Development of Predictive Algorithms for Electrical Stimulation Based Major Depressive Disorder Therapy

By
Alaa Mohammed Al-Kaysi

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

School of Biomedical Engineering,
Faculty of Engineering and Information Technology,

University of Technology Sydney

June 2018

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as part of the collaborative doctoral degree and/or fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Production Note:

Signed: Signature removed prior to publication.

Date: 20/06/2018

Acknowledgements

Foremost, my sincere thanks to God, who endowed me to complete this PhD research.

I would like to express my deepest gratitude and appreciation to my supervisor, Dr. Ahmed Al-Ani, for providing an invaluable effort of unwavering support, constant guidance and patient revision, have all been crucial in the realisation of this dissertation.

My gratitude is also to my external supervisor, Dr Tjeerd Boonstra, for providing scientific advice on the research's experiment, discussions and constant perusal of my work. Prof. Colleen Loo and Dr Tjeerd Boonstra are supported me the data that were used in this research in collaboration with Black Dog Institution's team, thank you for all the hours of collaboration.

Special thanks to my parents in Iraq. They provided me with a fine education, continuous support, prayers, as well as ceaseless encouragement in my research. I wish to also thank my best friends Yamama Shafiq and Sara Sami, for their encouragement during my research.

I would also like to acknowledge the Iraqi Government and the University of Technology-Baghdad, I acknowledge their support by granting me the scholarship to finish my PhD degree.

I would also thank Dr. Rifai Chai in centre of health technology and my colleagues at University of Technology Sydney, in particular Dr. Karthick Thiyagarajan.

Finally, I would like to thank the University of Technology Sydney.

Contents

List of Figures

List of Tables

Acronyms

AIM: Alpha inactivity mechanism

ANOVA: One-way analysis of variance

BCI: Brain computer interface

BM: Boltzmann machine

CRBM: Convolutional restricted Boltzmann machine

CDBN: Convolutional deep belief network

CD: Contrastive divergence

CGI: Clinical global impression

CDBNs: Convolutional deep belief networks

CNN: Convolutional neural network

CRBM: Convolutional restricted Boltzmann Machine

DBN: Deep belief network

DBNs: Deep belief networks

DCNs: Deep convex networks

DNN: Deep neural network

DLPFC: Dorsolateral prefrontal cortex

ECG: Electrocardiogram

ECT: Electroconvulsive therapy

EEG: Electroencephalography or Electroencephalogram

ELM: Extreme learning machine

EMG: Electromyography

EOG: Electrooculography

ERPs: Event-related potentials

FIS: Fuzzy inference system

KNN: K-nearest neighbours

ICA: Independent component analysis

LDA: Linear discriminant analysis

LDAA: left-domain alpha asymmetry

MCHDBN: Multichannel deep delief network

MCMC: Markov chains monte carlo

MDD: Major depressive disorder

MDBN: Multimodal deep belief network

MDRS: Montgomery-Åsberg depression rating scale

MLP: Multi-layer perceptron

MLPs: Multi-layer perceptron neural networks

MIMO: Multi-input and multi-output

MPH: Minimum peak hight *(MinPeakHeight)*

MPP: Minimum peak prominence *(MinPeakProminence*

PSD: Power spectral density

PSO: Particle swarm optimisation

QIDS-SR: Quick inventory of depressive symptomatology

RBM: Restricted Boltzmann machine

SD: Standard deviation

SDMT: Symbol digit modalities test

SIFT: Scale-invariant feature transform

sMRI: Structural magnetic resonance imaging

SVM: Suport vector machine

tDCS: Transcranial direct current stimulation

TMS: Transcranial magnetic stimulation

rTMS: repetitive Transcranial magnetic stimulation

TSLOW: Time to onset of slowing

WHO: World health organization

WE: Wavelet entropy

WT: Wavelet transform

Abstract

Major depressive disorder (MDD) is a brain disorder that is characterised by negative thoughts, mood and behaviour. Transcranial direct current stimulation (tDCS) has emerged recently as a promising brain-stimulation treatment for MDD. A standard tDCS treatment involves numerous sessions that are run over a few weeks, however, not all participants respond to this type of treatment. This delay could have negative impact upon patients that do not respond, being an inefficient use of staff time and exposing patients to ineffective treatment. The early identification of patients who respond to this type of treatment is needed. Electroencephalography (EEG) signal is a significant tool that can be used to study the modulatory effects of tDCS treatment. The significant part of this research aims to predict the clinical outcomes of tDCS treatment by analysing the patients' EEG signals.

EEG signal is a complex signal that has high sensitivity to noise. EEG signal records both neural and non-neural activities from a large number of electrodes, therefore, the analysis and classification of EEG signals proved to be quite challenging. Machine learning has attracted the attention of many researches as a powerful approach for analysing various types of signals, including EEG. Algorithms for channel/feature selection, classification, detection, prediction and fusion have been developed. Recently, deep neural networks, particularly deep belief networks (DBNs), have emerged as a new hierarchical technique for modelling high level abstractions of data, and have been successfully applied to a number of classification problems. Similar to most classification algorithms, the existing DBNs have been mainly designed to handle single stream data, and there are hardly any attempts to generalize those to suit multi-channel signals.

Accordingly, the first part of this research investigates the utilisation of DBN to differentiate between tDCS sessions based on classification EEG signals, particularly the implementation of multi-channel DBNs. One of the important attributes that needs to be carefully studied is considering the multi-channel nature of EEG in the design and training of deep networks.

A second part of this research aims to predict which patients improve in mood and cognitive in response to tDCS treatment based on EEG data that were collected at the start of tDCS treatment. Classifying power spectral density (PSD) of resting-state EEG is achieved using support vector machine (SVM), linear discriminate analysis (LDA) and extreme learning machine (ELM). Participants were labelled as improved/not improved based on the change in mood and cognitive scores. The obtained classification results of all channel pair combinations are used to identify the most relevant brain regions. The frontal area is found to be particularly informative for the prediction of the clinical outcome of the tDCS treatment. Subject independent results reveal that our proposed method enables the correct identification of the treatment outcome for seven of the ten participants for mood improvement and nine of ten participants for cognitive improvement. This represents an encouraging sign that EEG-based classification may help to tailor the selection of patients for treatment with tDCS brain stimulation.

The second line treatment of depressive disorder is electroconvulsive therapy (ECT). ECT is an effective and widely used treatment for major depressive disorder, in which a brief electric current is passed through the brain to trigger a brief seizure. The second main aim in this research is to identify seizure quality rating by utilising a set of seizure parameters. Four seizure related parameters, (time to onset of slowing, regularity, stereotypy and post-ictal suppression) are used as inputs to decision tree and fuzzy rule-based classifiers to predict seizure quality ratings. The classification results show that the four seizure parameters provide relevant information about the rating of seizure quality. Automatic scoring of seizure quality could be beneficial to clinicians working in electroconvulsive therapy.

Chapter 1

Introduction

Depression is a very significant mental disorder, affecting many people around the world. Depression is expected to be the second leading cause of disability in the world by 2020 [1]. Depression is described by a group of symptoms such as sadness, insomnia, anxiety, lethargy and, in critical cases, suicidal tendencies. Depression can be developed based on many factors such as psychological, social and environmental risks. There are several types of depression depending on the risk degree of symptoms, such as major depressive disorder (MDD), psychotic depression and persistent depressive disorder. Recently, transcranial direct current stimulation (tDCS) has emerged as a neuro-stimulation treatment for major depressive disorder (MDD) [2, 3]. tDCS is a safe and non-invasive treatment method that does not require anaesthesia.

The second line treatment for major depressive disorder is electroconvulsive therapy (ECT). ECT involves an electrical pulse that is applied to the scalp under short term anaesthesia. These pulses excite the brain's cells, which produce a seizure [4]. In general, the seizure is detected from the sudden surge or uncontrolled electrical activity in the brain, minor physical signs or a series of symptoms. A seizure is defined as subclinical when the electroencephalography (EEG) pattern is changed in a characteristic way without changing the patient's behaviour [5].

Depression is a highly disabling mental illness that can be diagnosed by analysing EEG signals [6]. An EEG is the best tool for analysing changes in cortical activity following tDCS [7, 8]. The behaviour of each EEG channel in this process needs to be studied. Recently, machine learning and computational intelligent techniques have come to play an important role in medicine, including in the classification, prediction and diagnosis of clinical data. Deep belief network (DBN) is a new technique of deep learning that can deal with time series data. A deep belief network is a probabilistic, generative model which is constructed of multiple layers of hidden units. DBNs are multilayer perceptron neural networks (MLPs) with many hidden layers; however, the network weights in DBNs are initialised using a greedy layer-by-layer pre-training algorithm. The greedy layer-by-layer training algorithm is significant for DBN and is used for learning in a deep way to build a hierarchical probabilistic model [9].

This chapter introduces the motivation and the problem statement of this research study. It will then describe the main objectives of this research and explains the principle contributions of this dissertation. Finally, to simplify the reading of the dissertation, this chapter will outline its structure.

1.1 Research Motivation

Major depressive disorder (MDD) is one of the most widespread mental illnesses around the world. It is considered a serious clinical syndrome as it can have fatal consequences. MDD is characterised by negative thoughts, mood and behaviour. The symptoms of MDD can vary based on age, gender and ethnicity. Treatment of MDD has been investigated by many mental health professionals and various types of medications have been developed to treat the disease. However, despite many noticeable advancements, not all patients respond to antidepressant medications and, therefore, other options have to be considered [10–12].

Recently, transcranial direct current stimulation (tDCS) has attracted increasing attention from psychology and neuroscience researchers. tDCS delivers a weak electric current through the scalp, which induces shifts in neuronal membranes that cause secondary changes in the cortical activity. A number of tDCS sessions are usually required, and therefore a long time could be needed to determine the clinical outcome; that is, whether a patient responds to the treatment or not. As this delay could have a negative impact on the patient and it consumes staff's time and resources, there is a need for the early identification of the patients who respond to this type of treatment from those who don't.

Another type of treatment for depressive disorder is ECT, particularly for patients who do not respond to antidepressant medication. ECT is the first choice treatment in critical cases, such as suicidal patients and patients who have a poor response to medication [13, 14]. There are many side effects of ECT treatment, such as headaches, nausea (queasiness), sore muscles and disruption (confusion). However, due to improvements in ECT, it is a safe and effective treatment for many psychiatric disorders [15]. Despite these improvements, the neurobiological mechanisms of therapeutic efficacy are still unclear. The seizure quality represents a useful tool for deciding the appropriate therapeutic treatment and is relevant for the clinical outcome [16, 17].

Mirowski et al. focused on the differentiation of brainwave synchronisation patterns in three states: inter-ictal, pre-ictal and ictal in order to complete the understanding of neurological brain states [18]. Long term EEG monitoring is the best technique for diagnosing seizures in order to evaluate the patient's state [19]. Tracking the seizure activity by diagnosing EEG signals in order to recognise the onset of seizures helps us to understand how seizures occur [20, 21]. Several parameters can affect the efficacy of ECTs, such as seizure duration, electric stimuli and seizure threshold.

Electroencephalography involves the recording, analysis and physiological interpretation of the electrical activity which is generated by large number of neurons and recorded through a number of electrodes (32, 64, 128 or 256) that are placed on the human scalp, where these electrodes aim at capturing the average neuronal activity [22]. The EEG signal is very sensitive to noise, as the electrical potential that is recorded on the scalp is very small (5-10 μV) [23]. Given the complex nature of EEG data, the analysis and classification of these signals is quite challenging.

Electroencephalograms (EEGs) play a vital role in analysing the chemical activities in the human brain and understanding the functional state of mental disease, especially for diagnosing disorders [24]. A clinical scalp EEG can contribute to understanding the complex inner mechanisms of the brain [25]. However, EEG signals are usually contaminated by artefacts, which are non-neural electrical activities from non-physiological and physiological sources [26].

In order to deal with the complex data, such as EEG data, machine learning techniques that have the ability to discover hidden rules in data and the capacity to handle missing data should be used. This is essential to help the decision-making process, especially in the medical field.

1.2 Research Problem

Transcranial direct current stimulation is a significant treatment for depressive disorder [7, 27, 28]. The antidepressant effects of the tDCS treatment are still not very well understood for the neurophysiological mechanisms [7]. The EEG signal has the ability to study the neuromodulatory effects of this type of treatment.

When dealing with complex data such as an EEG, it is reasonable to use machine learning techniques that have the ability to evaluate the behaviour of EEG channels based on their performance in classification tasks. Several techniques of machine learning have been applied to the classification of EEG data such as support vector machine (SVM), linear discriminate analysis (LDA), extreme learning machine (ELM) and Bayesian classifiers. However, most of these techniques have not been specifically designed to suit the nature of EEG waveforms. In recent years, deep learning has been introduced as a machine learning technique that uses multiple learning layers to model the complex and abstract representations of many real-life signals [29, 30].

In general, a deep learning algorithm is used to solve non-linear transformation issues. However, it has two main problems: a slow learning time and over-fitting. Deep neural networks (DNNs) require proper training mechanisms and the estimation of a number of hyper-parameters such as the number of visible and hidden layers, the number of units per layer, the learning rate, the momentum and the number of iterations [31]. Hinton et al. proposed a new fast learning method for deep generative models called deep belief networks (DBNs) that can overcome the complexity of training deep networks [9, 32] and has the ability to learn high-dimensional manifolds of the data [33].

Recent developments of DBNs have been successfully applied to a broad range of classification problems, such as image classification, speech recognition and natural language processing. Deep belief networks were proposed to reduce the training time and solve the over-fitting problem by pre-training DBNs in both unsupervised and supervised manners [32, 34]. DBNs have mainly been applied to image classification due to the hierarchical structure of images, where edges can be grouped to form segments, which, when grouped, form objects [35]. They have also been successfully applied to a number of other classification tasks, such as speech recognition and natural language processing [35]. Although DBN is a relatively new technique of deep neural networks that have the ability to deal with time series data such as EEG data, DBNs have not been fully explored for the classification of EEG data.

The second treatment of depressive disorder is electroconvulsive therapy. Based on the nature of ECT-induced seizure recordings, it is important to use interpretability techniques to identify the seizure quality. This facilitates the decision for experts as to whether the suggested treatment is suitable or not. In addition to this, a fuzzy set has the ability to deal with the uncertainty of biomedical systems through building rules [36]. The selection of good identification tools is very important to support medical systems and help clinicians make the right decisions. The best proposed methods for this purpose are decision tree and fuzzy rule-based classifiers.

After identifying the problems that associated with tDCS and ECT treatment based on EEG signals, it is obvious that the dealing with the complex nature of EEG signal is a common problem with both treatments. In order to improve the predictive performance based on depressive disorder therapy, three main objectives guiding this dissertation which will describe in the next section.

1.3 Research Objectives

The aim of this research is to improve and utilise machine learning techniques to predict the clinical outcomes of patients with major depressive disorder based on tDCS and ECT treatments. Correspondingly, the specific objectives of the dissertation are as follows:

The first objective is to improve a special type of deep belief networks for enabling it to deal with multi-channel EEG signals. The proposed deep belief network utilises the restricted Boltzmann machine (RBM) technique to classify and predict the antidepressant responses of patients. Accordingly, the utilisation of multi-channel deep belief network in distinguishing between the three EEG sessions of baseline, active tDCS and sham tDCS is investigated. The performance of the proposed method is compared to three standard machine learning classification techniques.

The second objective is to propose a method for predicting "How the patients will respond to the tDCS treatment?". In particular, analysing the neurophysiological effect of neuro-stimulation through identifying the relevant brain regions and EEG rhythms that are influenced by the tDCS treatment. This is achieved by using three machine learning classification methods that enable the differentiation between EEG data of patients undergoing the tDCS treatment.

The third objective is to estimate the ECT induced seizure quality rating. Two approaches were considered, one that utilises seizure parameters that were scored by expert clinicians, and another one that relies on the raw EEG signals of patients undergoing ECT sessions. In addition to this, two sets of features are extracted based on the DBN method and EEG bands using raw EEG data. These features are investigated for the identification of ECT induced seizure quality.

1.4 Principal Contributions of the Dissertation

This dissertation presents new classification and identification techniques for predicting the clinical outcomes of patients with major depressive disorder based on EEG oscillations. The principal contributions of this research study are enumerated as follows:

1. Three implementations of DBNs that fuse the information from multiple EEG channels at different levels are presented; in particular, the modified DBN architecture that deals with multi-channel EEG data. The first implementation is a traditional DBN that is fed with a concatenated feature vector of the multiple channels (fusion at the feature level). The second implementation uses a separate DBN for each channel and combines the classification results using another DBN (fusion at the decision level). The third implementation is based on a multi-channel model fusion that constructs a partial DBN with no decision layer for each channel, which is followed by higher level hidden and decision layers that combine the information provided by the partial DBNs. The classification task involves differentiating between baseline EEG, active tDCS and sham tDCS sessions. This will be implemented after identifying the most relevant channels for this classification task. The results obtained using these three implementations are compared to those obtained using support vector machine (SVM), linear discriminant analysis (LDA) and extreme learning machine (ELM) classifiers.

2. The features of resting-state EEG recorded at baseline that differentiate participants who respond to the tDCS treatment from those who do not are investigated. The improvement in mood and cognition scores are used to predict the clinical outcome of the tDCS treatment. Mood is assessed using the Montgomery-Åsberg Depression Rating Scale (MADRS), while cognition is scored using the Symbol Digit Modalities Test (SDMT). A technique is developed using three standard classifiers (SVM, LDA and ELM) to predict if a patient will respond to the tDCS treatment. This experiment consists of two investigations: an initial investigation and an advanced investigation. These are described further as follows:

    - Initial investigation: Three sessions (baseline, active tDCS and sham tDCS) are used to identify which patients are responding and non-responding to tDCS treatment.

- Advanced investigation: Identify and predict the antidepressant response of depression patients based on the baseline EEG data.

  In the two investigations of this research, 62-channel EEG data of 10 patients was used. The performance of each EEG channel is evaluated individually.

3. An ECT induced seizure quality rating is identified. This experiment consists of two studies as follows:

   - Study 1: An ECT induced seizure scoring system based on two classification methods is proposed: fuzzy rule-based system and decision tree. Inputs for the two classification methods were the four seizure parameters of TSLOW, regularity, stereotypy and post-ictal suppression that are recorded by expert clinicians. The two classification methods have a significant advantage as they are both considered transparent models that allow interpretable mapping from the inputs to the output, which makes them suitable for medical applications, unlike other black-box classification methods. The approach presented in this study is expected to help clinicians estimate the scoring of seizure quality and reduce the time taken to evaluate the effectiveness of ECT sessions.
   - Study 2: This study consists of two stages:
     * Two new sets of features are extracted based on a DBN technique and standard EEG bands using the EEG data of ECT induced seizures. The extracted features are then fed to the regression tree classifier to evaluate their discrimination capability. The obtained results are compared with the clinicians scoring.
     * A new estimation system is proposed to predict seizure parameters based on EEG data that are collected through ECT sessions. The study presents an algorithm for rating some of the relevant electroconvulsive therapy based seizure indices (parameters). In this study, the identification systems are constructed from two parts. In the first part, an estimation system is built to estimate the four seizure parameters

(time to onset of slowing (TSLOW), regularity, stereotypy and post-ictal suppression) from 2-channel EEG data. This part is built based on the peaks identification technique. In order to examine each EEG sample, each peak and location are tested. In the second part, a fuzzy rule-based system is designed based on the estimation values of four seizure parameters to predict the seizure quality rating.

This experiment is based on EEG data that was recorded during ECT sessions by Gálvez et al. [37]. The proposed rules in two studies are based on the literature through identifying the rating of seizure quality. This study presents a tool that will help clinicians to estimate the scoring of seizure quality and reduce the time taken to evaluate the effectiveness of ECT sessions.

1.5 Publications

Published papers

- **Alaa M. Al-kaysi**, Ahmed Al-Ani , Tjeerd W. Boonstra, "A Multichannel Deep Belief Network for the Classification of EEG Data", The 22nd. International Conference on Neural Information Processing (ICONIP), 2015.
- **Alaa M. Al-kaysi**, Ahmed Al-Ani , Colleen K. Loo, Michael Breakspear and Tjeerd W. Boonstra, "Predicting Brain Stimulation Treatment Outcomes of Depressed Patients through the Classification of EEG Oscillations", The 38th International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016.
- **Alaa M. Al-kaysi**, Ahmed Al-Ani , Colleen K. Loo, Tamara Y. Powell, Donel M. Martin, Michael Breakspear and Tjeerd W. Boonstra. "Predicting tDCS treatment outcomes of patients with major depressive disorder using automated EEG classification", Journal of Affective Disorders (Elsevier), Volume 208, 15 January 2017, Pages 597-603, Epub 2016 Oct 24.

- **Alaa M. Al-kaysi**, Ahmed Al-Ani, Verònica Gálvez, Colleen K. Loo, Steve Ling and Tjeerd W. Boonstra, "Estimating the quality of electroconvulsive therapy induced seizures using fuzzy inference system and decision tree classifiers", The 40th International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018 (Accepted).

Submitted papers:

- **Alaa M. Al-kaysi**, Ahmed Al-Ani, "Predicting an ECT induced seizure quality rating based on estimation seizure parameters system and construction of fuzzy rules" (Submitted).

1.6 Structure of the Dissertation

The dissertation consists of six chapters. These are organised as follows:

- **Chapter One:** This chapter introduces the thesis and provides a general overview of the rest of the dissertation.
- **Chapter Two:** This chapter presents the theoretical background and literature review on the depression disease and brain neurophysiology. The chapter also provides a background on EEG signals, including EEG rhythms, analysis techniques of EEG signals and the role of EEG as a diagnostic technique for depressive disorder. After that, the chapter focuses on tDCS as a special treatment of major depressive disorder. In addition to this, the chapter contains an overview of the second method for treating depression, electroconvulsive therapy. Finally, the chapter explores the use of special machine learning techniques for the classification, feature extraction and identification of EEG signals. Then, the technique that has the best ability to deal with multi-channel EEG signal, deep belief networks, is highlighted.
- **Chapter Three:** This chapter presents a proposed method for classifying multi-channel EEG signals (multi-channel deep belief networks) and contains the first

contribution of this dissertation. First, a brief description of the training of deep belief networks is provided and the principal operation of the proposed method is introduced. A technical overview of restricted Boltzmann machines is provided and the greedy learning algorithm as a training method is reviewed. Then, the three modification techniques of DBNs which are proposed to deal with multi-channel EEG data are briefly explained.

