out in two stages: feature extraction and feature detection (Mason and Birch 2003). A successful EEG-based BCI system mainly depends on whether the extracted features are able to differentiate MI-oriented EEG patterns. How to improve the recognition performance of MI

S. Siuly et al., *EEG Signal Analysis and Classification*,
Health Information Science, DOI 10.1007/978-3-319-47653-7_12

signals remains a vital issue for the development of BCI systems. The goal of this study is to develop an approach for detecting different MI EEG signals, thus improving classification performance. The present study proposes a methodology where the OA is employed for the feature extraction stage and a probabilistic classifier, NB is employed for the detection of the obtained features.

In this study, there are strong grounds for using an optimum allocation technique to obtain a representative sample from each group of a category of MI data. An optimum allocation technique is introduced to allocate the numbers of sample units into different groups with minimal variation, thus providing the greatest precision. This method is applicable when a dataset is heterogeneous and very large in size. When measuring an EEG, a large amount of data with different categories is obtained over a time period and this huge amount data is not directly usable in BCI applications. The dataset must be divided into several groups to make homogeneity within a group according to the specific characteristics, and used to select representative samples from the groups such that those samples reflect the entire data. Thus, instead of random sampling, this study intends to develop an optimum allocation technique based sampling method to select representative sample points from every time group. In optimum allocation based sampling, sample points are selected from each group considering the variability of the observations, while random sampling does not consider variability. To describe the original patterns of EEG signals more representatively, the variability consideration is the most important thing to provide the highest precision of a sample for the least cost during the selection of sample points from a group. In this study, a sample is defined as a subset (or small part) of observations from a group.

To the best of our knowledge, an optimum allocation based NB approach has not been used on MI data for the detection of MI tasks in BCIs. The reason for choosing the NB method as a classifier for this study is due to the simplicity of its structure and the speed of the learning algorithm it employs (Mitchel 1997, Wiggins et al. 2011). Another advantage is that a small amount of bad data or 'noise' does not perturb the results much. The proposed approach is evaluated on two datasets, IVa and IVb of BCI Competition III (BCI competition III, Blankertz et al. 2006), where both sets contain MI EEG recorded data. A popular k-fold cross-validation method ($k = 10$) is used to assess the performance of the proposed method for reducing the experimental time and the number of experiments in the MI task EEG signal classification. This cross-validation procedure is applied to control overfitting of the data. The performance of the proposed approach is also compared with five most recently reported methods.

12.2 Optimum Allocation Based Naïve Bayes Method

The proposed approach aims to develop a methodology for the detection of MI based EEG signals for application in BCI systems that can work automatically. The scheme developed in this study is labeled as four stages as described in Fig. 12.1.

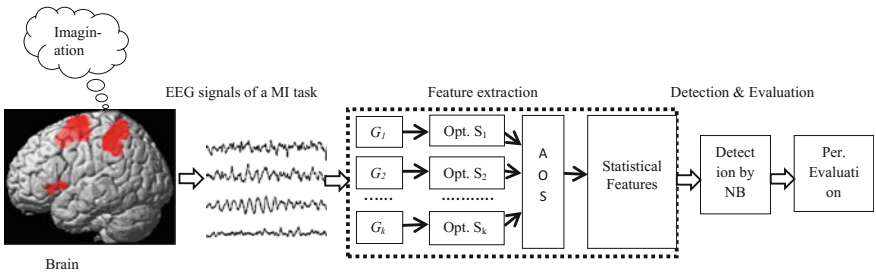


Fig. 12.1 Diagram of the proposed methodology for the detection of MI EEG signals. *Note* G_1 Group 1; G_2 Group 2; G_k Group k; Opt.S_1 = Optimal allocated sample 1; *Opt.*S_2 Optimal allocated sample 2: *Opt.*S_k Optimal allocated sample *k*; *AOS* All of the Optimal allocated samples together from the groups of a class

The first stage is the data acquisition, the second is feature extraction, the third is detection and the final stage is performance evaluation. These stages are discussed in the following sections.

12.2.1 Signal Acquisition

In this study, we used two datasets, IVa and IVb from BCI Competition III (Blankertz et al. 2006; BCI competition III, 2005), which were provided by Fraunhofer FIRST, Intelligent Data Analysis Group (Klaus-Robert Müller Benjamin Blankertz), and Campus Benjamin Franklin of the Charité-University Medicine Berlin, Department of Neurology, Neurophysics Group (Gabriel Curio). The descriptions of both datasets are provided in Chap. 3.

12.2.2 Feature Extraction

This study develops an optimum allocation based approach for feature extraction to identify a suitable representation of the original EEG recordings. The extracted features provide the inter-class discrimination information for detecting different categories or different classes (e.g., right hand movement; right foot movement) of MI tasks. The proposed optimum allocation based approach consists of the steps described below.

12.2.2.1 Data Partition

In this step, all the data for the EEG signals of each category (e.g., right hand movement) of MI tasks is partitioned into various groups to properly account for possible stationarities as signal processing methods require stationarity of signals. Although an overall EEG signal may not be stationary, smaller windows or parts of those signals will usually exhibit stationarity. The partitions of the observations are performed with respect to a specific time period. The time period is determined to view the signals' periodic patterns in each class. In this work, each partition is called 'group' and the groups of data for a particular MI task are denoted as G_1, G_2, ..., G_k as shown in Fig. 12.1. The number of observations of k groups are denoted as N_1, N_2, …, N_k, respectively. It is worth mentioning that the groups must be non-overlapping.

Based on the data structure, we segment the recorded EEG signals of every MI task in each subject into seven ($k = 7$) groups, such as G_1, G_2, ..., G_7 for dataset IVa and into ten ($k = 10$) groups, such as G_1, G_2 ..., G_{10} for dataset IVb. For the RH class of dataset IVa, we get the number of observations for each of the seven groups as 11,627 which means that $N_1 = N_2 = \ldots = N_7 = 11{,}627$, while the RH class holds 81,389 data points of 118 dimensions. For the RF class of the same dataset, we determine that the sizes of each group is 15,689, meaning that $N_1 = N_2 = \ldots. = N_7 = 15{,}689$ while the RF consists of 109,823 observations of the same dimension. For dataset IVb, we get 9743 data points in each of the groups for the LH class, e.g. $N_1 = N_2 = \ldots = N_{10} = 9743$ and 11,065 data points in each of the groups of the RF class, e.g. $N_1 = N_2 = \ldots = N_{10} = 11{,}065$, while the LH class and RF class hold 97,430 and 110,652 data points of 118 dimensions, respectively.