- **Chapter Four:** This chapter presents the second and third contributions of this dissertation. Two case studies for predicting the clinical outcomes of depression patients, differentiating those who respond to tDCS treatment from those who do not, are investigated. First, an initial investigation that aims at differentiating between the three sessions of tDCS (baseline, active tDCS and sham tDCS) is conducted. Second, a more in-depth investigation to predict the clinical outcomes based on resting-state EEG recorded at baseline is demonstrated.

- **Chapter Five:** This chapter contains the fourth and fifth contributions of this dissertation. The chapter provides a review of seizure quality rating. Then, the chapter presents the first identification of a seizure quality rating score based on other seizure parameters. Two identification techniques are proposed in the chapter: decision tree and fuzzy inference system. In addition to this, the chapter presents feature extraction from raw EEG data based on the DBN method and EEG rhythm. The obtained features are compared with the clinician's EEG recordings and results are reported based on the regression tree method. Finally, the chapter also presents the proposed estimation system and fuzzy rule-based system, which represents the second identification technique for seizure indices based on raw EEG data.

- **Chapter Six:** This chapter provides a summary of the research, outlines the conclusions and the limitations of this research and also suggests several directions for future research.

Chapter 2

Background and Literature Review

2.1 Introduction

Depression is one of the most common mental disorders affecting people around the world. This chapter introduces the disease of depression and identifies the major types of depressive disorder. Then, the neurophysiology of the human brain will be outlined, before describing transcranial direct current stimulation (tDCS), which is one of the most common treatments of major depressive disorders. Another stimulation treatment, electroconvulsive therapy, will also be discussed, followed by a description of the electroencephalogram (EEG), which is the neurophysiological measurement of the electrical activity in the brain. Nowadays, EEG is a great tool to determine the functional state of the brain and it facilitates the diagnosis of disorders. EEG data is captured using multi-channel electrodes placed on the scalp to record activities of billions of neurons in the brain.

This chapter will describe the use of EEG to assess the efficiency of tDCS for treating depression. Additionally, the chapter will also highlight machine learning techniques and rule-based systems methods that can be used in medical diagnosis. Finally, this chapter will review deep learning, particularly deep belief network, for classifying multi-channel EEG data.

2.2 Depression

Depression is a common mental disorder that affects about 350 million people around the world [1, 38]. Many researchers believe that chemical imbalances in the brain can cause depression, which could be related to difficult events that people are exposed to during their lifetime [1]. Depression is characterised by a group of symptoms, such as sadness, insomnia, anxiety, lethargy, and in critical cases, suicidal ideation [39]. Depression has different types based on the risk degree of symptoms, including major depressive disorder (MDD), MDD with melancholia, psychotic depression, persistent depressive disorder and dysthymic depression. Major depressive disorder can be diagnosed by observing the following symptoms: depressed mood, dramatic loss or gain of weight, fatigue, marked reduction in interests and activities from day to day and, in bad cases, thinking of suicide [40].

2.3 Neurophysiology of Human Brain

The human brain is one of the most complex systems in the world and consists of around 100 billion neurons. A neuron as shown in Figure 2.1 [41], represents the basic unit for processing information and it consists about 10000 or more connections per neuron [42].

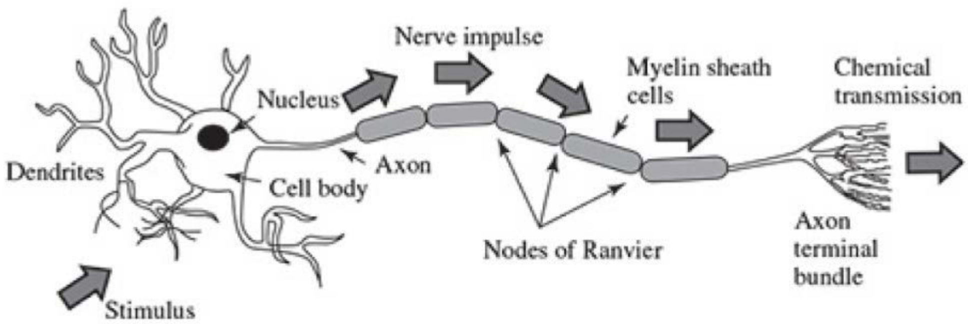


Figure 2.1: Structure of a neuron.

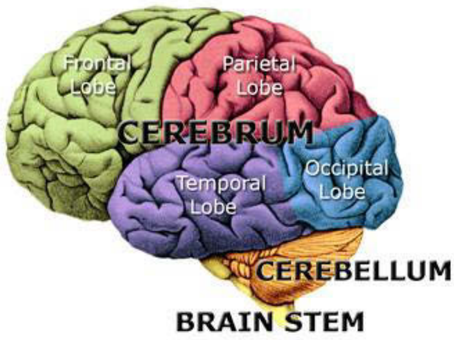


Figure 2.2: Brain regions.

According to the anatomical point of view, the brain can be divided into three parts as shown in Figure 2.2: cerebrum, cerebellum and brain stem [43]. The cerebrum is composed of two parts, the left and right hemispheres, and it represents the largest part of the brain. It is responsible for higher functions such as vision and hearing and, emotions and it holds memory, learning and controls of movements and interpreting touch [44].

The second part of the human brain is the cerebellum, which is located under the cerebrum. The cerebellum coordinates muscle movements and balance. The third part is the brain-stem, which is representing the centre connection between the cerebrum and the cerebellum and it is responsible for many automated functions such as heart rate, breathing, wake and sleep cycles, body temperature, blood pressure, swallowing and sneezing [45].

The cerebral cortex is the most significant part of the brain and it can be divided into four lobes: frontal, parietal, occipital and temporal. The frontal lobe is the largest lobe in the cerebral cortex and is located at the front of the brain. It is responsible for cognitive functions such as attention and short-term memory [44]. The second lobe is the parietal lobe. This is located between the frontal and occipital lobes and is responsible for sensory information. The third lobe is the temporal lobe, which is responsible for basic auditory and language processing. Finally, the occipital lobe is the visual processing centre of the brain and it is responsible for many functions such as vision and hearing. It is positioned at the back of the skull [44].

Analysing neurophysiological brain functioning requires attention from researchers outside of the neuroscience field and it is a significant interdisciplinary task [46]. Electroencephalography (EEG) is a commonly used technique to monitor changes in brain activities that results from brain stimulation.

2.4 Electroencephalogram

Clinical scalp EEG is useful in diagnosing and guiding therapy for different neurological conditions such as acute seizures and depressive disorders [25, 33]. EEG is a non-invasive technique that records the electrical activity of the brain using electrodes placed along the scalp [22]. For the last two decades, several neurophysiologists still manually analyse EEG signals in spite of a huge research in automated processing and classification [33]. In 1929, German neuropsychiatrist Hans Berger was recorded the tracing of human EEG using bipolar recording technique through 1-3 minutes' duration [22]. Berger was the first person to use an electroencephalogram to describe the electrical potential in the human brain and he laid the foundations for several present applications of electroencephalography. He observed that the brain's electrical activity changes based on the functional status of the brain, for example due to lack of oxygen, sleep, anaesthesia and, in particular, epilepsy [22]. Adrian and Matthews diagnosed human brain waves and they identified regular oscillations around 10 to 12 Hz, which they called "alpha rhythms" [22]. Due to the established relationship between alpha waves and a number of brain activities, EEG data is captured from the frontal brain hemisphere [39].

The most common placement of EEG electrodes is based on the 10/20 international system, where electrodes are separated by 10 to 20 percent of the total distance around the scalp [47], as shown in Figure 2.3. The number of electrodes are different based on the application, and the most common are 32, 64, 128 or 256 electrodes. The first letter of the electrode's name (capital) represents the positions of electrodes on the EEG scalp (F (frontal), P (parietal), T (temporal) and O (occipital)) and the letter z represents the midline. There are two types of EEG recordings: monopolar recordings (recording an

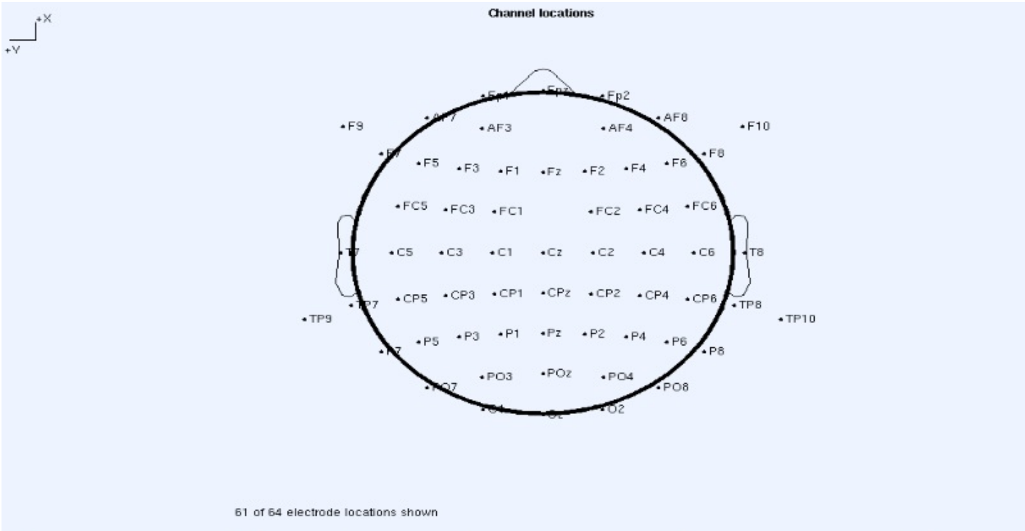


Figure 2.3: The locations of EEG channels

EEG with a reference to a common passive electrode) and bipolar recordings (recording an EEG in a differential way between pairs of contiguous electrodes) [47].

EEG signals are noisy and the traditional technique for reducing the noise of an EEG signal is through signal averaging of event-related potentials (ERPs) [43]. ERPs are significant voltage fluctuations that results from evoked neural activity. The relationship between EEGs and ERPs is quite simple, where every EEG waveform recorded with a stimulus includes an ERP waveform with some artefacts [23].

EEG signals are contaminated by artefacts, which come from external sources [47]. There are two sources of artefacts. The first source is associated with physiologic artefacts such as head movements, blinking, muscle activity and electrocardiograms. The second source on the other hand is associated with non-physiologic artefacts such as kinesiologic artefacts that are caused by electrode movement and mechanical artefacts that caused by body movement [48]. One of the most common physiological artefacts is head movements, which are correlated with a low frequency that result in a fluctuating baseline in the EEG signal [47]. The effect of artefacts can be reduced by filtering the signals; however, due to the overlap in frequencies, filters should be carefully designed in order to filter out important EEG information even if that means not substantially reducing the effect of artefacts [47]. Muscle artefacts are correlated with high frequency activities, which can be eliminated by using standard digital filter. A high-pass filter with a cutoff frequency of 1Hz is usually applied to reduce the impact of low frequency activities such as head movement [47]. It should be mentioned that low frequency could be applied to remove high frequency components, such as muscle artifacts, while a notch filter could be used to remove power line noise.

There are five main EEG frequency bands: delta, theta, alpha, beta and gamma which reflect the different brain states [48]. Brain waves represent the regularly recurring waveforms that are similar in shape and duration [49]. Brain waves and the functions of each EEG band [48] are described below:

1. Delta waves (0.1-3 Hz): appear in dreamless sleep and unconscious states.
2. Theta waves (4-7 Hz): observed in different states such as intuitive, creative, imaginary and drowsy states.
3. Alpha waves (8-12 Hz): the first EEG waves that were discovered by Berger [22]. Alpha waves appear in the relaxation state but not drowsy, tranquil and conscious states. Alpha waves become attenuated in several states such as eyes opening, hearing sounds, anxiety or attenuation.
4. Beta waves(12-30 Hz): observed in the active state and anxious thinking. There are three different bands of beta waves:
   - Low beta waves (12-15 Hz): appear in relaxed yet focused and integrated cases.
   - Midrange beta waves (16-20 Hz): appear in thinking and awareness of self and surroundings.
   - High Beta (21-30 Hz): observed in alertness and agitation states.
5. Gamma waves (30-100 Hz): observed in higher mental activity.

The EEG oscillations of the same frequency may have different functions; for example, delta oscillations are normal and abnormal based on states. Normal through slow wave sleep and clearly signing abnormality during awake state [47].

2.5 Pre-processing EEG Signal

EEG signals can be analysed by three methods: visual inspection by an expert person, an automatic analysis such as signal processing algorithm [50] and independent component analysis. The independent component analysis (ICA) is one of the most popular tools for analysing EEG signals. ICA is an analytical and computational technique that is usually used for detecting signal's sources that cannot be identified by the level of raw data. ICA is used to extract the independent components from the mixed signals. A description of the ICA technique can be found in [51, 52]. Appendix A shows a pattern of ICA analysis EEG data that is used in this research study. ICA has two significant features:

- Enables the removal of components that are highly influenced by artefacts.
- Helps to disentangle otherwise mixed brain signals

The EEG signal consists of a mixture of various brain and non-brain contributions, therefore the size of feature set can be quite large. There are different techniques for selecting features depending on mental disease, such as time domain, frequency domain and time-frequency distribution. In particular, time domain techniques include filtering. Frequency domain techniques depend on the estimate band power using different techniques such as the Welch method, autoregressive modelling and wavelet transform [53].

2.5.1 Spectral Analysis of EEG Signal

Power spectral density (PSD) is used to analyse and compute the characteristics of the EEG signals. Fast Fourier transform (FFT) is an analytical technique that transfers the EEG signal from time domain to frequency domain. FFT is used to compute PSD by estimating the autocorrelation sequence using non-parametric methods such as the Welch method [54, 55]. The Welch method is a power spectrum density estimator that applies the periodogram [56]. In addition, the other method that uses a parametric approach to estimate the PSD of the EEG signal is the autoregressive method (AR). AR computes the PSD by calculating the coefficients of the EEG system [54].

2.5.2 Time-Frequency Distribution of EEG Signal

EEG is a non stationary signal and the best tool to study the spontaneous fluctuation in oscillation is time-frequency distribution, which has the ability to quantify the spectral content of an EEG signal as a function of time. Time-frequency analyses of EEG signals can provide a lot of information about neural synchrony [57]. One of the most technique to compute time-frequency distribution is wavelet transform (WT). WT is a spectral estimation technique and it has a significant role in the recognition and diagnostic field [54, 58]. WT can be evaluated by calculating the relative wavelet function entropy [59]. WT uses two categories: continuous and discrete. The size of windows in the WT technique is variable, which means the time-frequency representation of a signal is more flexible. In addition, spectrogram is another common visual technique in time-frequency distribution.

2.5.3 Coherence Analysis of EEG Signal

The motivation of EEG coherence is to study the linear relationship of functional coupling between two channels on the EEG scalp [60, 61] and the potential interactions in the cortical layer. According to [62], coherence is defined as a measure of the variability of time differences between two time series. In particular, the coherence of EEG data represents the specific frequency that indicates the diverse neuroanatomical and neurophysiological factors in two electrodes of an EEG signal [63]. The significant index in EEG coherence is the phase angle and it represents the relationship between the time and frequency domains. Coherence is equal to 1 if the phase angle is stable and constant over time and coherence is equal to 0 if there is a time difference between two time series from moment-to-moment [62].

2.6 Diagnosing Depressive Disorder Based on EEG Signals

Depression is one of the mental illnesses that can be identified by analysing EEG signals [6]. For example, Li et al. analysed the EEG data of two groups; patients with depression and healthy controls using wavelet entropy (WE) and subband segmentation. They found that brain activity was more regular in depressed participants during mental arithmetic than the healthy control subjects [64].

Several neuro-imaging studies of major depressive disorder (MDD) have examined morphometric differences between patients and healthy controls [6]. For example, patients with depression have been found to have less activity in the left frontal cortex than the right frontal cortex [39, 65]. Other researchers used a multiparameteric classification technique and neuroanatomical characterisation to distinguish medication-naïve adults with first episode MDD from healthy controls [6]. They used a multivariate support vector machine to identify two groups of patients. They showed that MDD patients have a complex and multidimensional pattern which represent the spatial maps of the brain regions that discriminate MDD patients and healthy people, as shown in Figure 2.4 [6]. They attempted to discriminate the volumetric and geometric parameters between the patients that have MDD and the healthy controls [6].

Other researchers attempted to identify MDD by analysing abnormal oscillation of the theta and alpha EEG frequency bands [7, 66, 67]. In relaxed wakefulness, the alpha power is large and, when the brain conducts an active function such as intense mental arithmetic, the alpha power is decreased [68]. Hence, the activity in this brain region is taken to be inversely proportional to alpha band [39, 69]. Niemiec et al. called this technique the alpha inactivity mechanism (AIM) and many researchers in the brain field use AIM to deduce the difference in brain activity between left and right frontal cortex [39]. In contrast, Jatupaiboon et al. demonstrated that when they used different frequency bands of each channel pair to classify two emotions (positive and negative), high frequency bands gave better result than low frequency bands and frontal pairs of channels gave a higher accuracy than the other brain areas [70].

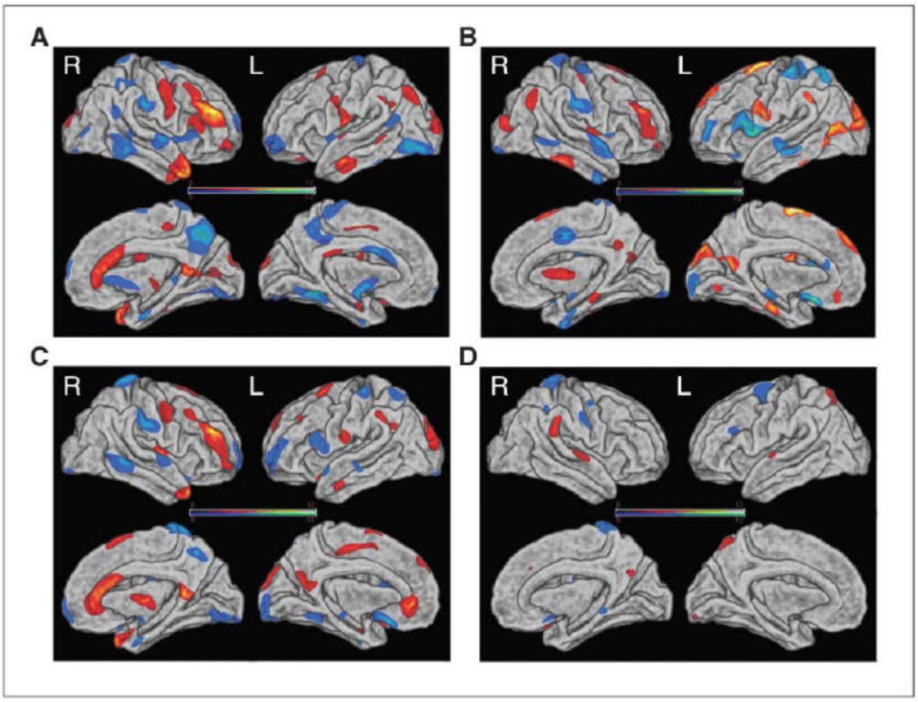


Figure 2.4: Discrimination maps of brain between depressive disorder patients and healthy controls.

Tabacaru analysed frontal lobes EEG signals to detect the effect of depression on alpha rhythms [71]. They calculated the left-dominant alpha asymmetry (LDAA) for each EEG recording for seven subjects. They showed that changes in the LDAA depend on the subject's emotion and that EEG activity in frontal lobes is responsible for these influential emotions. They demonstrated that the index of LDAA can be indcated the depression [71].

In addition, Puthankattil et al. developed two methods for analysing EEG signals, using a time-frequency domain and nonlinear method to distinguish between patient with depression and healthy control [59]. They used the wavelet entropy (WE) method for time-frequency analysis and approximate entropy for nonlinear analyses. They indicated that the WE indicator can be used to distinguish between healthy and patients with depression based on the different brain states. Also, they used the approximate entropy to examine the complexity of the brain activity, and they noted that the brain activity in normal people are more complex than the brain activity in depression patients [59].

Patients that have a depressive disease may require treatment at highly specialised mental health services [6]. There are two types of treatments for depressive disorders: medication techniques and non-pharmalogical techniques. Electroconvulsive therapy (ECT), transcranial magnetic stimulation (TMS) and transcranial direct current stimulation are non-pharmalogical interventions techniques that aim to improve the treatment of neuropsychiatric disorder [72].

2.7 Transcranial Direct Current Stimulation Treatment

Transcranial direct current stimulation (tDCS)is a neuro-modulator technique that can be used for treating depression. tDCS is a safe and non-invasive method, which does not require anaesthesia for treatment [73, 74]. A direct current which is generated by an electrical stimulator is delivered bilaterally by a pair of rectangular saline-soaked surfaces with sponge electrodes. The anode placed over the left dorsal lateral prefrontal cortex (DLPFC) and cathode located over the lateral aspect of the contralateral orbit.

In this research study, the transcranial direct current stimulation treatment was administered by an Eldith DC-stimulator (NeuroConn GmbH, Germany) with anode and cathode electrodes which are placed based on the international 10/20 electroencephalogram system [3]. The experiments of the diagnosis of the acute after effect of tDCS were conducted during a virtual working memory task on cortical activity of an EEG through a depressive episode [7]. Active stimulation was given at 2 mA for 20 min, with a gradual ramp up/down of the current over 30 seconds. A sham stimulation was recorded on 1 mA current for 30 seconds with ramp up/down over 10 seconds, giving an initial sensation of tDCS while mining stimulatory effects. The duration of the two sessions (active and sham) tDCS were performed seven to eight days apart [7].

Fregni et al. showed that the tDCS induced antidepressant effect appeared after five sessions for a limited number of patients (10) in a sham-controlled design [75, 76].

Boggio et al. used the same experiment with 40 patients and they also reported positive outcomes [76, 77]. In addition to these, many researchers showed that tDCS has good efficacy after six weeks. Rigonatti et al.[78] and other researchers [76] reported a fast and sustained response of tDCS at week two, while, Loo et al. showed that tDCS had no efficacy after five days but had better outcomes after 10 days of stimulation double-blinded with 40 patients during the open-loop phase [8, 76].