12.2.2.2 Determination of an Optimal Allocated Sample (Opt.S) Size and the Selection of the Opt.S by the Optimum Allocation

This step aims to select a representative sample from every group of a MI task in each subject considering minimum variance. Generally, in a random sampling process, variability is not considered within a group but it is the most important thing to provide sample precision. This study presents optimum allocation to determine the number of observations to be selected from different groups considering minimum variability among the values. If the variability within a group is large, the size of a sample from the group is also large. On the other hand, if the variability of the observations within a group is small, the sample size for that group will be small. Furthermore, this optimum allocation is also used to determine how a given total sample size (denoted as n) for an entire dataset of each MI task in a subject, should be allocated among k groups with the smallest possible variability. To obtain an optimum sample size for each group, we employ Eq. (12.1). In this study, the observation of EEG signals of each MI class (e.g. movement of right hand) is considered as a population. The detailed discussion of the optimum

allocation technique is available in Chap. 6 and also in references (Siuly and Li 2014a; Siuly et al. 2015, 2016).

$$n_i = \frac{N_i \sqrt{\sum_{j=1}^{h} s_{ij}^2}}{\sum_{i=1}^{k} \left(N_i \sqrt{\sum_{j=1}^{h} s_{ij}^2} \right)} \times n;\ i = 1, 2, \ldots, k \text{ and } j = 1, 2, \ldots, h \qquad (12.1)$$

Thus Eq. (12.1) is derived to calculate the best sample size for the *i*th group solving a set of equations by optimum allocation. Using Eq. (12.1), a sample selected from a group of a MI task in a subject is called the 'optimum allocated sample' denoted as Opt.S. All of the Opt.S (s) from the groups of a MI task together makes a matrix called AOS as described in Fig. 12.1. For example: if we select three Opt.S from three groups of a MI class with the sizes OF 10, 12, 11, respectively, then the size of the AOS will be 33. In Eq. (12.1), the total sample size, n is determined (Cochran 1977; Islam 2007) by using Eq. (12.2).

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{PS}} \qquad (12.2)$$

Here, $n_0 = \frac{z^2 \times p \times q}{d^2}$, where n_0 means the initial sample size, z is the standard normal variate (Z-value) for the desired confidence level; p is the assumed proportion in the dataset estimated to have a particular characteristic; q = 1-p and d is the margin of error or the desired level of precision; and PS denotes the population size which considers the total number of data points in a class.

Generally, the sum of all of the Opt.S sizes from all the groups in a MI class should be approximately equal to the total sample size (n) of that class (for example, $n = n_1 + n_2 + \ldots + n_k$) as all the groups come from individual MI classes. Sometimes, the calculated n may be a little larger than the given n due to the rounding figure of the calculated sample size. In this research, we get Z = 2.58 considering 99% confidence level; d = 0.01 for 99–100% confidence interval. If the estimator p is not known, 0.50 (50%) is used as it produces the largest sample size. The larger the sample size, the more sure we can be that their answers truly reflect the entire set of data. Thus, we consider p = 0.50 so that the sample size is the maximum and q = 1-p = 0.50 (50%). With Eq. (12.2), for dataset IVa, we obtain, n = 13816 for the RH class with a data size of 81,389 and n = 14,451 for the RF class with a data size of 109,823. For dataset IVb, we have n = 14,213 for the LH class with a size of 97,430 and n = 14,466 for the RF class with a size of 110,652.

The sizes of the Opt.S (n_i) for each group of every MI class in every subject for datasets IVa and IVb are calculated by Eq. (12.1) as presented in Tables 12.1 and 12.2, respectively. As the number of data points in each of the five subjects in dataset IVa is the same, the calculated sample sizes for each group of every class in Table 12.1 are applicable for every subject. As shown in both tables, the sample

Table 12.1 Calculated sample size by the optimum allocation approach for dataset IVa

Groups	Sizes	Obtained sizes of the Opt.S in each of the seven groups for every two classes	
		RH	RF
G_1	n_1	3786	1702
	N_1	11,627	15,689
G_2	n_2	1895	1473
	N_2	11,627	15,689
G_3	n_3	1674	2360
	N_3	11,627	15,689
G_4	n_4	1567	3945
	N_4	11,627	15,689
G_5	n_5	1344	2429
	N_5	11,627	15,689
G_6	n_6	2150	1476
	N_6	11,627	15,689
G_7	n_7	1401	1067
	N_7	11,627	15,689
AOS	Total n	13,817	14,452
	Total N	81,389	109,823

sizes are not equal in every group in a class, due to a different variability of the observations in different groups. Using the obtained sample size of each group (displayed in Tables 12.1 and 12.2), we select a sample from every group in each class in both datasets. As previously mentioned, a sample selected from a group is called Opt.S and all Opt.S in a class for a subject are integrated and denoted as the AOS set of that class. For example, as shown in Table 12.1, for the RH class of dataset IVa, we obtain seven Opt.S for each subject with the sizes of 3786, 1895, 1674, 1567, 1344, 2150, 1401 (e.g. $n_1 = 3786$, $n_2 = 1895$, $n_3 = 1674$, $n_4 = 1567$, $n_5 = 1344$, $n_6 = 2150$ and $n_7 = 1401$), while they are 1702, 1473, 2360, 3945, 2429, 1476 and 1067 for the RF class. Thus, the AOS set for the RH class and the RF class in every subject of dataset IVa consists of 13,817 and 14,452 observations respectively, as displayed in Table 12.1. Again, for dataset IVb, it can be seen in Table 12.2 that the sizes of the AOS set for the LH class and the RF class are 14,822 and 14,468, respectively. Then these AOS sets are used to extract representative characteristics. Note that in both datasets, the dimension of each AOS set for every subject is 118.

12.2.2.3 Statistical Feature Extraction

Choosing good discriminating features is the key to any successful pattern recognition system. It is usually hard for a BCI system to extract a suitable feature set

Table 12.2 Calculated sample size by the optimum allocation approach for one subject for dataset IVb

Groups	Sizes	Obtained sizes of the Opt.S in each of the seven groups of every two classes	
		LH	RF
G_1	n_1	1219	1506
	N_1	9743	11,065
G_2	n_2	1052	1312
	N_2	9743	11,065
G_3	n_3	2445	924
	N_3	9743	11,065
G_4	n_4	1024	2141
	N_4	9743	11,065
G_5	n_5	1031	1473
	N_5	9743	11,065
G_6	n_6	1922	1705
	N_6	9743	11,065
G_7	n_7	1291	2218
	N_7	9743	11,065
G_8	n_8	1625	869
	N_8	9743	11,065
G_9	n_9	1922	949
	N_9	9743	11,065
G_{10}	n_{10}	1291	1371
	N_{10}	9743	11,065
AOS	Total n	14,822	14,468
	Total N	97,430	110,652

which distils the required inter-class discrimination information in a manner that is robust to various contaminants and distortions. This study considers eleven statistical features: *mean*, *median*, *mode*, *standard deviation*, *maximum*, *minimum*, *first quartile* (Q_1), *third quartile* (Q_3) (75th percentile), *inter-quartile range* (*IQR*), *skewness and kurtosis*. These features are calculated from each AOS set of every class to achieve representative characteristics that ideally contain all the important information of the original signal patterns. The reasons for considering those features are described here. *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. In a tabular form, the *mode* is the value with the highest frequency. *Mean* and *median* are the measures irrespective of data being discrete or continuous, however, the *mode* is most suitable for discrete data but is tricky in the case of continuous data. The *mode* of a continuous probability distribution is defined as the peak of its histogram or density function. *Mean*, *median* and *mode* are the most used features that describe almost all distributions with a reasonable degree of

accuracy (Siuly and Li 2012; Siuly et al. 2016; Cochran 1977; Islam 2004) and provide a fairly good idea about the nature of the data.

Standard deviation gives information about the spread of data or how close the entire set of data is to the average value in the distribution. *Maximum and minimum* values are used to describe the range of observations in the distribution. Q_1 and Q_3, measure how the data is distributed in the two sides of the median. *IQR* is the difference between Q_3 and Q_1 and is used to measure the spread of a data set that excludes most outliers. *Skewness* describes the shape of a distribution that characterizes the degree of asymmetry of a distribution around its mean (Siuly et al. 2014c). *Kurtosis* measures whether the data are peaked or flat relative to a normal distribution.

In this step, we calculate a feature set of eleven features from each AOS set in each class from a subject in both datasets. From every AOS set of each MI class, we acquire a feature vector set of size 118 with 11 dimensions. Thus, we obtain a vector set of size 236 with 11 dimensions for two-class MI data of every subject in datasets IVa and IVb. In each subject, the obtained feature vector set is divided into a training set and a testing set using the 10-fold cross-validation approach. The training set is applied to train a classifier and the testing vectors are used to verify the accuracy and the effectiveness of the classifiers for discriminating MI tasks. In our experiments, the proposed method is separately trained on a single subject in the both datasets, as MI-based EEG signals are naturally highly subject-specific depending on the physical and mental tasks to which they are related. In this research, we present all experimental results from the testing set.

12.2.3 Detection

This study employs the Naive Bayes (NB) classifier to detect two-class MI tasks for the application of BCI systems as it provides a flexible means of dealing with any number of attributes or classes, and it is the fastest learning algorithm that examines all its training inputs. The NB (Mitchel 1997; Wiggins 2011; Richard et al. 2000; Bhattacharyya et al. 2011) is a straightforward and frequently used probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. The NB classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. Depending on the precise nature of the probability model, the NB classifier can be trained very efficiently in a supervised learning setting. In practical applications, the parameter estimation for naive Bayes models uses the method of the maximum likelihood. In this classifier, each class with highest post-probability is addressed as the resulting class.

Suppose, $X = \{X_1, X_2, X_3, \ldots, X_n\}$ is a feature vector set that contains C_k (k = 1, 2, … m) classes' data which must be classified. Each class has a probability $P(C_k)$ that represents the prior probability of detecting a feature into C_k and the values of $P(C_k)$ can be estimated from the training dataset. For the n feature values of X, the

goal of the classification is, to find the conditional probability $P(C_k|\ x_1, x_2, x_3, \ldots, x_n)$. By Bayes's rule NB (Mitchel 1997, Wiggins 2011; Richard et al. 2000; Bhattacharyya et al. 2011), this probability is equivalent to

$$P(C_k|X_1,\ X_2,\ X_3, \ldots,\ X_n) = \frac{P(C_k)P(X_1, X_2, X_3, \ldots X_n|C_k)}{\sum P(C_k)P(X_1, X_2, X_3, \ldots,\ X_n|C_k)} \tag{12.3}$$

Using the chain rule for the reaped application of conditional probability, we have,

$$\begin{aligned} P(C_k, X_1, X_2, X_3, \ldots, X_n) &= P(C_k).P(X_1, X_2, X_3, \ldots, X_n|C_k) \\ &= P(C_k).P(X_1|\ C_k).P(X_2|C_k, X_1).P(X_3|C_k, X_1, X_2)\ldots P(X_n|C_k, X_1, X_2, \ldots X_{n-1}) \end{aligned} \tag{12.4}$$

For the joint probability and for the independent assumption of Naïve Bayes theorem, we get

$$\begin{aligned} P(C_k, X_1, X_2, X_3, \ldots, X_n) &= P(C_k).P(X_1|\ C_k).P(X_2|C_k).P(X_3|C_k)\ldots P(X_n|C_k) \\ &= P(C_k)\prod_{i=1}^{n} P(X_i|C_k) \end{aligned} \tag{12.5}$$

Thus from Eq. (12.3) we have,

$$P(C_k|\ X_1, X_2, X_3, \ldots, X_n) = \frac{P(C_k)\prod_{i=1}^{n} P(X_i|C_k)}{\sum_{j=1}^{k} P(C_j)\prod_{i=1}^{n} P(X_i|C_j)} \tag{12.6}$$

Equation (12.6) is the fundamental equation of the NB classifier. If we are interested only in the most probable value of C_k, then we have the NB classification rule

$$C_k \leftarrow \arg\max_{C_k} \frac{P(C_k)\prod_{i=1}^{n} P(X_i|C_k)}{\sum_{j=1}^{k} P(C_j)\prod_{i=1}^{n} P(X_i|C_j)} \tag{12.7}$$

which simplifies to the following because the denominator does not depend on C_k

$$C_k \leftarrow \arg\max_{C_k} P(C_k)\prod_{i=1}^{n} P(X_i|C_k) \tag{12.8}$$

Thus the NB classifier combines this model with a decision rule. The decision rule for the NB classifier is defined as below:

$$\text{classify}\,(X_1, X_2, \ldots, X_n) = \arg\max_{C_k} p(C_k) \prod_{i=1}^{n} P(X_i|C_k) \tag{12.9}$$

In this work, we use the obtained feature vector set as the input in Eq. (12.9). In the training stage, $P(X_i|C_k)$ is estimated with respect to the training data. In the testing stage, based on the posterior probability $P(C_k|X_i)$, a decision whether a test sample belongs to a class C_k is made. For dataset IVa, C_k (k = 1, 2) is treated as RH = −1 and RF = +1 and for the dataset IVb, C_k (k = 1, 2) is considered as LH = −1 and RF = +1. Thus in this research, we achieve the detection results of each fold for each subject from the both datasets.