Another study evaluated the efficacy of tDCS for major depressive disorder based on the meta-analysis technique. For example, Berlim et al. used the clinical outcomes such as the response of patients and their remission rates and they showed that sham tDCS remains unclear using meta-analysis technique [79]. In contrast, Powell et al. investigated the effect of tDCS using the EEG signals of patients with MDD and they evaluated the activities of EEG after each session of tDCS using event related potentials and visual working memory tasks. They noted that behavioural effects of tDCS was not demonstrated on the virtual working memory [7].

The propagation during brain network may increase the cognitive and behavioural effects of tDCS [7, 80]. Attempts to explore the effects of tDCS have been conducted on mood and depressive symptoms in human since 1960 [81]. Between 1960 and 1970, a limited number of studies examined the effect of tDCS on depression. However, recent studies are used protocols in their experiments that fundamentally differ from early studies in terms of current strength, electrode position and the duration of stimulation, the lack of a control group, concomitant medications and the criteria for diagnosing depression in subjects [77, 82–84].

Accordingly, tDCS has mainly been investigated in clinical studies. A standard tDCS treatment involves numerous sessions that takes few weeks, however, not all participants respond to the treatment. Based on literature, any attempt to utilise machine learning in predicting the treatment outcome of tDCS is not proposed. Hence, this thesis attempts to address this issue, and accordingly open the door for future studies in this direction

2.8 Electroconvulsive Therapy

Electroconvulsive therapy (ECT) is the most common treatment for major depressive disorder [13]. During ECT, a brief electric current is passed through the brain to trigger a brief seizure. A seizure is a neurological state where there are irregular electrical signals in nerve cells of the cerebral cortex in the brain. These electrical signals can be observed from the EEG waveforms of patients. There are many symptoms that are associated with seizure state such as hand trembling, loss of consciousness and body spasms [85, 86]. While the seizure state can be identified from processing EEG signals.

Previous studies show that the quality of an ECT seizure is affected by the age of the patient and the treatment period [37, 87–89]. Most ECT studies focus on the effect and type of anaesthetic on seizure quality [90, 91], with some studies investigating the anaesthetic technique aspects [92]. Gálvez et al. evaluated the effect of the anaesthetic-ECT time interval on seizure quality, which is the time between the starting of anaesthetic injection and the ECT stimulus [37]. They evaluated seizure quality based on four indices: slow wave onset, amplitude, regularity, stereotypy and post-ictal suppression. They also rated the duration of the seizure using a single blinded trained rater and demonstrated that the seizure quality increases with longer anaesthetic-ECT time intervals [37]. They investigated the effect of the anaesthetic-ECT time interval for two ECT seizure ratings: the duration of seizure and seizure quality. They noted that longer anaesthetic-ECT time intervals caused increases in all seizure indices and durations, except for the time to onset of slowing (TSLOW), which was not affected. They also demonstrated that the seizure quality increases with longer anaesthetic-ECT time intervals [37].

Seizure quality is a useful tool for deciding the appropriate therapeutic for the patients [16, 17, 37]. However, there are hardly any studies that investigate automatic estimation of ECT induced seizure quality rating. This research study attempted to bridge this gap in the literature.

2.9 Seizure Prediction

In general, seizures are detected from the sudden surge of uncontrolled electrical activities in the brain [5], minor physical signs, or a series of symptoms. A seizure may start in one cerebral hemisphere or both cerebral hemispheres simultaneously [19] and it can be divided into three periods: the beginning, the middle and the termination of the episode. The automatic detection and prediction of seizures is expected to have numerous benefits such as enhancing the quality of epileptic patients' lives. Hence, it is important to enhance the performance of seizure detection methods [93].

The first seizure prediction method was proposed in 1970 by Viglione and colleagues [94]. Wang et al. detected epileptic seizure based on the classification of EEG signals. They compared healthy people with their eyes open and epilepsy patients during an epileptic seizure [95]. The significant part of their study was selecting the best features for characterising non-stationary EEG signals. They proposed a new intelligent EEG-based diagnosis system. Their proposed model consists of three stages. In the first stage; they represented the EEG signal using wavelet packet entropy and they extracted features from the EEG data using the same method. In the second stage, they trained data using cross validation and a k-Nearest Neighbour (k-NN) classifier. The last stage was the testing stage. Their classification results demonstrated that the efficiency of their proposed method for early detection any changing in electroencephalographic [95].

2.10 Machine Learning Classification Techniques

Over the last few decades, engineering principles such as machine learning techniques and rule-based system techniques have come to play an important role in medicine and the provision of healthcare. In particular, support vector machine (SVM), extreme learning machine (ELM), and linear discriminant analysis (LDA), and decision tree have become widely used in diagnosis and treating different types of illnesses [96–98].

Various tasks have been investigated; such as modifying the adjustable parameters (weights) of a model to minimize the error rate between the output and desired pattern scores [99]. The selection of a classification technique that has the ability to deal with multi-channel EEG data and the nature of prediction tasks is a challenging problem. Selection of appropriate classifier can be the help to improve the classification accuracy. Three standard classification methods that differ in their theories and training mechanisms are used in this research. These three classification methods are support vector machine (SVM), extreme learning machine (ELM), and linear discriminant analysis (LDA). Below is a description of these three classifiers.

2.10.1 Support Vector Machine

SVM is a supervised learning model with associated learning algorithms [100, 101] that was developed to solve classification and regression problems [102, 103]. SVM is a non-linear classification method that has successfully been applied in different fields, including medical diagnostics and recognition tasks. There are several advantages of this method: less effective by the overfitting problem and robustness to noise.

Accordingly, SVM achieved very promising classification results when applied to various classification tasks. However, the biggest limitation of the SVM techniques is their high computational cost, especially when dealing with large datasets. In addition to this, the lack of transparency of the classification results is a common disadvantage of SVM. On the other hand, the classification accuracy can be interpretative in certain cases using graphical visualization.

2.10.2 Extreme Learning Machine

ELM is an effective and fast learning method, which is proposed for training single hidden layer feed-forward neural networks [30, 104]. When comparing ELM with other traditional neural networks and SVM classifiers, ELM is faster. ELM has also shown to have good generalization capability. The random initialisation of the single hidden layer nodes (internal weights and biases) is the principal concept of ELM [105]. Despite this property, ELM keeps the universal approximation capability of single hidden layer feed-forward neural networks [105].

2.10.3 Linear Discriminant Analysis

LDA is the common linear method that is used in different classification tasks, especially with EEG and BCI-EEG. LDA can also be used for feature extraction [106]. LDA is a statistical machine learning method for classifying two or more classes to find the optimal linear combination of features based on the discriminant hyperplanes techniques [1]. LDA is a simple and a highly efficient method and it separates the data represented based on hyper-planes [105]. Linearity is the major drawback in LDA technique, which could result in a poor classification accuracy with complex nonlinear EEG.

2.11 Rule-Based Systems Methods

The rule-based systems have witnessed recently great attention from the researchers in classification tasks, particularly for diagnosing diseases. There are several techniques of rule-based systems such as fuzzy inference system, decision tree and genetic algorithms. Due to the huge amount and different sources of medical data, the process of making decisions is very complex and difficult. The literature shows that decision trees and fuzzy logic maintain interpretability with good classification accuracy [107–109]. Several researchers believe that these methods have a higher predictive power than signal analysis techniques [110–114]. The following two subsections describe the two rule-based system methods that are proposed to achieve the aim of this research: decision tree and fuzzy rule-based system.

2.11.1 Decision Tree

The decision tree is a hierarchical learning model and represents one of the most important classification methods. It is a graphical model for analysing decisions for classification and regression tasks. The concept of this method follows the tree metaphor, which consists of roots and leaves. The roots of tree represent the initial splitting of data attributes and the decisions are made based on the path from root to leaf [98]. A decision tree is fast and intuitive decision rules [33] and represents a visualised method that is easy to interpret [107]. There are three stages of building decision tree: initial constrained tress, splitting (branches and leaves), and pruning tree (reducing the number of leaves) using one of the pruning methods.

In general, the initial tree, which is generated from the tree building part is large and complex, which could increase the chance of overfitting, especially if training data is noisy or contains limited samples. Pruning is used to improve the capability of the decision tree [98]. Classification tree are divided into two groups: training and testing. Training data is used to construct the tree and testing data is used to test the misclassification rate of the classification method. The classification tree consists of two main parts: node and leaf [115], the labels of nodes represent the input features and the labels of the leaves represent the probability distribution over the classes.

2.11.2 Fuzzy Rule-Based Systems

A fuzzy set has the ability to handle the uncertainty that appears in medical applications when making decisions [110]. Therefore, fuzzy sets are very useful in various applications such as pattern recognition, diagnostics and decision making in the medical field and modern information technology [110, 116]. The structure of fuzzy inference systems consists of four main components: knowledge base, fuzzification interface, inference engine and defuzzification interface [117]. In general, fuzzy rules are built based on human expert knowledge. Due to the simplicity and speeding up of the formulation of fuzzy rules and the flexible design of medical diagnostic algorithms, fuzzy inference systems have become acceptable tools to describe medical knowledge [118].

There are two types of fuzzy system: Mamdani and Takagi–Sugeno (T-S) or fuzzy dynamic models. The T-S fuzzy model is used with the complex systems, for example, multi-input and multi-output systems (MIMO) [117].

There are several successful applications of fuzzy models such as data analysis [119], image processing [120] and data mining [121]. Gadaras and Mikhailov used numerical labelling of medical data to learn fuzzy classifiers in different manners [122]. The method represents a simple process that does not need any initial knowledge. They were dependent on diagnosis from medical data and they sought to find the matching between symptoms and diseases. Their proposal was focused on two main criteria: classification accuracy and interpretability of the model. Their fuzzy classification system aimed to improve the efficiency of medical diagnosis systems [122]. Other researchers have used observational data to build fuzzy systems [123]. For example, Levashenko et al. proposed a decision making system based on fuzzy logic for diagnosing breast cancer disease based on cumulative information estimates. They used three fuzzy decision trees: non-ordered, ordered and stable. They created a fuzzy decision tree based on cumulative information estimates [107].

In addition to the above mentioned methods, another study [124] designed a fuzzy rule-based classifier to predict coronary artery disease noninvasively based on a myocardial perfusion scan test and medical variables. They selected membership functions based on clinical variables and they obtained a high quality of precision based on these membership functions. They built weighted rules based on expert decision and supervisor classifiers with a fuzzy inference system to get results with a high accuracy. Due to the interpretable property of the fuzzy rule-based system, they noted that the classification rate could be improved [124].

Due to the capability of fuzzy sets and fuzzy systems to accommodate non-linearity and impressive information [125], fuzzy inference systems are suitable for treating the uncertainty, complexity and noise-contaminated systems in the real world. The fuzziness has the ability to describe the vagueness in the description of model or system.

2.12 Deep Learning Techniques

The motivation of deep learning is to learn abstract representations of the input data in an unsupervised manner then use the same training set as input for supervised learning. It is therefore applicable for both classification and regression tasks[126]. The original idea of deep learning came from the study of artificial neural networks, and is based on distributed representations. Deep learning has two significant advantages: first, treating the partially overfitting problem; and second, the non-random initialisation parameters of the network are based on the pre-training data [127].

Deep learning techniques started to play a significant role in several areas such as handwriting classification [128], single image super resolution [129], and brain computer interfaces (BCI) [130]. During 1950 and 1960, the researchers were focused on multilayer neural networks. However, in 1980, the researchers suggested that the perceptron was a special type of neural network. It is a supervised algorithm that consists of input and output layers [131, 132]. Another study proposed a supervised algorithm to train neural networks with hidden layers called backpropagation [131, 133]. In general, multilayer neural networks are special examples of models with deep architectures.

Many researchers [32, 134] demonstrated that deep learning has a promising capability to improve the accuracy of multi-layer perceptrons (MLPs) [135]. Deng and Hu proposed a deep architecture called deep convex networks (DCNs). DCNs consist of a set of layered modules and the authors proposed to use them to fix the problem that appears in learning scalability. The learning technique for DCNs is batch-mode based and this method can be used in image classification [136].

Due to the ability of deep learning architectures to handle high dimensional unlabelled data [137, 138], it is powerful tool to analyse big data such as speech recognition, computer vision and medical diagnosing. However, studies rarely examine the classification of EEG data. In this study, a special type of deep learning is used to classify EEG data.

There are many types of deep learning, including deep neural networks such as the deep Boltzmann machine, convolutional neural network (CNN), deep belief network (DBN) and stacked de-noising autoencoders. The following sections will discuss two special types of deep learning: DBN and CNN. It will also present the special and common techniques of deep learning.

2.12.1 Deep Belief Networks

The deep belief network (DBN) was first proposed by Hinton et al. [139]. It is a special type of deep neural network. DBNs are probabilistic generative models which are composed of many hidden layers [140]. DBNs represent a hierarchical model that is implemented using a greedy layer-by-layer method that can be repeated several times [140]. The building block of a DBN is the restricted Boltzmann machine, which is a bipartite undirected graphical model.

The training process of DBN consists of two stages. The first stage is the pre-training stage, which contains unsupervised learning based on down-up training to extract features. Training the network weights in an unsupervised manner is a more effective and efficient method than using random initialisation to select the best neural network weights [141]. The initial values of network weights are selected by an unsupervised greedy layer-by-layer pre-training algorithm. Backpropagation or other discriminating algorithms can be applied to fine tune weights. The second stage is the fine tuning stage, where supervised learning is used to adjust network parameters based on an up-down training algorithm. Hinton et al. proposed a deep belief network as an efficient unsupervised learning algorithm to overcome the complexity of training deep generative models [9, 142].

There are several applications of DBNs, such as the classification and detection of electroencephalography waveforms [33], documenting classification, musical emotion recognition [143, 144], phone recognition [145], modelling images and some

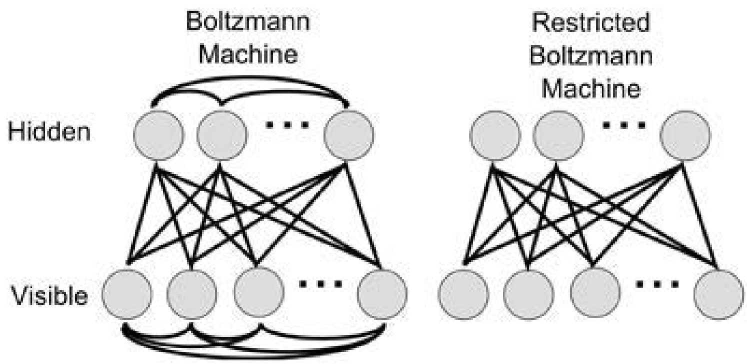


Figure 2.5: Restricted Boltzmann Machines.

applications for image transformation, extracting features from image data and extracting relevant features from audio [143, 146, 147]

Recently, Liao et al. proposed a new image retrieval method based on DBNs and a Softmax classifier for binary images. Based on simulating the architecture of the human visual system, DBN-Softmax is provide a more effective representation and extraction measurement in comparison with existing image retrieval models [148].

2.12.2 General Boltzmann Machine

Boltzmann machines (BMs) were introduced in the 1980 [149, 150] as graphical models [149, 151], along with undirected graphical models called Markov random fields [149]. Boltzmann machines have two types of units, visible neurons and hidden neurons [152]. Boltzman machine can be used to learn an unknown probability distribution based on the distribution of labels. The purpose of the Boltzmann machine is to estimate the prediction of two connected units [153]. The principle of the log likelihood gradient is used in the learning process of Boltzmann machines with any pattern of connection units [154]. However, the difficult and time-consuming learning process [149] can impose restrictions on the network topology of the BM. Restricted Boltzmann machine (RBM) is a special type of machine learning which is proposed based on these limitation of BM. Figure 2.5 shows the difference between a BM and an RBM [154].

2.12.3 Restricted Boltzmann Machine

A restricted Boltzmann machine (RBM) is an undirected, bipartite graphical model that consists of two layers. The first layer is a visible layer that contains the visible units and the second layer is a hidden layer that contains latent variables [155]. There are no connections between units of the same layers but there is a connection between the visible and hidden layers via undirected weights. RBM has many applications such as time series modelling [156], information and image retrieval [157] and collaborative filtering [140].

The Markov chain plays a vital role for training RBM because it has the ability to draw samples from complicated probability distributions such as a Gibbs distribution [152]. In 2008, researchers proposed a discriminative restricted Boltzmann machine to fix the selection of parameters [158]. In 2012, other researchers proposed conditional restricted Boltzmann machines to improve the performance of multi-label classification tasks [159].

The traditional learning algorithm for training RBMs is the contrastive divergence (CD) that was initially proposed by Hinton [142, 160]. This method requires expert skills to select the best values of numerical meta-parameters [160].

RBM plays a significant role in several applications such as dimensionality reduction, collaborative filtering, feature learning and classification tasks [127].

2.12.4 Convolutional Neural Networks

Due to the rapid growth in the amount of the annotated data and the strength in the graphic processor units, the convolutional neural network was developed to assist with different tasks [161]. The idea of a convolutional layer is weight sharing, so the property of weight sharing is to reduce the number of parameters and provide faster training [130]. A convolutional neural network is a technique designed to process data in the form of multiple arrays.

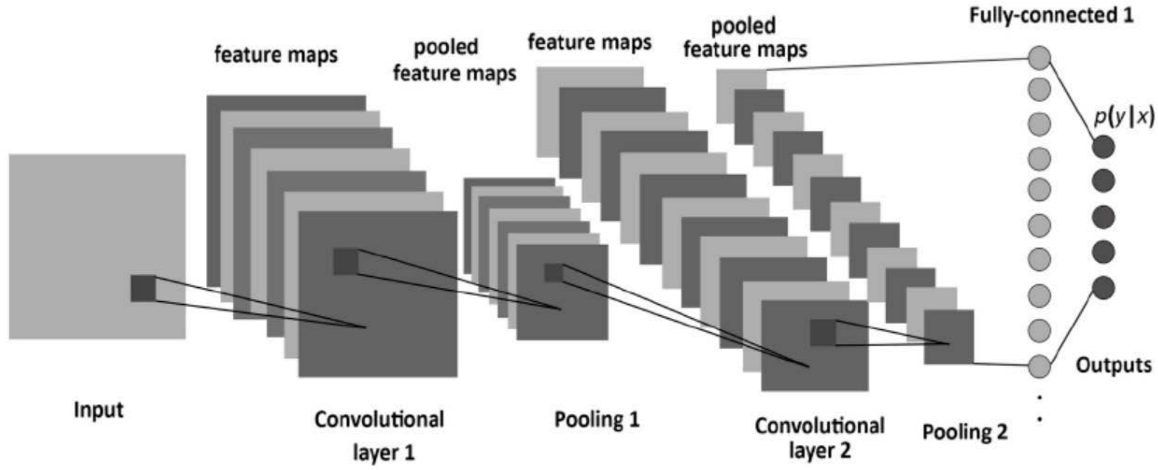


Figure 2.6: The structure of convolutional neural network

CNN is composed of two main layers: convolutional layer and pooling layer, as illustrated in Figure 2.6 [162]. The aim of the convolutional layer is to learn the feature representations of the input data and it is composed of several convolutional kernels that compute different features maps [161]. In particular, the convolutional layer is used to examine the previous layer to detect the features' local conjunctions; however, the pooling layer is used to combine similar features into one feature [99]. The pooling layer is a significant layer in CNN as it is used to reduce the multiple connections between convolutional layers.

There are different aspects that can improve the CNNs, such as convolutional layer, pooling layer, shared weights, local connections, activation function, loss function, regularisation and optimisation. One of the common solution for optimising CNN is stochastic gradient descent [161].

There is another type of deep learning called; the convolutional restricted Boltzmann machine (CRBM), which is consist of two layers, input layers and hidden layers. The visible units are binary or real valued, but the hidden units are binary valued [163]. A CRBM is similar to an RBM except for the weight sharing properties between hidden and visible layers [164]. In a CRBM, weights are shared between the hidden and visible layers through all locations in an image [165]. CRBM is used to learn high level representation by including a probabilistic max-pooling. Hidden layers consist of two types of layers the detection layer and the pooling layer, and the unit in the pooling layer represents the maximum probability of the units [166].

CRBM is designed to address the scaling model issue. In particular, CRBM is used to compute the spatial relationship between image patches, particularly the large images that RBM cannot be used for. The weights between the visible and hidden layers are shared in a CRBM. CRBM are building blocks of convolutional deep belief network (CDBN). A CDBN can be used with multi-channel data, such as EEG. The architecture of a CDBN is composed of several maximum pooling of CRBMs stacked on top of one another [164]. CDBN uses unsupervised learning to learn features from unlabelled data [167].

2.13 Classification of EEG Signals Using Machine Learning Techniques

The typical classification systems contains four main components: pre-processing, feature extraction, dimensionality reduction and classification. The first three components were described in the previous sections. This section discusses the classification of EEG signals. In general, the process of classification techniques involves dividing the data into two sets: training and testing.

Hosseinifard et al. analysed the nonlinearity of the EEG signals to distinguish between depression patients and healthy people [1]. They used three classification methods; K-nearest neighbours(KNN), linear discriminant analysis (LDA) and logistic regression, to discriminate between these two groups and they selected features using a genetic algorithm. They analysed two types of features: the power of EEG frequency bands (delta, theta, alpha and beta) and the max-Lyapunov exponent. Alpha power achieved better accuracy than other features and they demonstrated that there is a significant difference between electrodes in the left hemisphere of depression and healthy subjects in terms of the alpha band. Logistic regression had better classification performance than the other two classifiers (LDA and KNN). It is claimed that the method could be a complementary tool to help psychiatrists to diagnose depression disorder [1].

Another study considered the classification of two EEG datasets. The first set was collected from healthy volunteers with eyes open, while the second set was collected from epilepsy patients during epileptic seizures. They used a combination of model-based and least squares SVM [168]. They extracted three sets of features based on three spectral analysis methods: Burg autoregressive, moving average and least squares modified Yule–Walker autoregressive moving average. They showed that features obtained using the Burg autoregressive coefficients produced the best classification accuracy [168].

In contrast, Murugavel and Ramakrishnan proposed a novel hierarchical multi-class SVM with ELM as kernel to identify epileptic seizure based on classification of EEG signals [169]. They used wavelet transform to extract features and they fused binary SVM into multi-class SVM in a hierarchical way. They demonstrated that when they compared it with ANN, the proposed model was efficient based on classification accuracy with a lower execution time [169].