12.3 Experiments, Results and Discussions

This section presents experimental outcomes of the proposed optimum allocation based NB approach for two datasets, IVa and IVb of BCI Competition III, and a comparison of the present method with five recently 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. In this research, the experiments of the proposed method are performed separately on one single subject in the both datasets as MI based EEG signals are naturally highly subject-specific depending on the related physical and mental tasks. This study uses a 10-fold cross-validation (Siuly and Li 2012; Siuly et al. 2011b, 2014b) process to assess the performance of the proposed approach using the most standard statistical measures such as accuracy, true positive rate or sensitivity and true negative rate or specificity (see their descriptions in Chap. 3) and also available in references (Siuly et al. 2011a; Gu et al. 2009; Siuly et al. 2013). All of the experimental works of this research are executed in MATLAB (version 7.14, R2012a). In this study, all the experimental results are presented based on the testing set.

12.3.1 Results for BCI III: Dataset IVa

Table 12.3 presents the accuracy for each of the 10 folds and the overall performances over the ten folds for dataset IVa. The overall performances for each subject are reported in terms of mean ± standard deviation (SD) of the accuracy over the ten folds. As shown in Table 12.3, most of the accuracy values for each of the folds are close to 100. The overall performances for Subjects, **aa**, **al**, **av**, **aw** and **ay,** are 97.92, 97.88, 98.26, 94.47 and 93.26%, respectively, and the average of the performances for all of the subjects is 96.36%. Table 12.3 also reports that there is no significant variation in the accuracies among the different folds, indicating the stability of the proposed method.

Table 12.3 Experimental outcomes for the proposed approach for dataset IVa of BCI Competition III

Subject	Accuracy for each of the 10 folds (%)										Overall performance
	1	2	3	4	5	6	7	8	9	10	
aa	83.33	100	100	100	95.83	100	100	100	100	100	97.92 ± 5.29
al	95.83	95.83	100	100	100	95.83	95.83	100	95.45	100	97.88 ± 2.24
av	95.83	100	95.83	100	100	100	100	100	100	90.91	98.26 ± 3.11
aw	95.83	95.83	100	83.33	100	91.67	100	91.67	95.45	90.91	94.47 ± 5.26
ay	79.17	95.83	95.83	87.5	100	100	91.67	91.67	95.45	95.45	93.26 ± 6.24
Total											96.36 ± 2.32

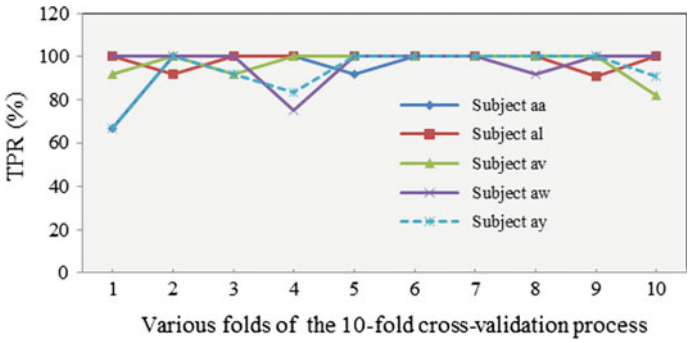


Fig. 12.2 Patterns of the true positive rate (TPR) for each subject in dataset IVa

Figure 12.2 presents the patterns of the true positive rate (TPR) for each subject of dataset IVa. Here the TPR is the correct detection rate for the RH class for this dataset. This figure shows the individual TPR against each of the 10-folds for the five subjects, **aa**, **al**, **av**, **aw** and **ay**. As can be seen in Fig. 12.2, most of the values of the TPR for the proposed approach are close to 100 for each of the folds of each subject, and the variations of the TPR among the 10-folds for each subject is not substantial, thus indicating that the proposed approach is fairly stable.

The contour of the true negative rate (TNR) for each of the five subjects is provided in Fig. 12.3. For dataset IVa, the TNR refers to the correct detection rate for the RF class. This figure displays the separate TNR for each of the ten folds for each of the five subjects. From Fig. 12.3, it is observed that the TNR in most of the folds for each subject is approximately 100%, and there are no significant variations of the TNR among the ten folds for each subject. This indicates that the proposed method is reliable and robust. Along with Table 12.3, Figs. 12.2 and 12.3, it can be concluded that, although there is a little variability in performances over the

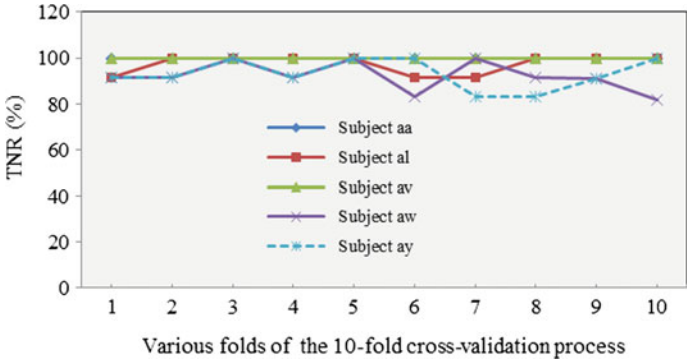


Fig. 12.3 Patterns of the true negative rate (TNR) for each subject in dataset IVa

subjects, generally the proposed approach provides higher performances for all of the subjects, and it is consistent and fairly stable.

12.3.2 Results for BCI III: Dataset IVb

The experimental outcomes for the proposed optimum allocation based NB approach for dataset IVb are presented in Table 12.4. As mentioned in the data description (in Chap. 3), this dataset holds data for one male subject. This table displays the individual accuracy rate for each of the ten folds of that subject and the overall performance of the proposed method in terms of mean ± standard deviation of the accuracy over the ten folds. As shown in the table, the method provides higher accuracy values for most of the folds and the variation among the different folds is not significant. The overall performance for this dataset is 91.97% and the standard deviation is 7.02%.

Figure 12.4 shows the pattern of the TPR of each of the ten folds for a healthy male subject in dataset IVb. Here the TPR is the correct detection rate for the LH class of this dataset. From this figure, it can be seen that most of the values of the TPR lie in approximately 95–100. There are no significant differences in the TPR values among the ten folds, indicating the consistency of the proposed method.

Table 12.4 Experimental outcomes for the proposed approach for dataset IVb of BCI Competition III

Subject	Accuracy for each of the 10 folds (%)										Overall performance
	1	2	3	4	5	6	7	8	9	10	
One healthy male	79.17	83.33	91.67	91.67	95.83	95.83	100	95.83	100	86.36	**91.97 ± 7.02**

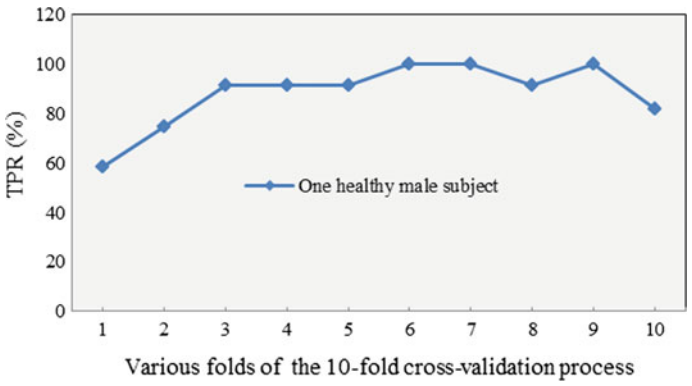


Fig. 12.4 Pattern of the true positive rate (TPR) for dataset IVb

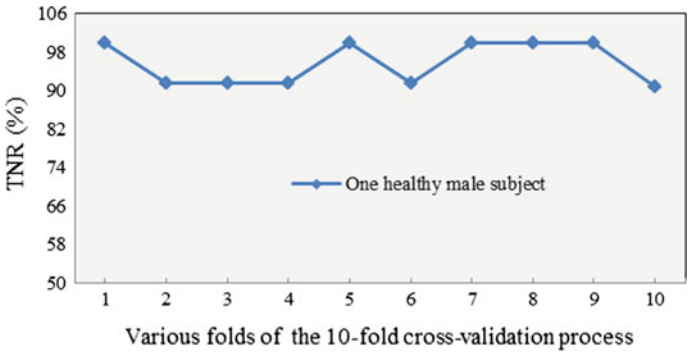


Fig. 12.5 Pattern of the true negative rate (TNR) for dataset IVb

The shape of the TNR for each of the ten folds for one healthy subject is illustrated in Fig. 12.5. Here the TNR refers to the correct detection rate of the RF class for dataset IVb. This figure demonstrates that most of the values of the TNR are close to 100 and the variation among the TNR values of the ten folds is not substantial. This proves the reliability of the proposed approach. Thus, it is obvious from Table 12.4, Figs. 12.4 and 12.5 for dataset IVb, that the proposed algorithm produces a good performance for both individuals and the overall.

12.3.3 Comparison to Previous Work

To further examine the efficiency, this section provides a report for the comparison between the proposed approach and five recently developed methods for dataset IVa. As mentioned before, we cannot present the comparison results for dataset IVb as there were no reported research results available in the literature. Table 12.5 presents the comparison results of the performances for the proposed method and the five existing algorithms for dataset IVa. This table shows the classification performance for each of the five subjects, as well as the overall mean and the SD of the performances for all the subjects. As shown in Table 12.5, the proposed method yields the excellent performances of 97.92 and 93.26% for Subjects, **aa** and **ay**, respectively. The performance of the proposed method for Subject **av** is also very high (98.26%), and this is close to the highest performance (98.75%). The CC based LS-SVM (Siuly and Li 2012) and Z-LDA method (Zhang et al. 2013) provides better results for Subjects **al** and **aw**, respectively. The highest mean for all five subjects is obtained by the proposed approach, which is 96.36%, and the SD value is the lowest (2.32%). Therefore, it can be stated that generally the proposed approach significantly outperforms the five existing methods.

Table 12.5 A comparison report over five most recent reported methods for dataset IVa

Authors	Methods	Detection performance (%)						
		aa	**al**	**av**	**aw**	**ay**	Mean	SD
Proposed approach	OA & NB based approach	97.92	97.88	98.26	94.47	93.26	**96.36**	**2.32**
Suk and Lee (2013)	BSSFO	79.46	94.64	57.65	91.96	53.57	75.46	19.06
Zhang et al. (2013)	Z-LDA	77.7	100.0	68.4	99.60	59.9	81.1	18.2
Siuly and Li (2012)	CC based LS-SVM	97.88	99.17	98.75	93.43	89.36	95.72	4.35
Siuly et al. (2011)	CT based LS-SVM	92.63	84.99	90.77	86.50	86.73	88.32	3.22
Lu et al. (2010)	R-CSP with aggregation	76.80	98.20	74.50	92.90	77.00	83.90	10.86

Looking further into the performance comparison in Table 12.5, it is noted that the proposed algorithm is ranked first in terms of overall performance (96.36%), while the CC based LS-SVM method (Siuly and Li 2012) comes into second position (95.72%), and the CT based LS-SVM algorithm (Siuly et al. 2011) is third (88.32%). The Bayesian spatiospectral filter optimization algorithm (Suk and Lee 2013) is the last (75.46%). The results indicate that the proposed method achieves up to 20.90% improvement over the five existing methods for dataset IVa of BCI competition III.

12.4 Conclusions

There are many challenges in detecting EEG signals of MI activities for the applications of BCIs. In this study, we propose an automatic approach that interprets how EEG signals are organised to detect different categories of MI tasks. Our proposed approach develops an optimum allocation based algorithm to determine representative sample points from every group of the original data considering the minimum variation within each group. Eleven statistical features are extracted from a group of sample points for a particular MI activity. A probabilistic model, NB classifier is employed to detect different MI tasks based on the extracted features. In our experiments on two public databases, IVa and IVb of BCI Competition III, the proposed method outperforms the state-of-the-art methods in terms of the overall detection performances. The adoption of the optimum allocation technique with the NB resulted in an improvement of performance up to 20.90%, compared to the other five reported methods. The performances also show that two-class MI based EEG signals can be reliably identified using the proposed approach.

References

BCI competition III, 2005, http://www.bbci.de/competition/iii.

Bhattacharyya, S. et al. (2011) 'Performance Analysis of Left/Right Hand Movement Classification from EEG Signal by Intelligent Algorithms', *Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB) IEEE Symposium*, 2011.

Blankertz, B, Muller, K. R, Krusierski, D. J, schalk, G, wolpaw, J. R, Schlgl, A, Pfurtscheller, G. and Birbaumer, 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, 153–159.

Cochran, W.G. *Sampling Techniques,* Wiley, New York, 1977.

De Veaux, R. D. Velleman, P.F. and Bock, D.E. *Intro Stats*, 3rd ed., Pearson Addison Wesley, Boston, 2008.

Gu, Q. Zhu, L. and Cai, Z. (2009) 'Evaluation measures of the classification performance of imbalanced data sets', *ISICA 2009, CCIS 51*, pp. 461–471.

Islam, M.N. *An Introduction to Sampling Methods: Theory and Applications,* Book World, Dhaka, 2007.