In addition, other researchers used SVM to predict a subject's response, who can perform task or not [170]. They examined different kernels to select the best kernel function. In order to select the most relevant features, they used four different methods: two filters and two wrappers. They demonstrated that the information on movement planning were low in the lower frequency bands (7-13) Hz [170]. On the other hand, Bhardwaj et al., classified EEG signals to detect real human emotions and ignore the other external emotions like facial expressions based on SVM and LDA classifiers [171]. They classified emotions into seven classes; happy, sad, anger, disgust, neutral, fear and surprised. They demonstrated that the happy and sad emotions were the best real emotions based on the class-wise accuracy analysis [171].

In general, shallow learning algorithms such as SVM and LDA require short training time and could produce good classification performance. However, when the classification task becomes very complex [100], the performance of these classifiers may not be that good.

2.14 Classification using Deep Learning Techniques

Recently, deep learning and, in particular, deep belief network (DBN), has witnessed increased attention from researchers as a new classification platform that has the ability to deal with big data. Several papers have successfully applied DBNs to a number of classification problems, such as image classification, speech recognition and natural language processing. However, few studies have used DBNs as semi-supervised paradigm for the classification and detection of EEG waveforms. Wulsin et al. used a semi-supervised DBN for the classification and anomaly detection of EEG signals. They used raw EEG data to evaluate the performance of four classifiers (decision tree, SVM, DBN and KNN) based on hand chosen features [33]. They tested each feature individually by evaluating the classification performance of each feature and then selected the feature set that produced a good performance. They also evaluated the execution times of these classifiers. Finally, they compared the performance of DBN anomaly detection based on two cases: using raw EEG data as input and using extracted features as input. They demonstrated that the DBN performed better than other classifiers in high dimensional data. They also noted that raw data had a better performance than extracted features for signal anomaly detection [33].

Zheng et al. introduced a deep belief network to search the neural signature associated with positive and negative emotional categories from EEG data. They trained the DBN using different entropy features that were extracted from multi-channel EEG data. Finally, they showed that the performance of DBN was better in higher individual frequency bands [30]. Furthermore, Shim and Lee improved the DBN's algorithm to recognise the patterns of multi-channel electromyography (EMG). They demonstrated that DBN outperforms shallow learning algorithms (SVM, LDA) and backpropagation neural network [172].

Several researchers use DBNs as a feature extraction technique in different applications. A feature extraction scheme is a procedure for choosing features which represent the most significant information for classification and regression tasks [54] and it represents a particular form of dimensionality reduction. Ngiam et al. proposed a deep network to learn unsupervised features for text, image or audio. They showed that the performance of a bimodal DBN based autoencoder was worse in cross modality and shared representation tasks [35]. Srivastava and Salakhutdinov proposed a DBN of a multimodal system for learning a joint representation of data. They used images and text bi-model data to demonstrate that a multimodal DBN (MDBN) can learn good generative attributes [173].

Moreover, Cheng et al. proposed multi-model DBN (MDBN) as a new deep learning model to fit large images (e.g. 1024×768). The authors used a scale-invariant feature transform (SIFT) descriptor to process the images, which are then sent to MDBN to extract features. They also adapted the Markov sub-layer to reflect the neighbouring relationship between the inputs [174]. Wang and Shang applied DBNs to analyse features that were learned from raw physiological signals. The raw data was collected from two EEG and two EOG channels and they focused on using a single DBN to train all subjects. They depended on classification accuracy to evaluate the DBN's performance [175].

In addition to these, Leng et al. used DBN to extract features of a 3D model. As the data was unlabelled, they applied a semi-supervised method to recognize 3D objects. They therefore demonstrated the powerful ability of their proposed method to represent the distribution of input data [176]. Kuremoto et al. proposed a three-layer DBN algorithm based on the original idea of DBN [142] to predict time series forecasting [177]. When they compared their algorithm with the MLP, they demonstrated that DBNs with RBMs have better forecasting accuracy than the MLP. They modified RBMs by using a classical particle swarm optimisation (PSO) algorithm [177].

Other researchers used another RBM to extract features, for example, Hu et al. modified a restricted Boltzmann in an unsupervised learning and used it as a feature extraction [178]. They proposed a class information or class preserving RBM to evaluate the usefulness of the features for classification task. They made the class information more clear by adding two constraints to the RBM: intra-class affinity and inter-class repulsion. One of these constraints is used to decrease the distance between features in the same class and the other one to increase the distance between the features of different classes [178].

It is worth to observe the advantages of deep learning techniques as classification and feature extraction tools. However, limited studies that are classified and examined EEG signals, and there is no study deal with multi-channels EEG signals. Accordingly, this study is modified deep belief network as a special technique of deep learning to deal with EEG signals (multi-channels inputs).

2.15 Summary

In this chapter, a background of the neurophysiology of the human brain was presented. The chemical imbalances in the brain that can cause depression was discussed. The chapter then briefly reviewed the literature on electreoncephalogram signals, then highlighted the special methods to extract features from EEG data. This chapter also reviewed treatment of depressive disorder. The procedure of the tDCS tool was explained and the literature concerning the its use in depression was briefly reviewed. The ECT, which is the second treatment of depressive disorder was also discussed and the need for estimating the ECT induced seizure quality was highlighted.

The second part of this chapter outlines classification and feature extraction methods. It details the standard machine learning and rule-based systems techniques. In addition, standard types of machine learning and rule-based systems classification methods were presented, these include LDA, SVM, ELM, decision tree and fuzzy rule-based system. This was followed by description of deep learning methods that include deep belief networks and convolution neural networks. Finally, this chapter review briefly the feature extraction and classification techniques with their applications in medical and non-medical fields.

Chapter 3

Multichannel Deep Belief Networks for the Classification of EEG Data

3.1 Introduction

Deep belief networks (DBNs) have emerged recently as new learning algorithms composed of multilayer neural networks that have the ability of training high dimensional data. A DBN is a probabilistic and generative model that is constructed of multiple layers of hidden units. The undirected layers in the DBNs are called restricted Boltzmann machines (RBMs), which are trained layer by layer. The principal operation of a DBN depends on using a hierarchical structure of multiple layers of RBM. RBMs have been used in machine learning as a generative model for different types of data [160].

First, this chapter will describe the basic concepts of the DBN architecture, including the procedure of training DBNs. It will then present the applications of DBN classification. This will be followed by a description of the proposed DBNs for the classification of EEG data, including individual and multichannel DBNs.

3.2 Deep Belief Networks

Deep belief networks (DBNs) are composed of several levels of neural networks and each level of these networks represents nonlinearity in training data. The building blocks of these levels are called restricted Boltzmann machines (RBMs) [143]. RBMs are energy-based models formed by two layers of probabilistic binary units: visible and hidden layers [131]. Hinton et al. introduced a new greedy layer-wise training algorithm for DBNs that learns layer by layer [142]. An unsupervised greedy layer-by-layer pre-training algorithm is used to initialise the values of network weights. According to [141], training the network weights in an unsupervised manner is more effective and efficient than using random initialisation for selecting the optimal weights. Backpropagation or other discriminating algorithms can be applied for fine tuning weights.

3.2.1 Training Deep Belief Networks

The process of training DBNs involves the individual training of each RBM one after another and then stacking them on top of each other. Each two consecutive layers in DBNs are treated greedily as an RBM [179]. The greedy layer-wise algorithm for training DBNs is described briefly by Hinton et al. and Bengio et al. [134, 180].

An RBM consists of a weight matrix $W_{i,j}$, where i represents a visible node and j a hidden node. The visible units in RBMs are Gaussian or Bernoulli stochastic, while the hidden units are usually Bernoulli [142]. The energy function of a joint configuration (**v**,**h**) of the visible and hidden units is described by [134, 160, 181]:

$$E(\mathbf{v},\mathbf{h}) = \sum_{i \in visible} a_i v_i - \sum_{j \in hidden} b_j h_j - \sum_{i,j} v_i h_j W_{ij} \tag{3.1}$$

where v_i, h_j are the binary states of visible and hidden units i, j respectively, W_{ij} is the weight between visible and hidden units and a_i, b_j are their biases. The model parameters are composed of $[W, \mathbf{v}, \mathbf{h}]$, $a = [a_1, a_2, ..., a_V]^T$, $b = [b_1, b_2, ..., b_H]^T$. The probability for every possible visible and hidden pair is assigned by the following function:

$$p(\mathbf{v},\mathbf{h}) = \frac{1}{Z} e^{E(\mathbf{v},\mathbf{h})} \tag{3.2}$$

where Z is the partition function that represents the summing of all possible pairs of visible and hidden vectors:

$$Z = \sum_{\mathbf{v},\mathbf{h}} e^{E(\mathbf{v},\mathbf{h})} \tag{3.3}$$

The derivative of the log probability of the training vector with respect to weights is defined by following equation:

$$\frac{\partial \log p(\mathbf{v})}{\partial W_{ij}} = \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model} \tag{3.4}$$

The learning rule for performing stochastic steepest ascent in the log probability of the RBM training data results in the following change of weight:

$$\triangle W_{ij} = \varepsilon(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}) \tag{3.5}$$

where ε is a learning rate. In each RBM, there are direct connections between units of the two layers, but there are no connections between units of the same layer. Not having connections between visible-visible and hidden-hidden units leads to an unbiased sample of the state of hidden and visible units. Updating the visible and hidden units can be implemented using Gibbs sampling [134, 160].

The traditional learning algorithm for training RBMs is contrastive divergence (CD), which was initially proposed by Hinton [142, 160]. This method requires expert skills to select the best values of numerical meta-parameters [160]. If the training vector v is randomly selected, then the binary state, h_j of each hidden unit is set to 1. The logistic function $\sigma_x = (1/(1+\exp(-x)$ is used to compute the probability of turning a hidden unit, j:

$$p(h_j = 1 \mid \mathbf{v}) = \sigma(\sum_i v_i W_{ij} + b_j) \tag{3.6}$$

Then, applying the same probability to get an unbiased sample of the binary state of each visible unit will produce:

$$p(v_i = 1 \mid \mathbf{h}) = \sigma(\sum_i h_j W_{ij} + a_i) \tag{3.7}$$

Using a sigmoid function in deep belief learning and Gibbs sampling represents the basis of the inference and learning algorithm that is used to update the visible and hidden units [160].

Gibbs sampling is a particular type of Markov Chains Monte Carlo (MCMC) technique that is used in the joint probability distribution of multiple random variables to produce samples, and it can be classified into the class of Metropolis-Hastings algorithms [134]. The strategy of this method is to update all variables respectively depending on the conditional distribution of each unit that gives the state of other units [149].

3.2.2 Contrastive Divergence

Contrastive divergence (CD) is the objective function used in the learning process that gives the difference between two Kullback−Liebler divergences [160, 182]. The CD learning algorithm is a significant method for training RBMs [160]. This method requires expert skills to decide the best values of numerical meta-parameters such as initial values of the weights, weight cost, learning rate, number of hidden units, momentum, sparsity target and size of each mini-batch [160]. It is important to know how to deal with the learning procedure and how to control and terminate the training at a specific time. This learning algorithm only deals with binary visible and hidden units.

The following stages describe the learning procedure of a generative model:

* Use RBMs for DBN learning and the backpropagation for fine-tuning.

* Update the hidden states.

* Update the visible states.

* Collect statistics needed for learning.

3.2.3 Bernoulli-Bernoulli RBM

The first type of RBM is a Bernoulli-Bernoulli RBM. In this method the hidden and visible units have binary values [183]. The function of energy model can be represented by:

$$E(\mathbf{v},\mathbf{h}) = -\acute{b}v - \acute{c}h - \acute{h}W_v \tag{3.8}$$

where $\acute{b}$ is the bias vector for visible units, $\acute{c}$ is the bias vector for hidden units, $\mathbf{v}$ is the vector of visible units, $\mathbf{h}$ is the vector of hidden units, and W is the weight matrix, where W_{ij} represents the edge value between visible unit i and hidden unit j.

3.2.4 Gaussian-Bernoulli RBM

The second type of RBM is a Gaussian-Bernoulli RBM. The hidden and visible units in this type are different to the Bernoulli-Bernoulli RBM in that the visible units follow a Gaussian activation probability while the hidden units are normal binary victors that follow a Bernoulli distribution [131, 183]. The energy function of the Gaussian Bernoulli RBM can be represented by:

$$E(\mathbf{v},\mathbf{h}) = \sum_{i \in visible} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j \in hidden} b_j h_j - \sum_{i,j} \frac{v_i}{\sigma_i} h_j W_{ij} \tag{3.9}$$

where σ_i is the standard deviation of visible unit i and a_i is the mean of the i^{th} visible unit. The data set is normalized to have a zero mean and variance of 1.

3.3 Related Work

Deep learning, and in particular deep belief network, has recently witnessed increased attention from researchers as a new classification platform. DBNs have mainly been applied to image classification due to the hierarchical structure of images, where edges can be grouped to form segments, which when grouped form objects. They have also been successfully applied to a number of other classification tasks, such as speech recognition and natural language processing [35].

DBNs have been proposed for the classification of single-stream data. Multi-modality DBNs have also been investigated by a number of researchers. Srivastava and Salakhutdinov proposed a DBN of multi-modal system for learning a joint representation of data. They used images and text bi-model data to demonstrate that a multimodal DBN (MDBN) can learn good generative attributes. [173].

However, DBNs have not been fully explored for the classification of electroencephalogram (EEG) data, which records the electrical activity of the human brain using multiple electrodes on the scalp. Zheng et al. introduced a DBN to search the neural signature associated with positive and negative emotional categories from EEG data. They trained the DBN using different entropy features that were extracted from multichannel EEG data. They showed that the performance of DBN was better in higher individual frequency bands [30]. In addition to this, Wulsin et al. (2011) evaluated the classification and anomaly measurement of EEG data using an autoencoder produced by unsupervised DBN learning [33].

In the next section, the best DBN structure for EEG classification will be discussed. Three implementations of deep belief network are proposed. One is similar to the single-stream DBN, the second implements a single-stream DBN for each channel and then combines the classification outcomes of the different channels using another DBN, and the third utilises the concept of MDBN in proposing a new multi-stream (or multichannel deep belief network (MCHDBN)). In the third case a "partial" DBN with no decision layer is constructed for each channel, and the last hidden layers of the partial DBNs are combined using higher level hidden and decision layers.

An EEG dataset of depression patients that was recorded to study the modulatory effect of transcranial direct current stimulation (tDCS) is used to evaluate the proposed method. In order to evaluate its efficiency, the performance of the proposed method is compared with the three standard classification methods of support vector machine (SVM), linear discriminant analysis (LDA) and extreme learning machine (ELM).

3.4 Proposed Method: A Multichannel Deep Belief Network for the Classification of EEG Data

The study proposes three implementations of DBNs to classify multichannel EEG data based on different channel fusion levels as shown in Figure 3.1. The following points represent the three implementations of DBN classification methods:

- The first architecture (Figure 3.1a) is a traditional single-stream DBN that trains features extracted from all channels using a single network that consists of two or more RBMs. The process of training the RBM models is as follows: the hidden layer of the first RBM is the visible layer of the second RBM, and the hidden layer of the second RBM is the visible layer of the third RBM, and so on until arriving at the top RBM. This architecture is used to evaluate the ability of a single-stream DBN to extract high level information from features of the different channels.
- The second architecture 3.1b) consists of multiple DBNs, with one for each EEG channel, each of which is trained in a way similar to that of the single-stream DBN. In other words, each channel is trained separately by providing its features as input to the corresponding of the sample. The produced labels of these individual DBNs are then trained using another DBN to obtain the final decision. This architecture evaluates the ability of individual channels to correctly classify the data and the advantages of combining the classification results of individual channels.
- The third architecture, or the multi-stream DBN, (Figure 3.1c) is implemented using partial DBNs that are used to process individual channels. The top hidden

layers of the partial DBN are combined using unified hidden layer(s) that is (are) followed by the decision layer. The rationale behind proposing this architecture is to extract "local" information from each channel using the partial DBNs, while the higher level information is extracted using the combined hidden layers that fuse the local information of the different channels.

3.4.1 Data Set

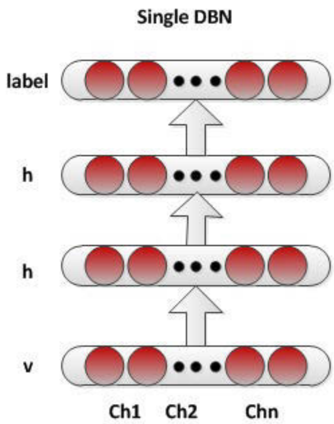
In this experiment, data was collected from 12 patients with depression who underwent transcranial direct current stimulation (tDCS) treatment [1]. The participants were asked to go through three EEG recording sessions [7]. Each session lasted for 20 minutes. In the first session, baseline EEG was recorded for each patient. This was followed by one session of active and one session of sham tDCS that were randomly ordered among the 12 patients. A scalp EEG was recorded while patients were sitting still with their eyes closed using 62 electrodes placed according to the international $10/20$ system.

The classification task aims to differentiate between the three session types using the recorded EEG data. The data of each EEG session was sampled at F = 2000 Hz. A Hanning window of 1024 samples with default an overlap of 0.5 of window length was applied and the power spectral density was estimated using the Fourier transform. The average PSD in the five conventional frequency bands, which are Delta $(0.5 \leq \delta < 4)$Hz, Theta $(4 \leq \theta < 8)$Hz, Alpha $(8 \leq \alpha < 13)$Hz, Beta $(13 \leq \beta < 30)$Hz, and Gamma $(30 \leq \gamma < 100))$Hz, were used as features.

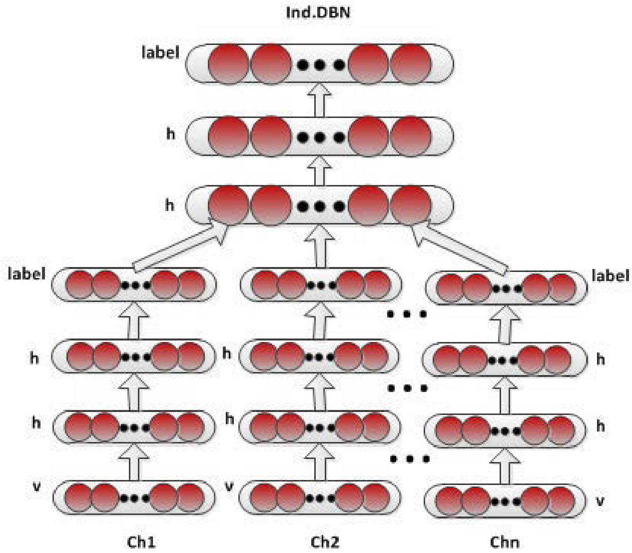
3.4.2 Experiments and Results

Two experiments were conducted, with two objectives, first, to rank channels based on their importance for this particular classification task, while the other aim is to evaluate the three proposed multichannel DBN implementations with the benchmark classifiers of linear support vector machine (SVM), linear discriminant analysis (LDA) and Extreme

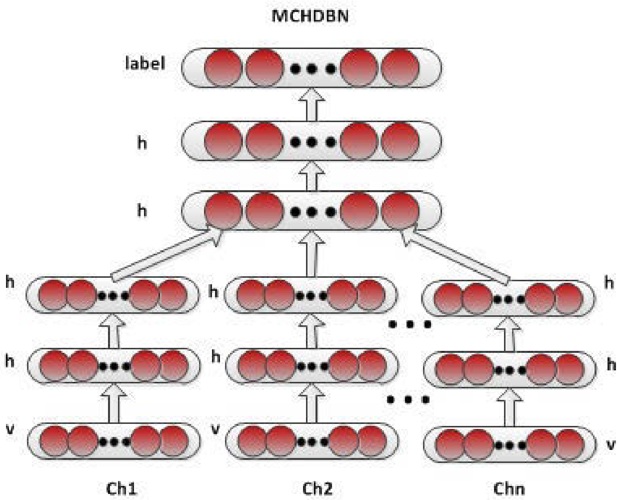
[1] Details about tDCS are given in Section 4.1.



(a)



(b)



(c)

Figure 3.1: DBN architectures: (a) Single-stream DBN, (b) Individual DBN for each channel and a top DBN to combine the channels' classification results, and (c) Multi-stream (or multichannel) DBN.

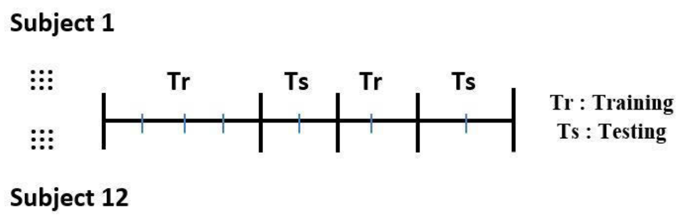


Figure 3.2: Testing and training divisions.

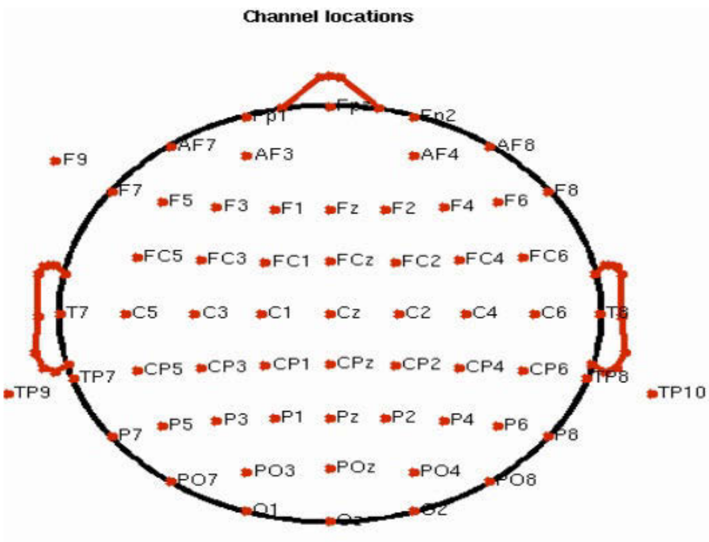
Learning Machine (ELM) classifiers to rank the performance of channels. The aim of using three classifiers is to identify channels that perform well for all three classifiers and, hence, reduce sensitivity to a particular classifier.

Data of each tDCS session was divided into 10 segments, and segments 5,6,9 and 10 were used for testing, while the remaining segments were used to train the classifiers, as shown in Figure 3.2. To increase the training data, each segment was divided into 10 windows.

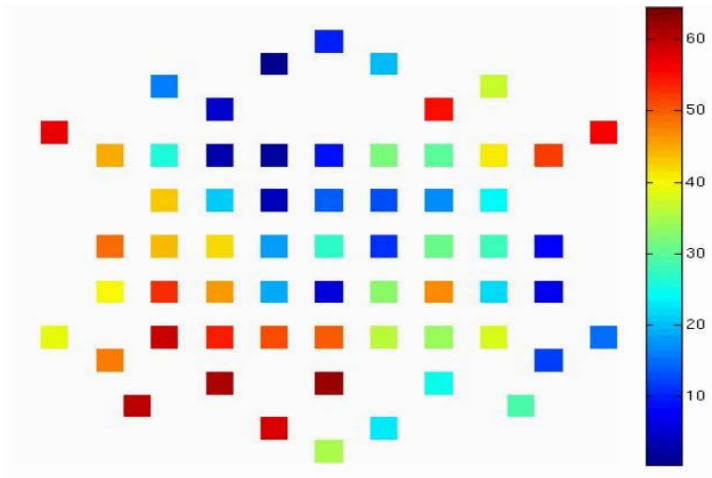
3.4.3 Evaluating the Performance of the EEG Channels

In the first experiment, the performance of individual channels was evaluated. In order to reduce fluctuations in the performance of channels, the evaluating subsets of four neighbouring channels is considered and then, the accuracy of the subsets that consist of that particular channel was averaged for each channel. For example, the accuracy of channel $C4$ is obtained by averaging the accuracy of four subsets $(C4, FC2, C2, FC4)$, $(C4, CP6, CP4, C6)$, $(C4, CP2, C2, CP4)$, and $(C4, FC6, FC4, C6)$. The 62 channels were ranked based on their classification performance, as shown in Figure 3.3. The color bar represents the rank order, and the best channels are a dark blue color. The figure shows that channels located over the frontal midline perform relatively better than the other regions of the brain and, hence, have the ability to differentiate between the three sessions of baseline, active tDCS and sham tDCS.

In the second experiment, the channel subsets of various sizes were constructed to get a better evaluation of the performance of the classifiers. These subsets were formed using frontal channels as well as the channels of other regions of the brain. The channel



(a)



(b)

Figure 3.3: (a) EEG electrodes montage. (b) Classification accuracy for individual channels (The color bar represents the rank order).

combinations are listed in Table 3.1. Unlike the last experiment, each channel is used by itself in this experiment.

In order to evaluate the performance of the three DBN architectures, they were used to classify the subsets of selected channels. For the single-stream DBN shown in Figure 3.1a, the two hidden layers consist of 200 and 150 units. For the other two DBNs, the four hidden layers consist of 200, 150, 200 and 150 units. The models were trained for 1200 iterations. The results obtained for the six classifiers shown in the Table 3.1 indicate that the ELM classifier performed noticeably better than both SVM and LDA. For the DBN architectures, the results obtained using the single-stream architecture were not found to produce very competitive results. The performance of the second DBN was slightly better than the first one, particularly for the large channel subsets. The multi-stream DBN, on the other hand achieved smaller error rates in most channel subsets in comparison to the other five classifiers, especially for the larger ones. Figure 3.4 shows a box-plot of the obtained classification error.

Table 3.1: Classification error rates of the six classification methods for each of the selected channel subsets

Channels	Error rates of the different classification Methods					
	SVM	LDA	ELM	DBN1	DBN2	DBN3
[F7,AF4]	0.507	0.534	0.243	0.396	0.361	0.264
[O1,AF3,FCz]	0.319	0.333	0.298	0.389	0.340	0.257
[TP9,CPz,POz]	0.423	0.416	0.312	0.359	0.403	0.299
[O1,TP9,P6,CPz]	0.340	0.354	0.305	0.375	0.389	0.257
[O1,O2,PO3,PO4,PO7,PO8]	0.465	0.423	0.312	0.382	0.306	0.278
[Fp2,TP9,AF3,AF4,CPz,POz]	0.306	0.305	0.229	0.340	0.340	0.243
[Fp1,Fp2,F7,F8,F9,F10,Fpz]	0.444	0.326	0.229	0.292	0.292	0.236
[Fp1,Fp2,AF3,AF4,AF7,AF8,Fpz]	0.333	0.263	0.277	0.271	0.243	0.215
[O1,O2,PO3,PO4,PO7,PO8,Oz,POz]	0.347	0.284	0.326	0.333	0.285	0.243
[P4,O2,P8,P2,PO4,P6,PO8,Oz]	0.368	0.368	0.263	0.368	0.278	0.257
[Cz,FC1,FC2,CP1,CP2,C1,C2,CPz]	0.410	0.314	0.256	0.361	0.312	0.271
[Fp1,F3,F7,F1,AF3,AF7,F5,F9,Fpz]	0.333	0.298	0.256	0.278	0.243	0.215
[Fp2,F4,F8,Fz,F2,AF4,AF8,F6,F10]	0.431	0.291	0.326	0.306	0.292	0.222
[P3,O1,P7,Pz,P1,PO3,P5,PO7,POz]	0.486	0.430	0.291	0.368	0.361	0.285
[C3,T7,FC5,CP5,TP9,FC3,CP3,C5,TP7]	0.424	0.416	0.333	0.319	0.299	0.229
[C4,T8,FC6,CP6,TP10,FC4,CP4,C6,TP8]	0.389	0.305	0.333	0.347	0.333	0.181
[Fp1,Fp2,F7,F8,AF7,AF8,F9,F10,Fpz]	0.396	0.222	0.259	0.299	0.222	0.167
[Fz,AF3,AF4,AF7,AF8,F5,F6,F9,F10,Fpz]	0.299	0.229	0.270	0.257	0.188	0.111
Average	0.390	0.340	0.284	0.336	0.305	0.235

The obtained results indicate that the DBN needs to be carefully designed when applied to the classification of EEG data, as traditional or single-stream architecture may not produce the expected outcomes. The same is also applicable to combining the classification results of individual channel DBN classifiers. The multi-stream DBN proved to be the most appropriate architecture, as it attempts to extract local attributes from each channel and combine those at a higher level. It is worth mentioning that this research did not attempt to optimise the computational cost of the DBN algorithms and, therefore, their execution time was noticeably higher than the other three classifiers.

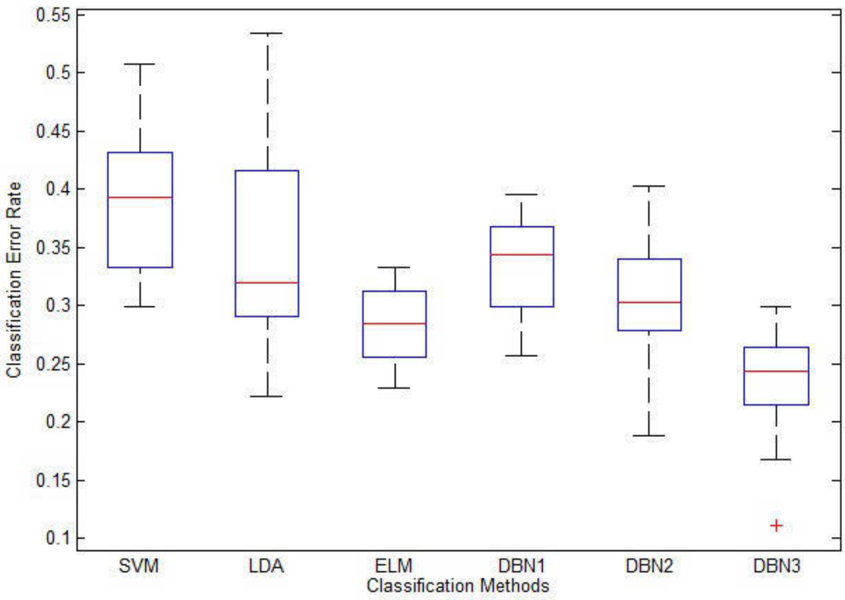


Figure 3.4: Classification error of the six classification methods.

3.5 Summary

In this chapter, the concepts and principal operation of the DBN as a special technique of deep learning was explained. EEG data of 12 patients with depression was used in this experiment and the classification task aimed at differentiating between baseline, sham tDCS and active tDCS sessions. Three proposed DBN implementations were presented to classify multichannel EEG data based on different channel fusion levels. Two experiments were presented in this chapter. In the first experiment, the performance of each EEG channel was evaluated by determining the average of the accuracy of subsets of four neighbouring channels. In the second experiment, the performance of classifiers were evaluated by using different subsets of EEG channels. The multi-channel deep belief network produced very promising results when compared to three well-established classifiers SVM, LDA and ELM especially when classifying subsets that consist of multiple channels.

Chapter 4

Predicting Transcranial Direct Current Stimulation Treatment Outcomes of Depression Patients Using Automated EEG Classification

4.1 Introduction

Recent years have witnessed the emergence of transcranial direct current stimulation (tDCS) is a promising non-invasive and safe method for treating neuropsychiatric disorders and a tool for modulating cortical activity [73, 74, 82, 184, 185]. Transcranial direct current stimulation involves applying a low current, typically 1-2 mA, across the brain through two or more electrodes placed on the scalp. Compared to other techniques of brain stimulation, tDCS has several practical advantages, such as cost effectiveness and minimal adverse effects [29, 186], as well as being clinically effective intervention [77].

This chapter presents an investigation of the ability of spectral power of resting-state EEG in the classification of mood and cognition improvements that are achieved following tDCS treatment. EEG data of the three sessions of baseline was used in the initial investigation, active tDCS and sham tDCS to perform this classification task. The classification accuracy of mood and cognition were averaged to find the most relevant channels/regions. A more in-depth or advanced evaluation was then conducted, where only baseline EEG for the estimation of mood and cognition improvement is considered,i.e., before commencement of the tDCS treatment sessions. Individual channels as well as pairs of channels are assessed to identify relevant brain regions for each one of two measures.

4.2 Related Work

A number of studies have shown that tDCS can reduce depressive symptoms and improve cognitive functioning of depressed patients [3, 187]. However, up to 80% of patients do not respond to current forms of tDCS treatment, which deliver the same intensity of brain stimulation to all participants [2]. Of patient factors determining treatment response, a recent analysis based on individual patient data pooled from several randomized controlled trials only identified medication resistance as significant [188].

The neurophysiological mechanisms involved in the antidepressant effects of tDCS remain incompletely understood. Hence, two recent techniques have proposed to use neuroimaging to identify changes in brain activity following tDCS treatment [189, 190]. Electroencephalography (EEG) plays a vital role in understanding the underlying neurophysiological states surrounding neuropsychiatric disorders such as MDD and for discovering biomarkers or diagnostic tools relating to these disorders [24]. The neuromodulatory effects of tDCS on cortical activity for the treatment of neuropsychiatric disorders can be readily studied using EEG [7]. In people with depression, EEG reveals an asymmetry in frontal alpha activity, i.e. lower alpha power in the right hemisphere compared to the left [191, 192]. Alpha asymmetry is disorder specific [193], and can be used to predict antidepressant treatment response [194, 195]. Powell et al. used EEG to study the modulatory effect of tDCS on changes in cortical activity in people with mood disorders [7]. A multi-channel deep belief network can be used to accurately classify EEG data that was recorded after active or sham tDCS [196], this represents the first contribution in this research. In addition, Wozniak-Kwasniewska et al. showed that EEG oscillatory activity was significantly different for depressed patients that responded to repetitive transcranial magnetic stimulation (rTMS) therapy compared to non-responders, suggesting that baseline EEG has predictive value for brain stimulation treatment outcomes [197].

Normally, treatment efficacy can only be evaluated after participants have completed all treatment sessions. This negatively impacts participants that do not respond to the treatment and is an inefficient use of staff time and resources. In this work, EEG data recorded at the start of tDCS treatment is proposed to predict which participants will improve in mood and cognition following treatment. Being able to predict the outcome of tDCS treatment from baseline measures may allow selection of patients most likely to respond to tDCS.

4.3 Methods

4.3.1 Participants

The experiment was conducted at the Black Dog Institute after obtaining approval from the Human Research Ethics Committee of the University of New South Wales. Data sets from 10 participants is used in this study. The participants meeting formal diagnostic criteria for major depressive disorder (MDD) through a clinical trial at the Black Dog Institute through investigating the efficacy of tDCS treatment for depression [3]. Written informed consent was obtained from all participants prior to study enrolment in accordance with the National Health and Medical Research Council guidelines and the Human Research Ethics Committee of the University of New South Wales and all research conducted abided by the Australian Code of Responsible Conduct of Research. The Mini-International Neuropsychiatric Interview (MINI, [198]) with the psychiatrist's confirmation was used to diagnose participants in a semi-structured interview. All participants had unipolar major depressive episode with a Montgomery-

Åsberg Depression Rating Scale (MADRS, [199]) of ≥ 20 at study entry. Table 4.1 represents the demographic and clinical information of participants.

Table 4.1: Demographic and clinical information

	n	mean	SD
Gender: male/female	5/5		
Age		41.8	13.3
MADRS		28.7	5.1
CGI		4.3	0.5
Age of onset		24.8	10.4
Current episode (weeks)		23.0	29.5
All prior episodes (weeks)		59.0	66.8
QIDS-SR		14.7	3.6
Melancholia	5/10		
Dysthymia	3/10		
Concurrent use of antidepressants	7/10		
Failed antidepressant medication trials in current episode		1.7	1.4
Failed antidepressant medication trials in past episodes		0.9	2.4

SD: Standard Deviation, MADRS: Montgomery-Åsberg Depression Rating Scale, CGI: Clinical Global Impression, QIDS-SR: Quick Inventory of Depressive Symptomatology

4.3.2 Protocol

All data were originally acquired from participants entering a double-blind clinical trial to investigate the efficacy of tDCS treatment [3]. These participants were formerly assessed for eligibility then randomised to receive either sham or active tDCS in 15 treatment sessions given over three weeks. All participants were then offered an additional 15 sessions of open-label active tDCS given over an additional three weeks. Psychiatric assessment of mood was assessed using the MADRS at baseline, session 8 and 15, 23 and 30, and at 1 week and 1 month after trial completion. Assessments at session 23 and 30 were part of the open-label phase. Neuropsychological assessment of acute cognitive effects was conducted using the Symbol Digit Modalities Test (SDMT; [200]) immediately before and after session 1. Each participant was assessed by the same blinded rater for mood evaluation using the MADRS.

Participants were invited to participate in an EEG study prior to starting the clinical trial. EEG activity was acquired during rest and during a cognitive task at baseline, after a single session of active tDCS and after a single session of sham. Data from the ten patients enrolled in the active arm of the clinical trial was used. That is, two patients diagnosed with bipolar disorder were excluded. For the current study, resting-state EEG recorded at baseline is used. The duration between this baseline EEG and entry into the clinical trial was approximately 12 days.

The labels were assigned based on mood and cognition improvement, which is used for machine learning classification based on data obtained during the clinical trial. Improvement in mood was determined based on MADRS obtained at baseline (S0) and after sessions 15 (S15) and 23 (S23). We did not use the MADRS after session 30, as not all participants completed this. The cognitive score is calculated using the Symbol Digital Modalities Test (SDMT) at baseline and after the first session of tDCS.

4.3.3 EEG Acquisition

For the EEG acquisition participants were seated in a light and sound attenuated room. Continuous eyes-closed resting-state EEG was acquired for 10 minutes. EEG data was acquired using a 64-channel BrainAmp MR Plus amplifiers (Brain Products, Munich, Germany, hardware bandpass filter $0.1-250$ Hz, resolution 0.1 μV, range $+/23.3$ mV) and custom electrode caps (Easy Cap, FalkMinow Services, Herrsching-Breitbrunn, Germany) with electrodes placed according to the international $10/20$ system). Figure 4.1 represents the EEG cap that was used to collect data at the Black Dog institution. All data were referenced against an electrode centred on the midline between Fz and Cz, and sampled at 5 kHz. Electrode impedance was kept below $5k\Omega$. The electrooculogram (EOG) and two electrocardiogram (ECG) channels were also recorded.

4.3.4 Transcranial Direct Current Stimulation Treatment

Each session of transcranial direct current stimulation treatment was applied using an Eldith DC-stimulator (NeuroConn GmbH, Germany) with the anode placed over the left dorsal lateral prefrontal cortex (DLPFC) (identified as F3 on the international 10/20

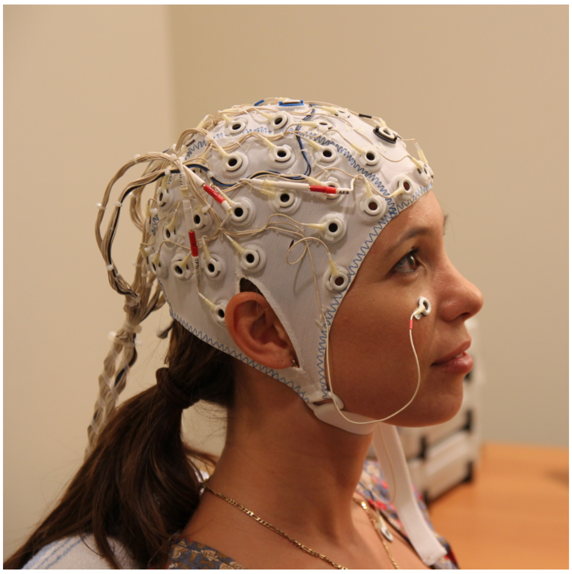


Figure 4.1: EEG cap.

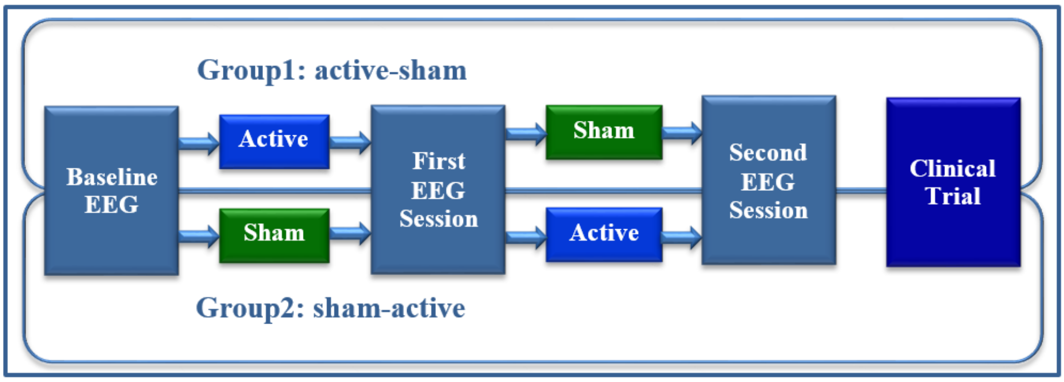


Figure 4.2: Processing of transcranial direct current stimulation treatment.

EEG system) and cathode located over the lateral aspect of the contralateral orbit (at the F8 position). The active stimulation protocol included applying 2 mA for 20 min, with a gradual ramp up/down of the current over 30 s.

The participants in tDCS experiment were divided into two groups, the first group received active tDCS followed by sham and the second group received sham followed by active tDCS as shown in Figure 4.2 [7]. These sessions were received randomly. The two sessions of active and sham tDCS were performed 7-8 days apart [7].

4.3.5 EEG Data Analysis

Data from the 61 EEG channels (excluding EOG and ECG channels) were pre-processed using Brain Vision Analyzer software (version 2.0.3). Data was down sampled to 2 kHz and high-pass filtered (Butterworth filter with 0.5 Hz cut-off frequency). Artefacts were removed using independent component analysis (ICA) using the infomax algorithm. Components containing extraneous eye, heart and electromyographic signals were rejected and remaining components back transformed to yield artefact cleaned EEG data. Cleaned EEG data was then re-referenced to the average reference resulting in 62 channels.

EEG data was segmented into 10 non-overlapping 1-minute segments. Welch method was used to estimate the power spectral density (PSD) in each segment and EEG channel using a Hanning window of 60 sec. Average PSD in conventional EEG frequency bands were used as data features: delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (13-30 Hz) and gamma (30-100 Hz). PSD was averaged in these frequency bands to reduce the feature space and avoid overfitting.

Alpha asymmetry in frontal, central and parietal cortex was also examined. Alpha asymmetry indices were calculated by subtracting log power of the left hemisphere electrode from log power of the homologous right hemisphere electrode [193, 201]. Given the inverse relationship between alpha power and cortical activity [202], positive alpha asymmetry scores are interpreted as left-frontal activity relative to right, while negative scores are interpreted as right-frontal activity. Analyses were conducted on a frontal pair (F3, F4), a central pair (C3, C4), and a parietal pair (P3, P4).