Islam, M. N. *An introduction to statistics and probability*, 3rd ed., Mullick & brothers, Dhaka New Market, Dhaka-1205, pp. 160–161, 2004.

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, 2936–2945.

Mason, S.G, Birch, G.E. (2003) 'A general framework for brain–computer interface design', *IEEE Trans Neural Syst Rehab Eng. Vol.*11, no.1, 70–85.

Mitchel, T, *Machine Learning*, McGraw-Hill Science, 1997.

Richard, D.G.S, Duda, O, Hart, P.E, *Pattern classification*, 2nd edn. Wiley, New York, 2000.

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, Vol. 20, no. 4, 526–538.

Siuly and Y. Li, (2014a) 'A novel statistical framework for multiclass EEG signal classification', *Engineering Applications of Artificial Intelligence*, Vol. 34,154–167.

Siuly, Li, Y. and Wen, P., (2011a) '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, 390–409.

Siuly, Li, Y. and Wen P. (2011b) 'Clustering technique-based least square support vector machine for EEG signal classification', *Computer Methods and Programs in Biomedicine*, Vol. 104, 358–372.

Siuly, Li, Y. and Wen, P. (2013) 'Detection of Motor Imagery Tasks through CC-LR Algorithm in Brain Computer Interface', *International Journal of Bioinformatics Research and Applications*, Vol. 9, no. 2, 156–172.

Siuly, Li, Y. and Wen, P. (2014b) 'Comparisons between Motor Area EEG and all-Channels EEG for Two Algorithms in Motor Imagery Task Classification', *Biomedical Engineering: Applications, Basis and Communications (BME)*, Vol. 26, no. 3, 1450040 (10 pages).

Siuly, Y. Li and P. Wen, (2014c) 'Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain computer interface', *Computer Methods and programs in Biomedicine*, Vol. 113, no. 3, 767–780.

Siuly and Y. Li, (2015), 'Discriminating the brain activities for brain–computer interface applications through the optimal allocation-based approach', *Neural Computing & Applications*, Vol. 26, Issue 4, pp 799–811.

Siuly, H. Wang and Y. Zhang (2016), 'Detection of motor imagery EEG signal employing Naive Bayes based learning process', *Measurement* 86, 148–158.

Suk, H. and Lee, S.W. (2013) 'A Novel Bayesian framework for Discriminative Feature Extraction in Brain-Computer Interfaces', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, no. 2, 286–299.

Wiggins, M. Saad, A. Litt, B. and Vachtsevanos, G. (2011) 'Evolving a Bayesian Classifier for ECG-based Age classification in Medical Applications', *Appl Soft Comput*, Vol. 8, no. 1, 599–608.

Zhang, R. Xu, P. Guo, L. Zhang, Y. Li, P. and Yao, D. (2013) 'Z-Score Linear Discriminant Analysis for EEG Based Brain-Computer Interfaces', *PLoS ONE*, Vol. 8, no. 9, e74433.

Part IV
Discussions, Future Directions and Conclusions

Chapter 13
Summary Discussion on the Methods, Future Directions and Conclusions

In this book, we intended to develop some computer-aided diagnostic methods for the analysis and classification of EEG signals, especially focused on *the diagnosis of epilepsy* and *the recognition of mental states for BCI applications*. All proposed methods are tested on several real-time EEG databases and their results give a valuable contribution to the understanding of brain dynamics. This chapter provides a summary discussion on each of the developed methods along with their findings. Furthermore, this chapter provides concluding remarks and suggestions for further research.

13.1 Discussion on Developed Methods and Outcomes

The EEG is an important measurement of brain activity and has great potential in helping 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 the identification and evaluation of brain activity. Identifying various types of EEG signals is a complicated task, requiring the analysis of large sets of EEG data. Representative features from a large dataset play an important role in classifying EEG signals in the field of biomedical signal processing. In this book, we studied and developed EEG signal processing and classification techniques to identify different types of EEG signals with three main objectives:

- Develop methods for the classification of epileptic EEG signals to improve the classification rate
- Introduce methods to identify EEG signals during MI tasks for the development of BCI systems
- Investigate which algorithm and which EEG data (motor area data or the all-channels data) is better for MI signal classification.

S. Siuly et al., *EEG Signal Analysis and Classification*,
Health Information Science, DOI 10.1007/978-3-319-47653-7_13

To achieve these objectives, we first developed four methods; *simple sampling technique based least square support vector machine* (*SRS-LS-SVM*), *clustering technique based least square support vector machine* (*CT-LS-SVM*), *optimum allocation* (*OA*) *technique based methodology* and *injecting PCA in the OA scheme* to contribute to epileptic EEG signal classification.

In the SRS-LS-SVM method (Siuly et al. 2011a), we introduced a simple sampling (SRS) technique in two stages for the feature extraction process (see Chap. 4). 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 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" a considering 99–100% confidence interval and a 99% confidence level. We employed the LS-SVM with the RBF kernel for classification where these features were employed as the inputs. First, we implemented this method to the EEG epileptic data (EEG time series 2005) to classify epileptic signals. In order to supplementary test the effectiveness of the method, we employed this method to the mental imagery task EEG data (Chiappa and Millán 2005) for the classification of different categories of 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 task 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. To further test the efficacy of the method on non-EEG data, we also applied this method to 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 the LS-SVM, which can be used as a new intelligent diagnostic system.

To reduce the experimental time and to improve the classification performance, we developed the CT-LS-SVM (Siuly et al. 2011b) for epileptic EEG signals classification (see Chap. 5). In this method, we proposed the clustering technique (CT) approach as a new process for feature extraction from 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 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 the method uses all data points for

the experiments. The second advantage is that, using the CT technique, much less time is taken to run the program. We evaluated the performance of this method through the tenfold cross-validation procedure.

The performance of the CT-LS-SVM algorithm was compared, in terms of 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 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 with less computational complexity. The CT-LS-SVM algorithm can help to provide clinical information about patients who have a neurological disorder, and mental or physiological problems.