4.3.6 Experiment

In this study, two investigations were conducted; initial investigation and advanced investigation. In the first investigation, EEG data of three tDCS sessions were used (baseline, active tDCS and sham tDCS). In the advanced investigation, only the resting-state EEG data that was recorded at baseline was used. In the two investigations, a

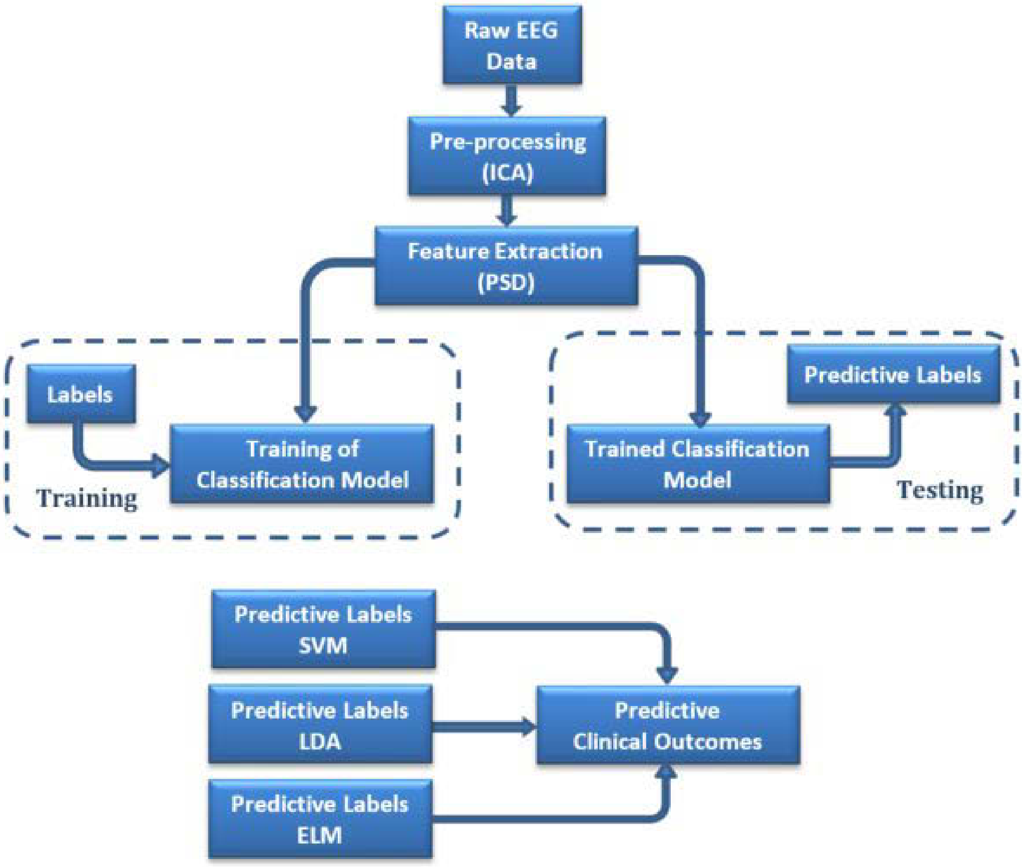


Figure 4.3: Flowchart of the classification framework.

subject independent approach was adopted to differentiate between participants who responded to the treatment from those who did not, as identified by the mood and cognitive scores. A leave-one-out cross validation method [203] was used to split the data into training and testing sets, i.e. training with nine participants and testing on the remaining one. The three classification methods of support vector machine (SVM), linear discriminant analysis (LDA) and extreme learning machine (ELM) were used to classify the features into one of two classes (low or high improvement). Then the results of these three classifiers were averaged, where the aim of using three classifiers that are based on different learning mechanisms was to minimise biases associated with a particular classification method. This bias may cause differences in the ranking of channels based on their performance. Hence, averaging the results of three classifiers helps in identifying channels that perform well by the three classifiers. A flowchart of the classification framework is shown in Figure 4.3.

Table 4.2: Scores and labels for individual participants

Participants	MADRS			SDMT		Labels	
	S0	S15	S23	S0	S1	Mood	Cognition
1	23	17	18	54	62	0	1
2	35	16	28	52	63	1	1
3	32	11	28	51	58	1	1
4	35	22	13	46	52	1	1
5	27	15	16	58	58	0	0
6	26	18	17	56	58	0	0
7	27	18	4	60	61	1	0
8	24	13	12	40	42	1	0
9	35	22	23	47	54	0	1
10	23	21	20	45	50	0	1

MADRS: Montgomery-Åsberg Depression Rating Scale, SDMT: Symbol Digital Modalities Test

To evaluate mood score, a label of 1 was assigned to those participants who had a MADRS at S15 or S23 that was less than half the MADRS at S0 (responders) and a label of 0 was assigned to the remaining participants (non-responders). A label of 1 was assigned to those participants who had an increase in SDMT of 5 points or more (high improvement) and a label of 0 was assigned to the remaining participants (low improvement). Table 4.2 lists the mood and cognitive scores and labels for the 10 participants. The mood and cognition scores did not always match: Two participants showed a low improvement and three participants showed high improvement on both. The remaining five participants showed mixed scores for mood and cognition. These findings are summarised in Table 4.3.

The cognitive score was initially divided into three groups based on the level of cognitive improvement: low, medium and high. However, as only one participant achieved a high improvement based on this score, we decided to combine that participant's data with the ones that were considered to have medium cognitive improvement. The classifiers were separately tested for the mood and cognitive scores.

Evaluating the performance of each channel pair was considered, as this enables the evaluation of cross-region performance and may further improve the classification accuracy, where the inclusion of two channels could provide useful information for this

Table 4.3: Description of Mood and Cognitive Improvement Scores

No. of Participants	**Mood Score**	**Cognitive Score**
2	> 0.5MADRS1	<SDMT1+5
2	<= 0.5MADRS1	<SDMT1+5
3	> 0.5MADRS1	>=SDMT1+5
3	<= 0.5MADRS1	>=SDMT1+5

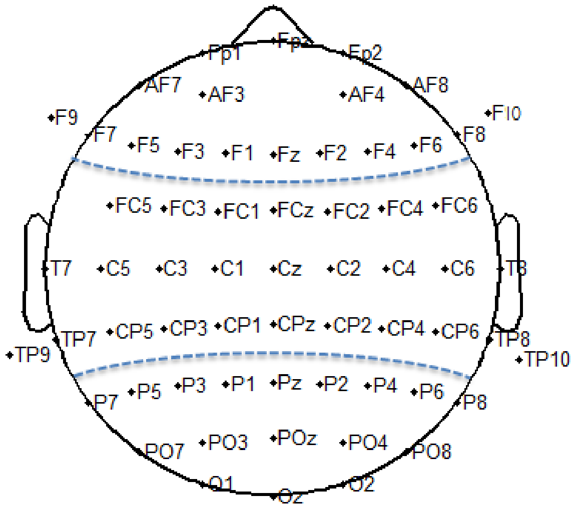


Figure 4.4: Brain regions and electrode placements.

classification task. To assess regional differences in classification accuracy, additional exploratory analyses was performed through dividing the brain into three main regions: frontal, central/parietal and parietal/occipital as shown in Figure 4.4. Firstly, the performance of the 62 EEG channels was ranked based on their ability to discriminate between responders and non-responders for the mood and cognitive scores. Then, the classification performance was calculated of all channel pair combinations within each brain region and between the different regions. As mentioned earlier, the classification accuracy was calculated by averaging the performance of the three classifiers for all ten participants based on the mood and cognitive scores individually.

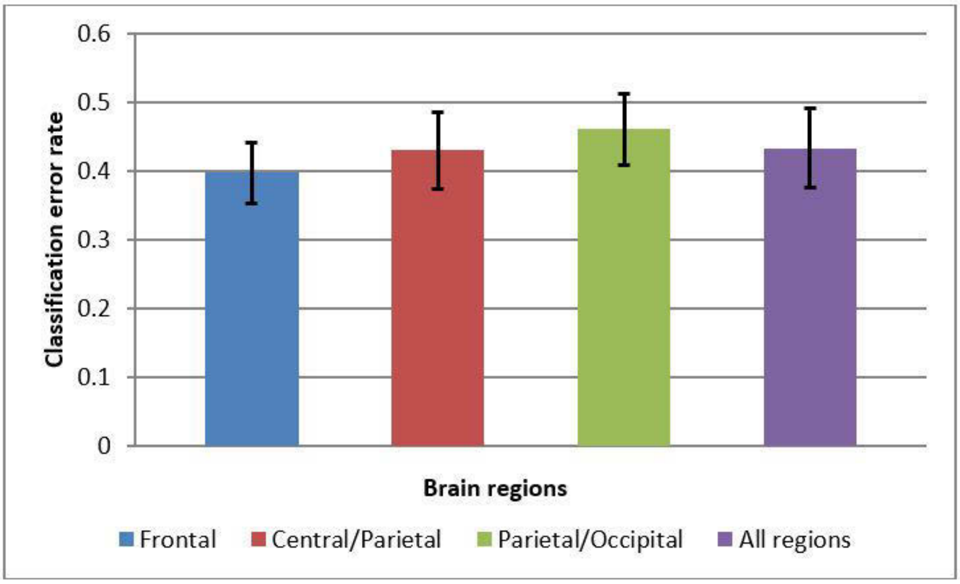


Figure 4.5: Error rates of the channel pair combinations

4.4 Initial Investigation: Results and Discussion

The average classification error rates of the channel pair combinations for the three brain regions as well as that of all regions are shown in Figure 4.5. The error rate of each channel pair is calculated by averaging the error rates of the three classifiers for all ten participants and for the mood and cognitive scores. The results indicate that the frontal area is performed better than the other regions of the brain. According to the pairwise ANOVA results listed in Table 4.4, there is a significant difference between the classification results of the brain regions, apart from the central/parietal and all regions.

The best two channel combinations within each of the three brain regions are shown in Figure 4.6. As for the case when considering all regions, the best two channel pairs were found to be {AF8 and Cz} and {AF8 and C1}, i.e., frontal and central channels. The error rates of the best two channel pairs of each region are shown in Table 4.5. These results confirm the superior performance of the frontal region for this classification task, and show that a lower error rate can be obtained by considering cross-region channels.

It is important to mention that the error rate was calculated by averaging the results of the thirty time segments for each participant (ten segments for each of the three sessions of baseline, active and sham). In order to obtain a discrete response for each

Table 4.4: Pairwise ANOVA of the obtained classification results

Brain Regions	Central/Parietal	Parietal/Occipital	All regions
Frontal	< 0.05	< 0.05	< 0.05
Central/Parietal		< 0.05	0.566
Parietal/Occipital			< 0.05

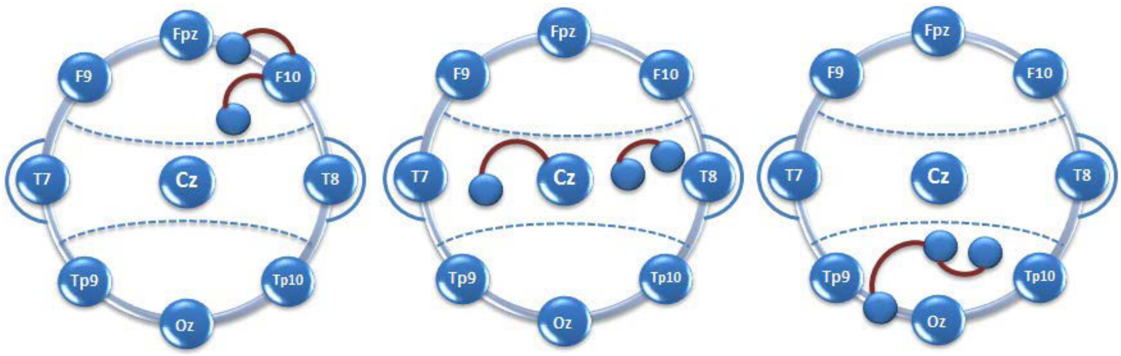


Figure 4.6: Best two channel pairs for each brain region.

participant, i.e., improved/not improved, a binary outcome for each participant based on a majority vote was calculated. For the channel pair {AF8 and Cz}, this has led to further reduction in the error rate, as it became 0.2167 (average of three classifiers and for both mood and cognitive scores). Also, rather than averaging their errors, if the classification results of the three classifiers for the thirty time segments are combined using a majority voting approach, then the averaged error rate (averaged across mood and cognition) of the channel pair {AF8 and Cz} becomes 0.2.

More specifically, one out of ten participants had a misclassified cognition label, while for mood; three participants had misclassified labels. These initial results could have an important influence on the adoption of tDCS treatment for depression, by enabling the early identification of patients more likely to respond, thus avoiding treatment delays and saving staff time and resources.

Table 4.5: Average Error rates of the best two channel pairs for the different brain regions.

Frontal		**Central/Parietal**		**Parietal/Occipital**		**All regions**	
Best channel pairs	Error	Best channel pairs	Error	Best channel pairs	Error	Best channel pairs	Error
(F4,F10)	0.2694	(C3,CPz)	0.3128	(Pz,P2)	0.2994	(AF8,Cz)	0.2411
(F6,F10)	0.285	(CP2,C1)	0.3144	(Pz,PO7)	0.3311	(AF8,C1)	0.2644

4.5 Advanced Investigation

4.5.1 Experimental Setting and Evaluation

Classification was separately performed for each of the three classifiers using ten data segments. A leave-one-out cross validation method was used, hence 3×10 tests for each participant were performed. The classification accuracy was quantified as the ratio of correctly predicted labels giving a number between 0 (predicted label was always incorrect) and 1 (predicted label was always correct). To determine whether the classification results were statistically significant, classification accuracy was compared for each participant against 0.5 (expected classification accuracy under the null hypothesis) using a one-sample Wilcoxon signed rank test. To compare the classification error across brain regions, classification accuracy was averaged across all channel pairs within each of the regions (frontal, central/ parietal and parietal/ occipital) and compared the average accuracy using a 3×2 (regions $\times$ labels) repeated-measures ANOVA.

4.5.2 Experimental Results

First, the classification accuracy using a single EEG channel was assessed. Channel Tp9 performed best for the mood labels ($accuracy = 71 \pm 11\%$) and channel Pz performed best for the cognition labels ($accuracy = 87 \pm 5\%$). The classification accuracy of individual channels for both labels is shown in Figure 4.7. These results indicate that the frontal channels produced on average better results than other brain regions for the mood labels. For the cognition labels the parietal channels yielded the highest classification accuracy.

Then, the classification accuracy using a pair of EEG channels was assessed. For the mood labels the highest classification accuracy was observed for channel pair FC4-AF8 ($76 \pm 11\%$; Figure 4.8.A). The accuracy was significantly higher than 50%($p = 0.034$). Using a majority vote, the mood labels were accurately predicted in 8 out of 10

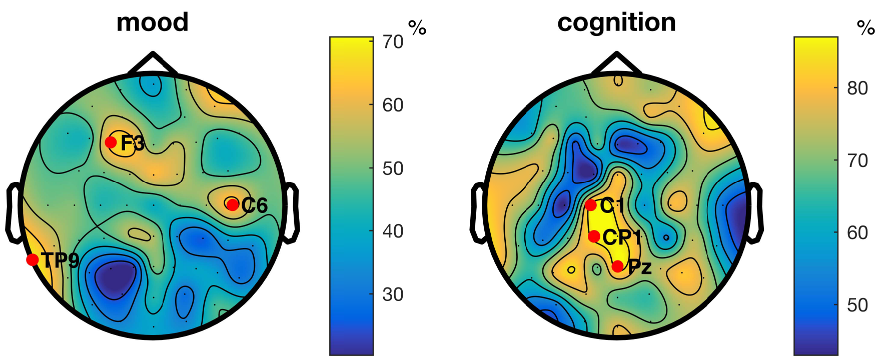


Figure 4.7: Classification accuracy of individual channels based on: (a) mood labels (left) and, (b) cognition labels (right).

participants. For the cognition labels the highest classification accuracy was observed for channel pair CPz-CP2 ($92 \pm 4\%$; Figure 4.8 A). The accuracy was significantly higher than 50%($p = 0.004$). Using a majority vote, the cognitive labels were accurately predicted in 10 out of 10 participants. The best performing channel pair for mood labels were located over frontal brain regions, while for cognition labels the best pair was located over parietal brain regions (Figure 4.8 A). Classification accuracy was not statistically different for mood and cognition labels (p = 0.31).

To investigate the brain regions that best predict the mood and cognition labels, the error rates across regions were compared. Table 4.6 shows the three channel pairs that produced the lowest classification error for each region. Also the average error rate was compared across all channel pairs within each of the three regions. Figure 4.8 A. A (3×2) repeated-measures ANOVA showed no main effect of region or label (mood or cognition). However, the interaction effect was statistically significant (F (2,18) = 5.8, p = 0.011). For mood labels the average classification was lowest for frontal channels, while for cognition labels the error was lowest for parietal-occipital channels. Figure 4.8 B.

As mentioned earlier, the classification results of the three classifiers (SVM, LDA and ELM) were averaged. Evaluation of the individual performance of these three classifiers is provided in Figure 4.9. The figure shows that the performance of the three classifiers are comparable, however, they are not consistent in terms of performance in different brain regions for both the cognition and mood scores. For example, the

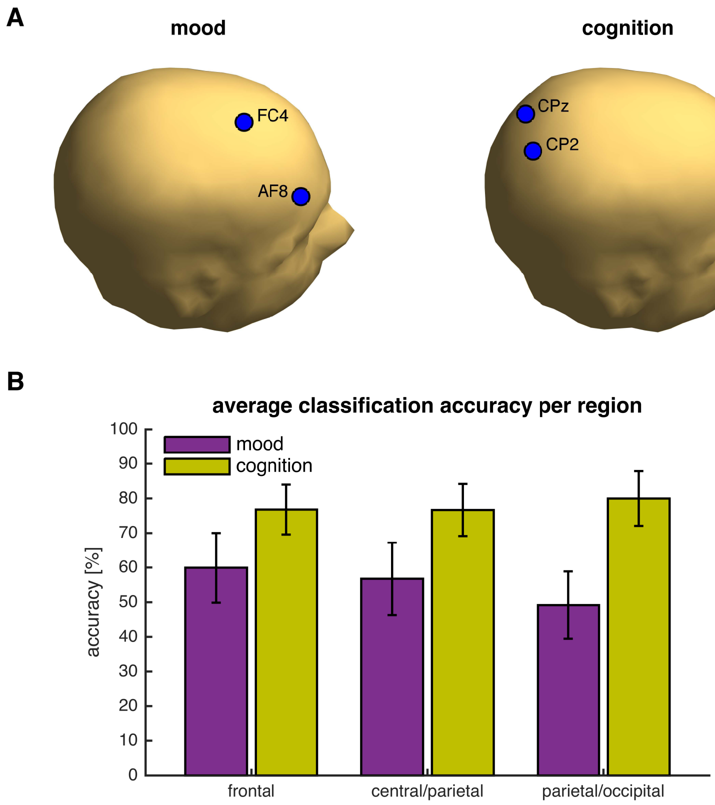


Figure 4.8: Classification results for pairs of EEG channels. A) Channel pair FC4-AF8 has the highest classification accuracy for the mood labels (left panel) and the best channel pair CPz-CP2 for the cognition labels (right panel). B) Average classification accuracy for each of the three brain regions (frontal, central/parietal and parietal/occipital).

ELM classifier is found to produce slightly lower accuracy compared to the other two classifiers for the classification of cognition, while it produced on average the highest accuracy for mood classification. Moreover, there are some differences in ranking across brain regions, for example the central/parietal channels were found to produce slightly better mood classification results than the frontal channels when considering the ELM classifier, which was not the case for the other two classifiers. Accordingly, we decided to average the results of the three classifiers to reduce the impact of inconsistency between them.

Finally, alpha asymmetry in three electrode pairs (a frontal, central and parietal pair) was examined. The asymmetry index was negative over frontal and central brain regions, but positive over parietal regions (Figure 4.10). This suggests greater right-lateralized frontal activity. The right-sided asymmetry seemed to increase in responders

Table 4.6: The average classification error for the three best channel combinations per brain regions

Brain Region	Mood		Cognition	
	Channel pair	Accuracy	Channel pair	Accuracy
Frontal	AF3-AF8	71%	F2-F10	88%
	AF8-F9	71%	F4-Fp2	85%
	F4-AF8	70%	Fp1-Fp2	83%
Central/parietal	T8-C1	73%	CP2-CPz	92%
	T8-CPz	71%	C4-CPz	89%
	T8-Cz	70%	FC6-CP6	87%
Parietal/occipital	Pz-P2	68%	P3-PO3	89%
	Pz-PO4	62%	P3-Oz	88%
	Pz-PO3	61%	P3-O1	86%
All regions	FC4-AF8	76%	CP2-CPz	92%
	T8-C1	73%	P3-PO3	89%
	T8-CPz	71%	C4-CPZ	89%

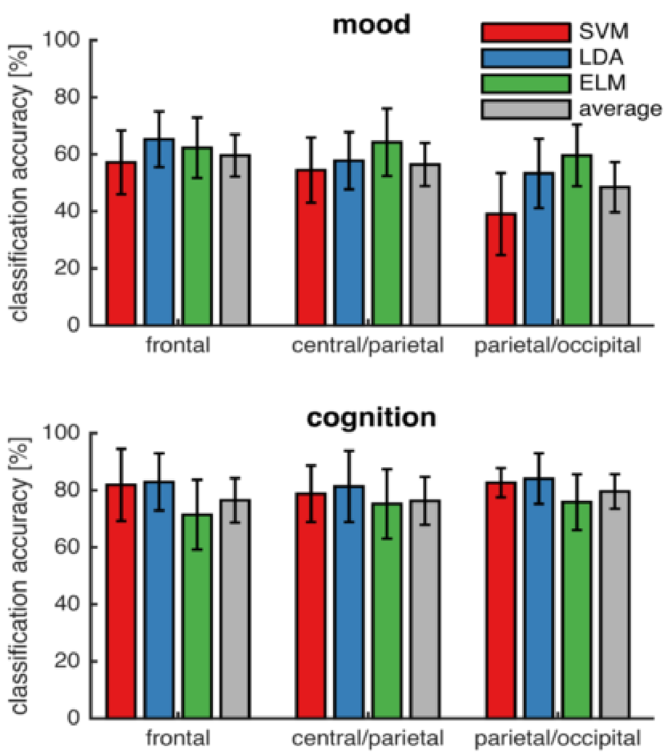


Figure 4.9: Classification accuracy of individual classifiers (SVM, LDA and ELM) and their average for the different brain regions. Top panel shows classification accuracy for mood labels and lower panel for cognition labels.

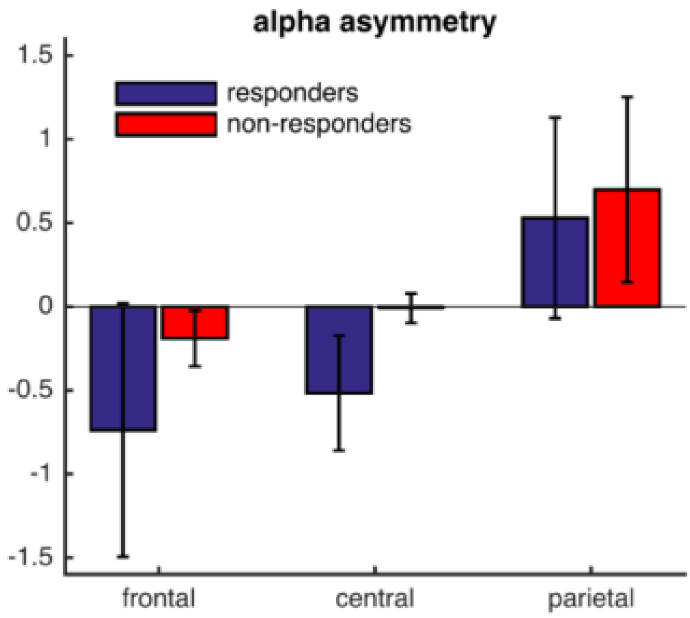


Figure 4.10: Alpha asymmetry in frontal (F4-F3), central (C4-C3) and parietal (P4-P3) regions for responders (n=5) and non-responders (n=5), as based on the mood labels. [Negative values indicate higher alpha power in the left hemisphere and hence greater right-lateralised activity].

compared to non-responders, although the standard errors are high given the low number of participant in each group (n=5).