The sample size is an important component of an empirical study's ability to obtain results that reflect the original population. Chapter 6 presents a new method for determining an appropriate sample size from each time-window, considering the variability of the values, to be reliable enough to meet the objectives of a study. In this book, we developed an innovative OA technique (Siuly and Li 2014a) for selecting representative sample points from each epoch considering the variability of the observations in the epileptic signal classification. In the proposed method, we first calculate an appropriate sample size for each class from the whole EEG data using the "*Sample size calculator*", with a desired confidence interval and confidence level. Second, the data of each class is segmented into different epochs, considering a specific time period. Third, the OA technique is employed to determine the best sample size for each epoch with a minimal variance. Fourth, the samples are selected from the epochs considering the size that is obtained by the OA procedure. Finally, these samples obtained from all epochs in each class are used as the input to the multiclass LS-SVM (MLS-SVM) with the RBF kernel. Four output coding approaches: MOC, ECOC, 1vs1, and 1vsA are applied in the MLS-SVM and their individual effectiveness is investigated. Before the classification process, the parameter values ($\gamma = 10$ and $\sigma^2 = 100$) of the MLS-SVM method are determined after an extensive experimental evaluation. To examine the consistency of the method, the experiments of the proposed algorithm are repeated 10 times for each of the four output coding approaches with the selected optimal parameter values. The experimental results show that our developed algorithm is very consistent in each repetition, and yields very high classification performances for each of the four output coding approaches. There are no significant differences among the performances by the MOC, ECOC, 1vs1, and 1vsA approaches. This research leads us to confirm that the OA is reliable for capturing valuable information from the original EEG data and the MLS-SVM is very promising for the classification of multiclass EEG signals.

In order to reduce the dimension of the obtained sample set, we develop another new approach injecting PCA in the OA scheme (Siuly et al. 2015) in the epileptic seizure detection from multiclass EEG signal data in Chap. 7. The PCA method is employed on the OA sample set to produce uncorrelated variables and also to reduce the dimensionality because there is a possibility of much correlation among

the brain electrical activities through the brain volume, and correlation may exist among different channels. The obtained principal components are treated as features in this study, and called OA_PCA feature set. To identify an efficient classifier for the OA_PCA feature set, we employ four prominent classifiers: MLS-SVM, NB, KNN, and LDA. To further evaluate the performances, we compare our proposed methods with other existing algorithms. The experimental results show that the LS-SVM_1v1 approach yields 100% overall classification accuracy, improving up to 7.10% over the existing algorithms for epileptic EEG data.

Concerning MI task classification, we further developed four algorithms: *Cross-correlation based logistic regression* (*CC-LR*); *Modified CC-LR with diverse feature sets*; *Cross-correlation based least square support vector machine* (*CC-LS-SVM*); *OA aided Naïve Bayes approach*. 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 four methods were tested on two benchmarks datasets, IVa and IVb of BCI Competition III (BCI competition III 2005; Blankertz et al. 2006). In both datasets, each subject was considered separately for experiments as MI tasks' EEG signals are naturally highly subject-specific, depending on physical and mental tasks being performed.

In the CC-LR algorithm (Siuly et al. 2013), we have combined the cross-correlation (CC) feature extraction and the logistic regression (LR) classification to identify MI tasks in BCI applications (in Chap. 8). In this algorithm, the CC technique follows three steps to extract features from MI task 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. Then, 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. Next, 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 threefold 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 of MI tasks, and can provide a positive contribution to the development of BCI systems.

To select a more suitable feature set to enhance the classification performance and give more reliable results for BCI systems, we modified the CC-LR algorithm with three diverse feature sets (see Chap. 9). In the modified CC-LR method (Siuly et al. 2014b), we provided 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 of 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 threefold cross-validation method. This study finally 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 task EEG data and the 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 task 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 is used for feature extraction and the LS-SVM classifier for the classification of the extracted features (see Chap. 10). 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 both datasets to report on the performance (Siuly and Li 2012). 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, *interquartile range* (*IQR*), 1⁄4 *percentile* (P_{25}) and 3⁄4 *percentile* (P_{75}), to our existing six feature set. From the experimental results, it was seen that there is no significant difference in performance between the existing six feature 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 tenfold 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. 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 does 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 MI signal classification in BCI applications. This study concluded that the CC based LS-SVM algorithm is a

promising technique for MI signal recognition and it offers great potential for the development of MI-based BCI analyses which will assist clinical diagnoses and rehabilitation tasks.

Finally, in Chap. 11, we investigated two issues for MI task based EEG signal classification. First, we sought to discover which algorithm performs better, and second which EEG data is more suitable for obtaining information about the MI. To answer these two questions, we applied three algorithms: the CC-LS-SVM, the CC-LR and the CC-KLR. These three algorithms were implemented on motor area EEG data and all-channels EEG data to investigate how well they performed, and to test which EEG area is better for 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: C5, C3, C1, C2, C4, C6, CP5, CP3, CP1, CP2, CP4, CP6, P5, P3, P1, P2, P4, and P6. Wang and James (2007) also considered the same electrodes for their research and their experimental results suggested that these electrodes are the best channels for obtaining MI information. These algorithms were also compared with some existing methods, which revealed their competitive performance in classification. In this algorithm (Siuly et al. 2014c), we used the C3 channel as the reference channel and employed the threefold cross-validation procedure for the evaluation of the performance. The results on 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 these results, it can be concluded that the CC-LS-SVM algorithm is the best algorithm for MI EEG signal classification and the all-channels EEG can provide better information than the motor area EEG for MI classification. Furthermore, the LS-SVM based approach can correctly identify discriminative MI tasks, demonstrating the algorithm's superiority in classification performance over most of the existing methods.

To improve performance, in Chap. 12, we presented a synthesis approach based on an optimum allocation system (OA) aided Naive Bayes (NB) probabilistic model (Siuly et al. 2016) for detecting mental states based on EEG signals. The proposed approach was developed to determine representatives sample points from every group of the original data considering the minimum variation within each group. Then eleven statistical features: *mean*, *median*, *mode*, *standard deviation*, *maximum*, *minimum*, *first quartile* (Q_1), *third quartile* (Q_3) (75th *percentile*), *interquartile range* (*IQR*), *skewness and kurtosis*, are extracted from a group of samples points for a particular MI activity. After that, the NB classifier is employed to detect different MI tasks based on extracted features. The experimental results on two public databases, IVa and IVb of BCI Competition III, demonstrate that the proposed method outperforms the state-of-the-art methods in terms of overall classification performance. The adoption of the OA technique with the NB resulted in an improvement of performance up to 20.90% compared to the other five reported methods.

13.2 Future Directions

EEG signal processing is a very difficult task, due to the noise, non-stationarity, complexity of the signals and the limited amount of training data available. Furthermore, the existing tools are still not perfect, and many research challenges are remain open. To facilitate further development in this research area, the authors have emphasized a few key below

- **Removal of artefacts for the advanced classification**

EEG brain signals often contain unwanted signals (artefacts) which may bias the analysis of the signals, and lead to incorrect conclusions. Although the area of artefact removal is ripe, very few methods have been employed to eliminate artefacts. Most of the investigations performed to date, have been performed offline due to the computational overheads. Artefacts are, however, a dramatic problem in real-life scenarios. Thus, further investigations are necessary to enable the removal of artefacts without contaminating the EEG signals. Therefore, before going to the feature extraction stage, removal of artefacts (called noise) is essential to have clean classification results. This requires a significant signal processing work to develop the appropriate real-time algorithms.

- **Advance features extraction algorithms**

Extracting representative features from EEG signals is a very difficult task due to the noise, non-stationarity, and complexity of the signals. Thus, it is essential to explore and design EEG features that are more *informative*, in order to obtain better performances. For example, in BCIs, it is important to derive signal features to robustly identify and distinguish the command-related activities because the recorded signals change over time due to technical, biological and psychological factors.

- **Improve machine learning techniques**

To achieve robust performances, improvements in machine learning algorithms are also necessary for excellent classification and diagnosis. In the literature, it has been observed that in most of the cases, the methods suffer from a trade-off between accuracy and efficiency. It is, therefore, necessary to be improved online as the experts require online computer-aided diagnosis (CAD) systems for real-time evaluation.

- **Diagnosis as an interdisciplinary challenge**

Research into the diagnosis of brain diseases/disorders is a multidisciplinary endeavour that requires a broad base of skills from several fields such as neuroscience, clinical rehabilitation, psychology, computer science, engineering and mathematics. To expedite the development of the technology, it requires successful collaborations between multiple disciplines.

- **Extension of goals to promote diverse application development**

The current BCI technology aims to develop a replacement communication medium for severely disabled people. If the systems can manage to improve the bit-rates, applications targeted at the broad population, such as multimedia, behaviour monitoring and video gaming applications may come to fruition. It is important that the community distinguishes between the goal of aiding people with disabilities and the development of BCI systems and related applications. The main direction for BCI research, for the present and near future, will remain the improvement of the communication and control capabilities for people who have severe neuromuscular disorders.

13.3 Conclusions and Further Research

13.3.1 Conclusions

The aim of this monograph is to introduce enhanced methodologies for the automatic analysis and classification of different categories of EEG signals. This book presents new methodologies combining statistical concepts and machine learning techniques which are quite different from the existing methods. In this book, we especially focus on the epileptic seizure detection and mental states recognition for BCI applications using EEG signal data. We developed a total of eight methodologies for EEG signal analysis and classification; four for epileptic seizures detection and another four for mental state recognition for BCI applications. We tested the proposed methods on several real-time databases. The results imply that the classification systems developed in this book are very promising and also are very helpful to efficiently manage and analyze the large size of EEG data. Taken together, it can be concluded that the research presented in this book provides new and successful methods for the reliable classification of EEG signals. These techniques will enable experts to diagnose brain degenerative diseases correctly and efficiently, and will also be 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. In addition, our proposed methods will also assist the experts with an online CAD tool for analyzing EEG signals.

13.3.2 Further Research

We believe that the methods presented in this book will provide promising outcomes in the EEG signal processing and classification areas. Extensive future work will examine the possibility of using the methods in the applications of EEG signal

classification. To facilitate further developments of those proposed methods, we have highlighted a few key issues which are addressed below.

Concerning the SRS-LS-SVM algorithm, it would be beneficial to study the distribution of different random samples, sub-samples and populations (whole EEG channel data) using a hypothesis testing procedure. In this process, first the distribution of the random samples and sub-samples in each EEG channel data will be tested to discover whether they are homogeneous or not. If they are homogenous, then sub-samples could be considered as representatives of the random samples, meaning the sub-samples could be used for feature extraction instead of the random samples. Then, the distribution of the sub-samples would be compared with the distribution of a 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 the best representatives for exploring the original EEG signals.

The CT-LS-SVM method can be improved following the same process as the SRS-LS-SVM algorithm. In addition, the SRS-LS-SVM and CT-LS-SVM algorithms can be improved by tuning hyper parameters of the LS-SVM method through a two-step grid search technique. These two algorithms could be extended for multiclass EEG signal classification. For the CC-LR and modified CC-LR algorithms; these can be developed using multivariate logistic regression instead of binary logistic regression for multiclass classification, and then compared with the kernel logistic regression.

As the frequency bands of EEG signals are important key issues in identifying EEG characteristics, in the near future, the SRS-LS-SVM, CT-LS-SVM, the CC-LR, modified CC-LR and CC-LS-SVM algorithms can be considered for classification based on frequency bands. To 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. In this book, our proposed methods were implemented on noise free EEG databases. We did not employ any methods for removing noise from the EEG data. Further studies are needed to successfully remove artefacts without contaminating the EEG signals for our proposed algorithms.

References

BCI competition III, 2005, http://www.bbci.de/competition/iii.

B. Blankertz, K. R. Muller, D. J. Krusierski, G. schalk, J. R. wolpaw, A. Schlgl, G. Pfurtscheller, and N. Birbaumer, "The BCI competition III: validating alternative approaches to actual BCI problems," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.

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.

EEG time series, 2005, [Online], Viewed 30 September 2008, http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html.

Ripley (1996), http://www.stats.ox.ac.uk/pub/PRNN/.

Siuly, Y. Li, and P. Wen, (2011a) '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, Y. Li, and P. Wen, (2011b) '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.

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*, 20(4), pp. 526–538.

Siuly, Y. Li, and P. Wen, (2013) 'Identification of Motor Imagery Tasks through CC-LR Algorithm in Brain Computer Interface', *International Journal of Bioinformatics Research and Applications*, Vol. 9, no. 2, pp. 156–172.

Siuly and Y. Li, (2014a), 'A novel statistical framework for multiclass EEG signal classification', *Engineering Applications of Artificial Intelligence*, Vol. 34, pp. 154–167.

Siuly, Y. Li, and P. Wen, (2014b) 'Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain computer interface', *Computer Methods and programs in Biomedicine*, Vol. 113, no. 3, pp. 767–780.

Siuly, Y. Li, and P. Wen, (2014c) 'Comparisons between Motor Area EEG and all-Channels EEG for Two Algorithms in Motor Imagery Task Classification', *Biomedical Engineering: Applications*, *Basis and Communications* (*BME*), Vol. 26, no. 3, pp. 1450040 (10 pages).

Siuly and Y. Li, (2015), 'Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification', *Computer Methods and programs in Biomedicine*, *Computer Methods and programs in Biomedicine*, Vol. 119, pp. 29–42.

Siuly, H. Wang and Y. Zhang (2016), 'Detection of motor imagery EEG signal employing Naive Bayes based learning process', *Measurement* 86, 148–158.

Wang, S. and C. J. James (2007) 'Extracting rhythmic brain activity for brain-computer interfacing through constrained independent component analysis', *Computational Intelligence and Neuroscience*, 41468, 9. doi:10.1155/2007/41468.