4.5.3 Discussion

This study investigated the feasibility of utilising EEG data in predicting the clinical and cognitive outcomes of patients with MDD undergoing the tDCS treatment. The EEG data was collected before tDCS treatment and was used to predict the level of improvement in both mood and cognition during a subsequent course of treatment sessions. Spectral power was observed in frontal scalp channels yielded above chance accuracy in classifying participants who showed improvement in mood following tDCS treatment. Parietal channels, on the other hand, are found to perform well in predicting cognitive improvement. When evaluating channel pairs across all brain regions, channel pair FC4-AF8 are correctly predicted the mood labels of 8 out of 10 participants, while channel pair CPz-CP2 predicted the cognition labels of 10 out of 10 participants. Comparing the average classifications across three brain regions, frontal channels performed best for mood labels and parietal-occipital channels performed best for cognition labels.

There is a scarcity of information regarding clinical or individual predictors of mood response for tDCS treatment for depression. In a recent meta-analysis of individual data for 289 patients, Brunoni et al. identified antidepressant treatment resistance was a negative predictor of response, though was unable to identify any other clear clinical or demographic predictors. Recently, Martin et al. showed that better pre-treatment letter fluency performance predicted better antidepressant response [204]. The current study showed that baseline resting-state EEG activity, specifically in frontal channels, held predictive information regarding patient subsequent mood response to a course of tDCS treatment. The channel pair identified corresponded with cortical regions stimulated by the electric field using this electrode montage as suggested by computer modelling [205]. Interestingly, abnormal EEG activity within the right frontal cortex, the region underlying the channel pairs identified, has previously been shown to differentiate depressed patients and healthy controls [206]. Specifically, depression was characterised by decreased right frontal alpha power and conversely increased left frontal alpha power (i.e., asymmetry), with greater asymmetry correlated with higher depression scores.

Baseline alpha asymmetry has further been identified as a possible discriminator of response to pharmacological treatments, with responders showing greater alpha power over the right compared to the left hemisphere [195, 207]. Also alpha asymmetry was assessed and found increased left frontal alpha activity consistent with previous studies [191, 193]. Responders appeared to have higher alpha asymmetry and slightly higher MADRS at baseline (30.6) compared to non-responders (26.8). However, both groups consisted of only 5 participants and the difference does not reach statistical significance $(p \geq 0.2)$. Larger studies are required to distinguish the relative predictive value of these markers of alpha asymmetry.

Our results further showed that EEG parietal channels were predictive of cognitive improvement following a single session of tDCS, as measured using the SDMT. SDMT assesses processing speed and working memory functioning and is commonly used as a screening tool for frontal lobe dysfunction [200]. Loo et al. found in a RCT that active compared to sham tDCS caused acute cognitive enhancing effects on this task in depressed patients, indicating acute effects on processing speed and working memory functioning. Moreno et al. similarly reported acute performance enhancing effects following a single session of active tDCS in depressed patients measured using an N-back working memory task. These findings are consistent with similar acute performance enhancing effects of tDCS in healthy participants [208, 209]. The current findings extend these prior results to indicate that these effects with prefrontal tDCS stimulation have electrophysiological correlates primarily at central/parietal sites. Indeed, decreased alpha activity at the precuneus with increased workload on a working memory task has been shown, suggestive of functional inhibition within this region with increased attentional demands [210]. Whilst the duration of acute cognitive effects of tDCS following stimulation remain unclear, the potential to predict improved cognitive enhancement following a single tDCS session may additionally have functional utility for patients, particularly given the impact that depression related cognitive deficits have on patient functioning [211]. The mood and cognition labels did not match in all participants, suggesting that improvements in mood and cognition scores are partly independent.

The current findings complement the results of Wozniak-Kwasniewska et al., who showed that pre-treatment resting-state EEG differentiated between responders and non-responders to rTMS therapy [197]. Although both studies involved small cohorts (10 MDD patients in the current study and 8 MDD and 10 BP patients in Wozniak-Kwasniewska et al.), together these promising results suggest that baseline resting-state EEG may have predictive value based on the treatment outcomes of non-invasive brain stimulation treatment of MDD [197].

It is important to mention that in the current study, due to the limited number of participants and following a subject independent testing approach, some channel pairs produced an error rate that is close to random guess (error of 0.5). The reasoning is that these channels could not generalise and differentiate between the two classes when trained with data of nine participants only and tested on data of the remaining participant. According to Raudys and Jain, one of the significant problems that face many classification tasks is the small number of training samples. For this highly complex EEG classification task, a subject-independent classification approach would require around 50 participants. Data from a similar number of participants has been used in [212]. Hence, the findings of this preliminary study warrant a larger study to determine the clinical utility of this approach.

4.6 Summary

In this chapter, prediction of the clinical outcomes of depressive disorder based on tDCS treatment was proposed. Improvement in mood and cognition was examined to predict response to the tDCS treatment by analysing EEG of MDD patients. Power spectral density was assessed in five frequency bands: delta, theta, alpha, beta and gamma. Improvements in mood and cognition were assessed using the Montgomery-Åsberg Depression Rating Scale and Symbol Digit Modalities Test, respectively. Three machine learning techniques (SVM, LDA and ELM) and a leave-one-out cross-validation approach were used to carry out this classification task. Ten participants with a current diagnosis of MDD were included. The obtained classification results of all channel pair combinations were used to identify the most relevant brain regions and channels for this classification task. This investigation study was presented in two cases: initial investigation and advanced investigation.

First investigation was achieved based on features of resting state EEG of three tDCS sessions to differentiate participants who respond to the tDCS treatment from those who do not. In the second investigation, the resting-state EEG that is recorded prior to the commencement of treatment (baseline EEG) was assessed to identify MDD patients that respond to tDCS treatment. Results showed that mood labels were accurately predicted in 8 out of 10 participants using EEG channels FC4-AF8. Cognition labels were accurately predicted in 10 out of 10 participants using channel pair CPz-CP2.

Due to the limited number of participants (n=10), the presented results mainly aim to serve as a proof of concept. These finding demonstrate the feasibility of using machine learning to identify patients that will likely respond to the tDCS treatment. These promising results warrant a larger study to determine the clinical utility of this approach.

Chapter 5

Estimating the Quality of Electroconvulsive Therapy Induced Seizures

5.1 Introduction

Electroconvulsive therapy (ECT) is an effective treatment for psychiatric disorders such as depression [13]. ECT is a procedure done under general anaesthesia in which small electric currents are passed through the brain, intentionally triggering a brief seizure. ECT is safe and is considered as one of the most effective treatment methods for depression [13]. The efficacy of electroconvulsive therapy may depend on the placement of the stimulating electrodes and the electrical dosage used, among other treatment parameters, and that the quality of the induced seizures is related to the efficacy of ECT [89, 213–215]. Well-defined criteria for seizure quality are therefore useful for the evaluation of the therapeutic effectiveness of ECT treatments.

This chapter will present two cases studies for predicting the seizure quality rating based on other seizure indices. The first study aims to identify the seizure quality rating by utilising a set of already rated seizure parameters. Four seizure related parameters (time of slowing, regularity, stereotypy and post-ictal suppression) are used as inputs to two classifiers, decision tree and fuzzy inference system (FIS). In addition to this, this chapter presents a new EEG based feature extraction method based on deep belief networks. Also, this chapter presents EEG rhythms features.

The second study proposes a complete system that consists of two stages: 1) an estimation of the seizure parameters from the raw EEG data, and 2) seizure quality prediction obtained by feeding the estimated seizure parameters into the constructed FIS.

5.2 Rating of Seizure Parameters and Quality

The onset of the seizure consists of two parts. The first part is the recruitment phase, where the amplitude gradually increases, and it is followed by the second part, which consists of chaotic polyspike activity, known as the 'polyspike phase' (high frequency, low amplitude) [214]. The next phase after the polyspike phase is referred to as the slow-wave phase, which represents a spike and wave activity. The slow wave phase of the ictal-EEG is a period of high amplitude polyspike and slow-wave activity. Following this, the termination phase starts when the slow-wave phase becomes irregular and is reduced in amplitude. In this phase, the amplitude and frequency of the seizure are reduced until it terminates and then the EEG trace will have a post-ictal phase [214].

This study focuses on the major seizure quality parameters that are identified in the literature. The first parameter is *time to onset* of slowing (TSLOW). TSLOW represents the time where the frequency (wave/sec) of the seizure starts slowing to ≤ 5 Hz [215, 216]. TSLOW is scored 0 if the frequency does not initially exceed 5 Hz [217]. The second parameter is the *regularity*, which is the score of the predominant morphologic pattern during the slow wave/spike and wave phase [216, 217]. Figure 5.1 describes the score (7 points) of slow wave phase regularity [218]. Another parameter that reflects the seizure quality rating is *stereotypy*, which is scored in the range 0 to 3 with half scores allowed [215]. Three variables are used to identify stereotypy scoring:

- High stereotypic: The progression from low amplitude (chaotic polyspike activity) to high amplitude (slow-wave/spike activity) is obvious without the reappearance of chaotic polyspike activity.
- Unclear spike wave morphology and the chaotic polyspike may appear after the onset of the slow-wave/spike and phase.
- The amplitude is changing during the slow-wave/spike and wave-phase (investigate the variability). A score of 0, 0.5 or 1 is assigned to each of the above mentioned three variables and their summation represents the stereotypy score.

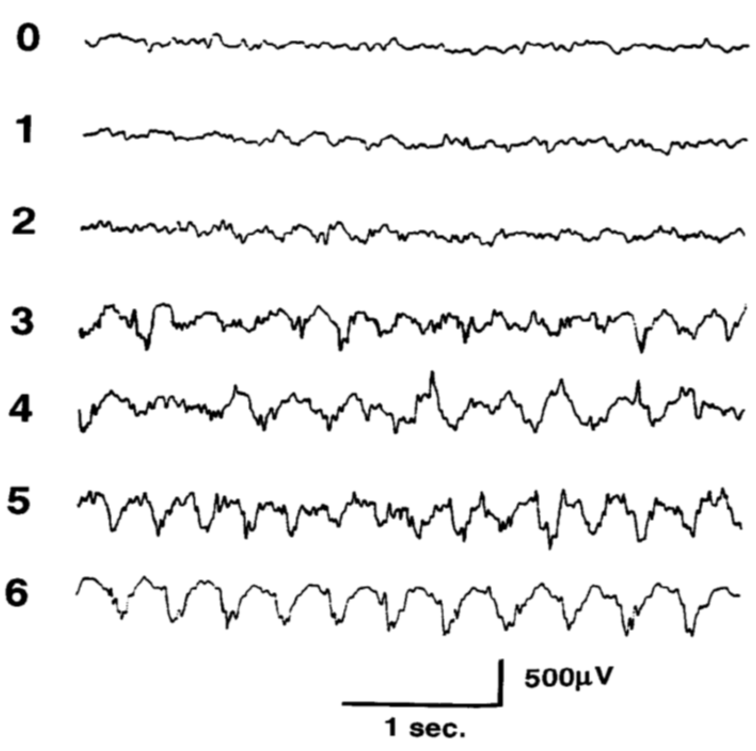


Figure 5.1: Regularity scoring.

Finally, the last parameter used to rate the seizure is *post-ictal suppression*. Post-ictal suppression is considered an essential parameter to predict therapeutic clinical outcomes when controlled by baseline Hamilton rating scale for depression and mode of stimulation [219]. The rating score of suppression ranges between 0 and 3 [220] with half-point scores allowed [215]. The time length for the decline in amplitude and

Table 5.1: Seizure quality score

Seizure quality rating	Score
Very poor	1
Poor	2
Average	3
Good	4
Very good	5

frequency from the last stereotypic spike/wave took into consideration. Table 5.1 lists the rating of global seizure quality based on five scores 1-5 [37].

5.3 Related Work

The detection and characterisation of seizures have been widely studied and several methods have been developed for the automatic detection of seizures and to predict their onset [93], which are mainly based on electroencephalography signal analysis [221]. However, most of these methods have not been developed for ECT-induced seizures, and there is hence a gap of knowledge in this area.

Recently, there has been an increased interest in utilising machine learning for the development of seizure prediction techniques. For example, a method is presented in [222] to predict seizures by classifying brain activity into four states: inter-ictal (baseline), pre-ictal (pre-seizure), ictal (during the seizure) and post-ictal (after the seizure). Inter-ictal and pre-ictal are the most significant states that can be used to predict seizures. Another seizure detection method consists of two major steps: the first one is the extraction of appropriate quantitative features from EEG data and the second step is classification of features into two classes (seizure and non-seizure) [223].

For ECT-induced seizures, a number of studies showed that the earlier mentioned seizure parameters are related to treatment efficacy; i.e., slow wave onset, regularity, stereotypy and post ictal suppression [89, 214, 215]. As a literature review revealed there is no automated method that attempts to automatically evaluate the quality of ECT-induced seizures, this study proposes a fuzzy rule-based system and decision tree for this purpose. The four seizure parameters were used as inputs for the two classification methods.

5.4 Estimation of ECT-Induced Seizure Quality

5.4.1 Participants

In this experiment, the EEG dataset recorded by Gálvez et al., which aimed at studying the impact of anaesthetic technique on EEG seizure quality [37], was used. Patients were at least 18 years old and had a clinical indication for acute ECT [37]. Patients with no medication information, especially PRN medication, 24 hours before the ECT session were excluded. The EEG data were collected at Wesley Hospital (Kogarah, Sydney, Australia) between April 2011 and April 2013 and was approved by the Human Research Ethics Committee of University of New South Wales. In total, 750 recordings from 84 patients were obtained.

5.4.2 Electroconvulsive Therapy Procedures

ECT treatment is safe and painless and its administrated during a short anaesthesia. ECT was delivered using a Mecta device (MECTA Spectrum 5000Q, maximum output 1152 mC Mecta Corp., Lake Oswego, OR). Anaesthesia induction was with Propofol ½ mg/kg.

EEG was recorded from two front-mastoid EEG channels; electrode-sites were cleaned adequately [37]. Expert trained clinicians scored the seizure parameters and rated seizure quality from the recorded EEG traces, using a previously described system of rating [217]. For each ECT session, the following treatment variables have been collected: absolute anaesthetic dose (propofol mg), ECT treatment number, time interval between sessions of ECT treatment (days), and initial seizure threshold (mC) [37].

5.5 Experiment 1: Classification of Seizure Parameters to Estimate Seizure Quality

This experiment proposes an approach that utilises the seizure parameters that were rated by clinicians to estimate an ECT-induced seizure scoring. These features are classified using the two methods of decision tree and fuzzy inference system, as shown in Figure 5.2. Decision tree is a machine learning method that has the ability to discover hidden rules in data. It produces reliable results when applied to different classification problems and has the capacity to handle missing data, which is important for the decision-making process in the medical field [224]. The fuzzy inference system is also fuzzy rule-based classifier and utilises fuzzy features, which makes it suitable for handling the uncertainties that commonly exist in medical applications. Moreover, unlike other black-box classification methods, the two classification methods are considered transparent models, as they allow interpretable mapping from the inputs to the output, which is considered an advantage for a number of medical applications. The approach presented in this study is expected to help clinicians to estimate the scoring of seizure quality as an aid in evaluating the effectiveness of ECT sessions.

5.5.1 Decision Tree Classification

A decision tree is constructed using the classification tree method. In order to reduce the tree complexity and enhance generalisation, (i) enforcing a limit on the number of splits (branches and leaves) and (ii) pruning the tree using the cost-complexity pruning method that pruned the tree based on optimal pruning scheme were both considered. A classification error rate was used to select the best splitting; i.e., limits for the number of branches at the various tree levels and fixing the number of leaves to 32. Tree-pruning aims to trim the initial branches of the tree. Pruning was used to improve the generalisation capability of the decision tree. An iterative approach is used in this process, which is terminated when the performance stops improving.

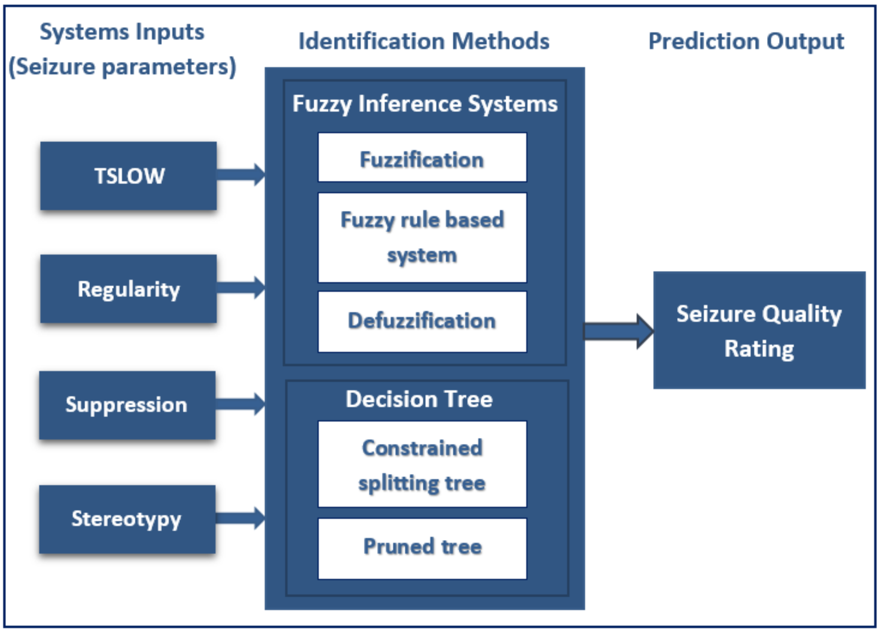


Figure 5.2: Predicting model.

5.5.2 Fuzzy Rule-Based System

The designed fuzzy rule-based system (Mamdani) contains four inputs and one output with 11 linguistic rules. The range of the fuzzy variables (inputs and output) are: TSLOW (1-10), regularity (1-6), suppression (1-3), stereotypy (1-3) and seizure quality (1-5). Three different membership functions were selected for each of the four fuzzy inputs (Low, Medium and High) and five membership functions for the fuzzy output, as shown in Figure 5.3. The membership functions of fuzzy output are selected based on scoring of seizure quality in Table 5.1:

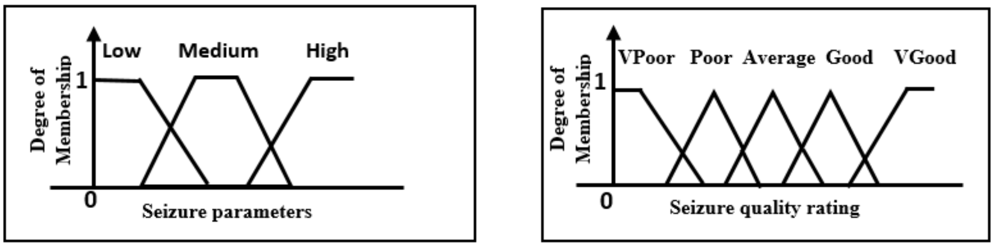
In general, the fuzzy rule-based system consists of sets of linguistic term rules in the form of "IF a set of conditions for the input variables are satisfied, THEN a consequence of the output variable is inferred." In this study, 11 rules are built based on the characteristics of the input variables and their effect on seizure quality rating, as outlined in the literature.

The proposed rules are:

- *Rule 1: If TSLOW is High and Stereotypy is Low, then seizure quality is VPoor (Very Poor)*
- *Rule 2: If Regularity is Low and Stereotypy is Medium, then seizure quality is VPoor (Very Poor)*
- *Rule 3: If Regularity is Medium and Stereotypy is Low and Suppression is Low, then seizure quality is Poor*
- *Rule 4: If Regularity is Medium and Suppression is Medium, then seizure quality is Average*
- *Rule 5: If Regularity is Low and Stereotypy is Medium and Suppression is Medium, then seizure quality is Average*
- *Rule 6: If Regularity is High and Stereotypy is Medium, then seizure quality is Good*
- *Rule 7: If Regularity is Medium and Stereotypy is Medium and Suppression is High, then seizure quality is Good*
- *Rule 8: If Regularity is High and Suppression is High, then seizure quality is VGood (Very Good)*
- *Rule 9: If TSLOW is Low and Regularity is High and Stereotypy is High, then seizure quality is VGood (Very Good)*
- *Rule 10: If TSLOW is High and Regularity is High and Suppression is Medium, then seizure quality is Good*
- *Rule 11: If Regularity is Medium and Stereotypy is High, then seizure quality is VGood (Very Good)*

5.5.3 Experimental Results and Discussion

In order to evaluate the performance of the decision tree and FIS, a cross-validation approach was used to split the data into training and testing. In the training phase, the



(a) (b)

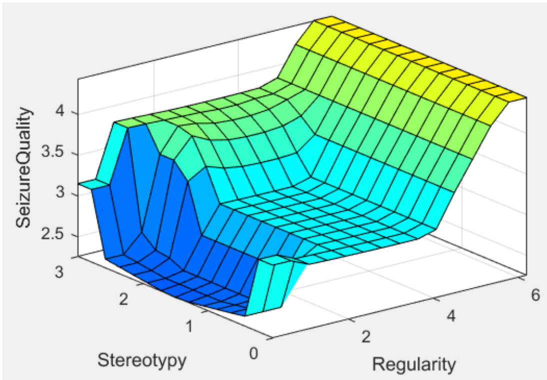
Figure 5.3: Membership functions: (a) fuzzy inputs, (b) fuzzy output (seizure quality).

Table 5.2: Confusion matrix

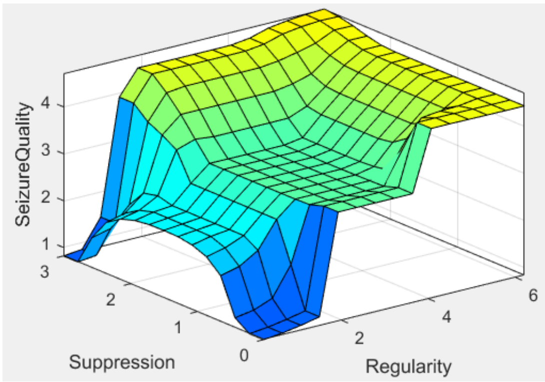
Fuzzy rule-based system					Decision tree				
3	6	4	0	0	10	3	0	0	0
0	1	70	8	0	2	43	33	1	0
0	2	133	51	0	0	19	133	34	0
0	4	23	220	16	0	1	29	183	50
0	0	0	84	124	0	0	0	57	151

Decision tree (Splitting)					Decision tree (Pruning)				
10	3	0	0	0	10	3	0	0	0
2	38	39	0	0	1	51	25	2	0
0	14	146	26	0	0	10	138	38	0
0	0	34	188	41	0	0	15	208	40
0	0	0	64	144	0	0	0	48	160

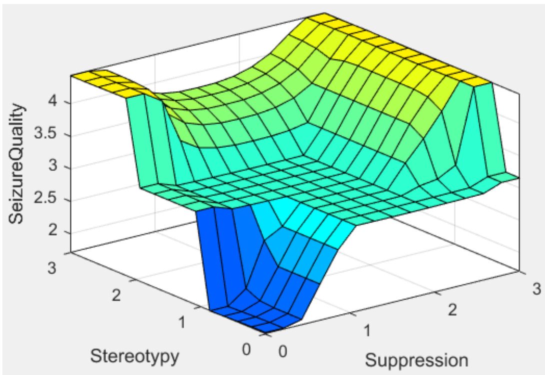
decision tree and identification the best parameters of FIS were constructed. When provided with the seizure parameters of the 750 recordings, FIS produced 481 matches (similar scores to those recorded by the clinician experts) and 252 that differed by one point only. The scores of the remaining 17 recordings differed by two points. As the FIS provides real numbers, the error rate was calculated between the FIS output and the true scores. Results show the error rate for this case = 0.36. Figure 5.4 shows the fuzzy surface between pairs of seizure parameters and the seizure quality. The results show that there is a monotonic relationship between the seizure parameters (regularity, stereotypy and suppression) and seizure quality, which indicate that these parameters are indeed good measures for the seizure quality.



(a)



(b)



(c)

Figure 5.4: Seizure quality based on: (a) Stereotypy-Regularity (b) Suppression-Regularity, (c) Suppression-Stereotypy.

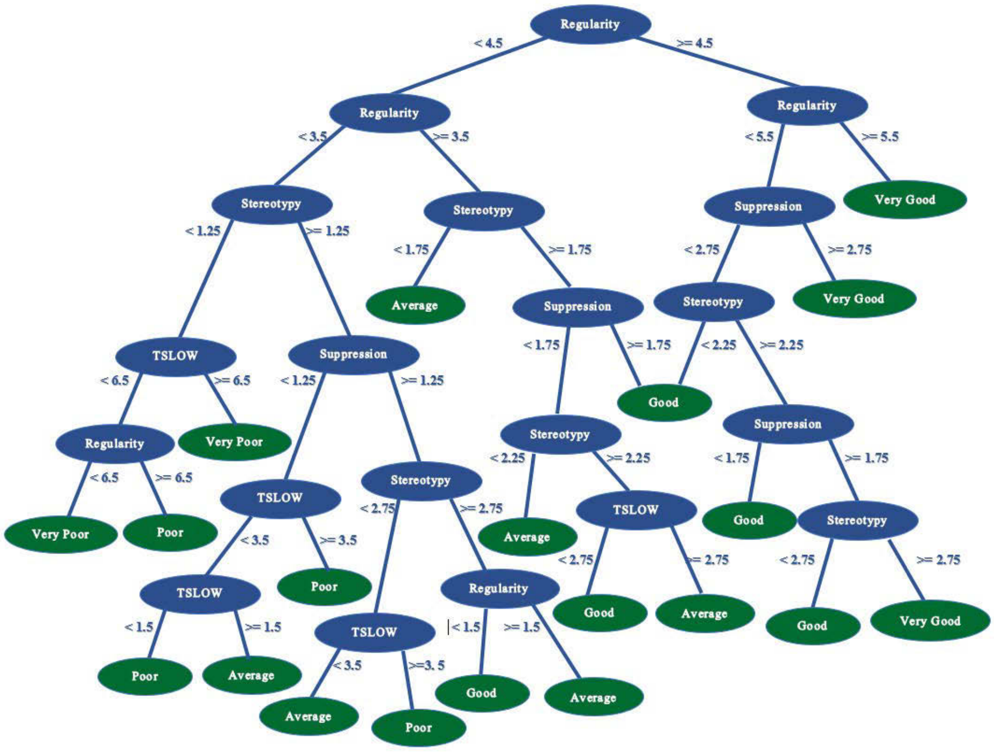


Figure 5.5: Tree pruning.

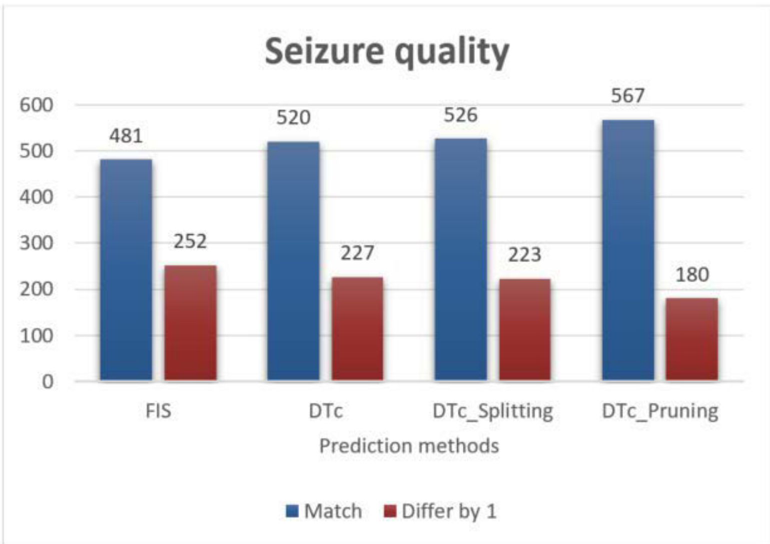


Figure 5.6: The prediction rate of seizure quality.

In the second case, the seizure ratings was predicted using the fully constructed decision tree without enforcing any conditions on splitting and without applying pruning. The tree produced 105 rules and the classification results of this tree are 520 matches, 227 that differed by one score only and the remaining three that differed by two scores. The classification error rate is found to be 0.31. To reduce the size of the tree and improve its classification accuracy, splitting and pruning techniques that resulted in noticeably simpler trees were used, as shown in Figure 5.5. The error rates of the constrained splitting tree and pruned tree are found to be 0.30 and 0.25 respectively. Figure 5.6 shows a comparison between the two prediction techniques: fuzzy inference system and decision tree (three cases; initial constrained, splitting and pruning). Table 5.2 shows the confusion matrices of FIS and three cases of decision trees. The diagonal of each matrix represents the number of correctly identified seizure quality rating, and average class-wise accuracy for two classification methods are: FIS=0.48 and decision tree = 0.69. The class-wise accuracy for FIS was as following: $C1 = 3/13 = 0.23, C2 = 1/79 = 0.01, C3 = 133/186 = 0.75, C4 = 220/263 = 0.84, C5 = 124/208 = 0.60$). Hence, the relatively higher error of FIS was influenced by the bad performance of the first two classes. particularly the second one. The overall classification accuracy for two classification methods were: FIS = 0.64 and the overall accuracy for decision tree was initial tree = 0.69, splitting = 0.70 and pruning = 0.76. Despite the fact that decision trees produced better results than the FIS, the performance of FIS could be improved through the optimisation of the fuzzy rules and membership parameters. More specifically, rules could be modified or weighted to enhance the accuracy of the first two classes.

5.6 Experiment 2: Identification of Seizure Quality Rating Based on EEG Data

5.6.1 Pre-processing EEG Signal

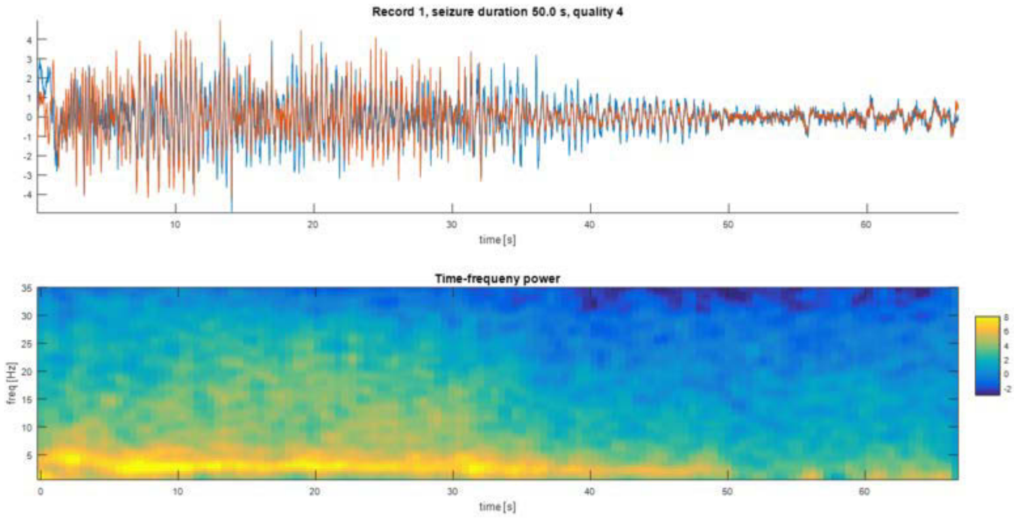
EEG data were recorded from two front-mastoid EEG channels; electrode-sites were cleaned adequately [37]. This study has EEG data of just 52 recordings. Each channel of EEG data was processed separately by removing high-frequency noise, which was achieved by down sampling to 140 Hz and applying a band-pass filter between 0.5 Hz and 35 Hz using a second order digital Butterworth filter. The distribution over the time and frequency (T-F) space was considered. In general, T-F distribution is an important tool for studying and analysing EEG signals and extracting useful features to improve the performance of seizure prediction. Figure 5.7 show the T-F distribution of two EEG recordings: Figure 5.7 (a) shows a recording of good seizure quality and Figure 5.7 (b) shows a recording of bad seizure quality.

5.6.2 Feature Extraction Based on Deep Belief Network and EEG Bands

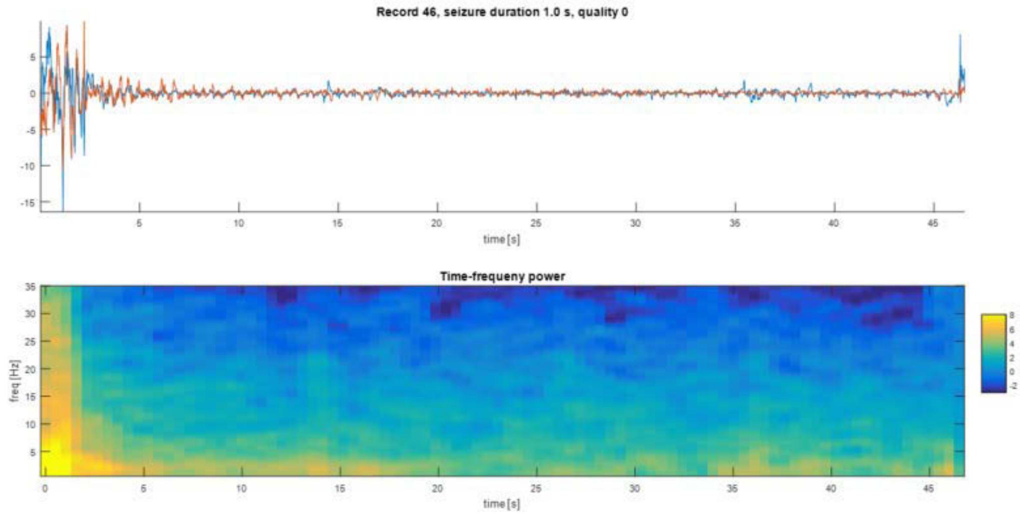
The DBN has been discussed in chapter 3 as a classification method. However, this part of the study discusses using of DBN as a feature extraction method and reducing dimension of EEG data. Unsupervised feature learning is still a challenging problem. In addition to this, another set of features was extracted based on EEG rhythms.

Each channel of the EEG signals was resized to a square matrix that can facilitate the utilising of the DBN feature extraction tools. DBN feature extraction technique was applied on each channel separately. Different subsets of features were extracted (8, 12, 24, 64). In order to evaluate these sets of features, a regression tree method was applied to map these features to seizure quality scores. The best set of features was found to consist of 12 features per channel (24-features for the two EEG channels) based on the performance of regression tree method. The results produced 16 matches, 24 recordings that differ by one point only, while the remaining 12 recordings differ by two points. Figure 5.8 represents the obtained scores using the DBN features and the coresponding clinician scores.

In addition to this, another set of features was extracted based on EEG rhythms. Power spectral density (PSD) was estimated using Welch method with a Hanning window. Five features were extracted from the EEG data based on PSD, which are Delta $(0.5 \leq \delta < 4)$Hz, Theta $(4 \leq \theta < 8)$Hz, Alpha $(8 \leq \alpha < 13)$Hz, Beta $(13 \leq \beta < 30)$Hz,



(a)



(b)

Figure 5.7: T-F distribution: (a) recording 1, (b) recording 46.

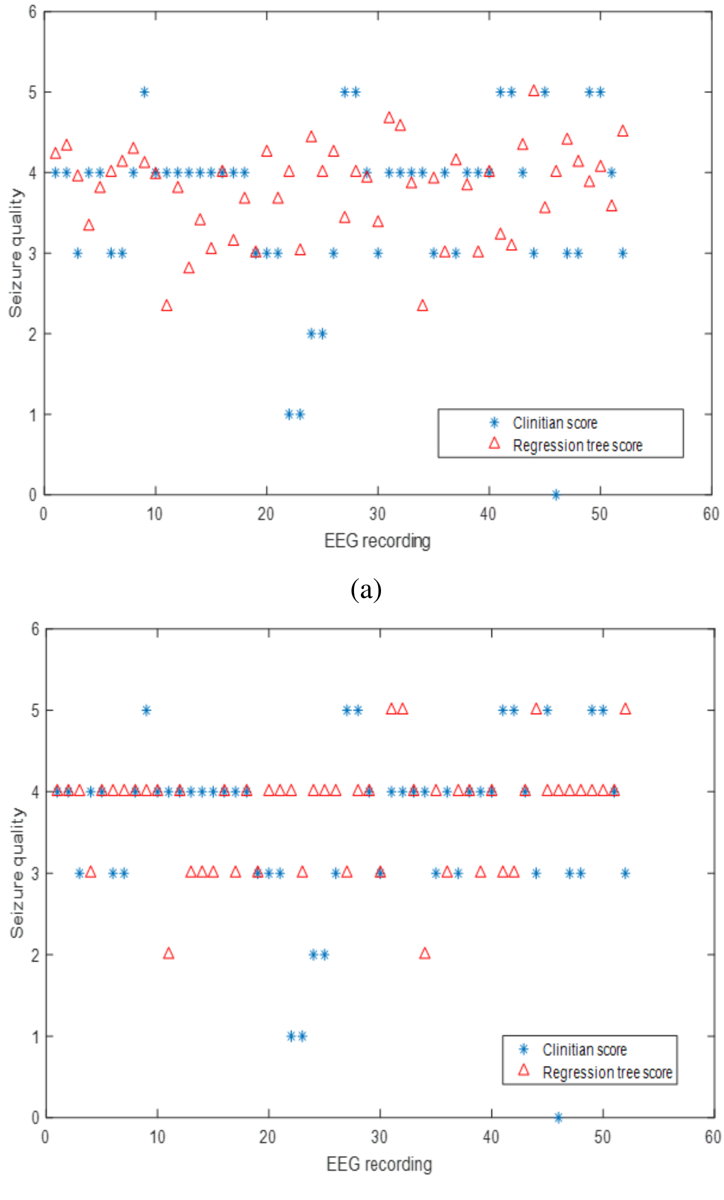
and Gamma $(30 \leq \gamma < 100)$Hz. The obtained seizure quality scores of these features based on regression tree is shown in Figure 5.9 with the corresponding clinician scores. Comparison between the two sets of features reveals that the 12-features of DBN outperform their EEG bands counterpart.

However, due to the limited number of training samples (52 EEG recordings), the scoring of seizure quality based on those two feature extraction methods and regression trees were found to be highly influenced by the dominated classes (3 and 4). These results are induced to propose a new technique that attempts to estimate the seizure parameters from the raw EEG data automatically.

5.7 Construction of Seizure Quality Estimation System Based on EEG Data

In Section 5.5, the seizure parameters that were manually scored by clinicians are used to estimate the seizure quality. In order to reduce the involvement of clinicians, a two-stage system to estimate seizure quality from the pre-processed EEG signals is proposed, as shown in Figure 5.10.

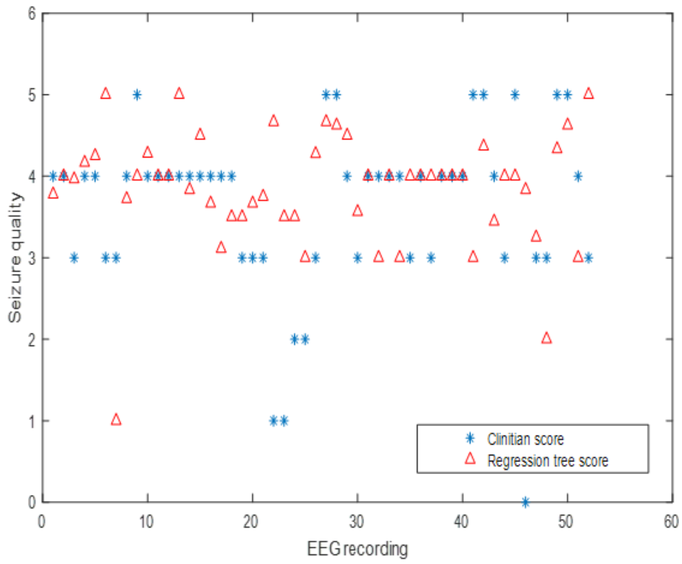
This study proposed a new rule-based system to estimate the seizure parameters based on clinician criteria. As described in Section 5.2, slow-wave/spike activities are found to play big role. Hence, the decision was made to identify peaks for the whole recording and analyse the height and prominence of peaks as well as the distances between them. Accordingly, peaks and their locations are tested for each EEG channel, by analysing the histogram to select the best peaks that can be used to estimate seizure parameters. From Figure 5.11, the best peaks are found around 0 and 1, then 1 and 0.1 were selected to evaluate peaks. The Matlab function MinPeakProminence was used to find all peaks that have a prominence of at least 1, and MinPeakHeight was used to find all peaks that are greater than 1 as shown in Figure 5.12. The same procedure is followed for all peaks at 0.1. The identified peaks are used to estimate the seizure



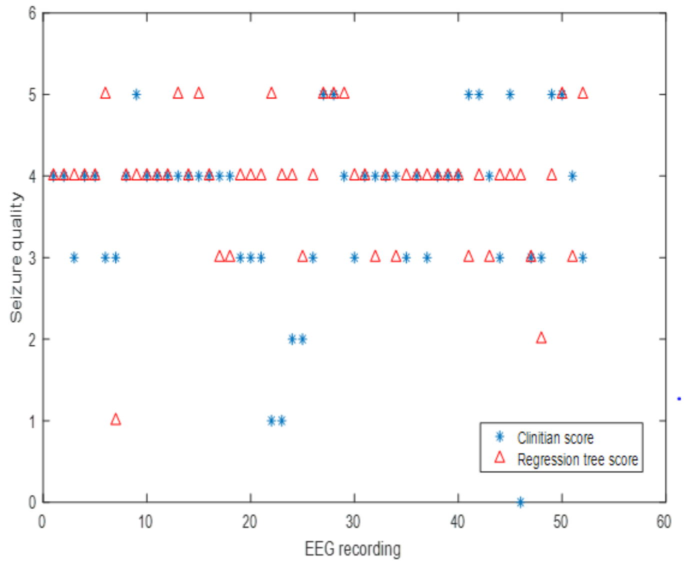
(a)

(b)

Figure 5.8: Seizure quality score based on DBN features and clinician recordings: (a) before round, (b) after round.



(a)



(b)

Figure 5.9: Seizure quality score based on EEG rhythm and clinician recordings: (a) before rounding to integers, (b) after rounding to integers.

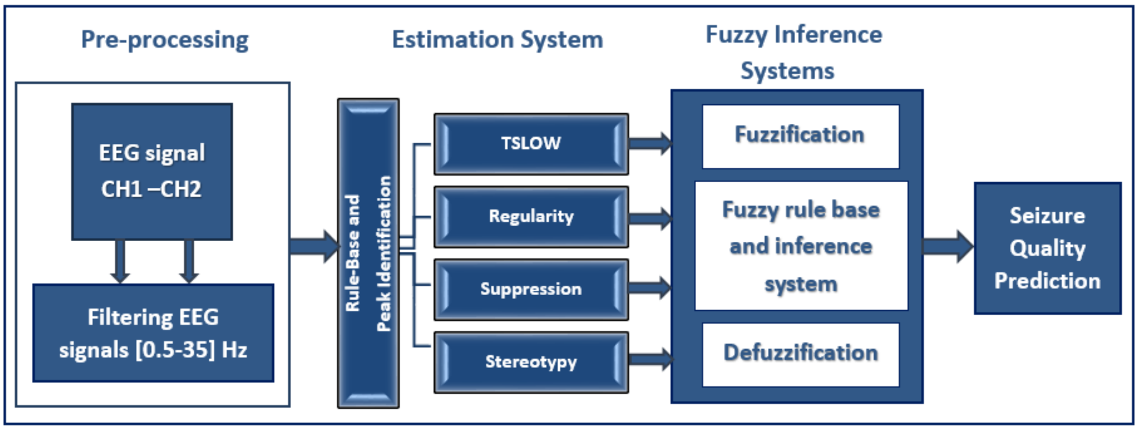


Figure 5.10: Prediction seizure quality rating model.

parameters described earlier; i.e., TSLOW, regularity, stereotypy and suppression, as will be described in Section 5.7.1.

5.7.1 The Proposed EEG-Based Seizure Quality Estimation Method

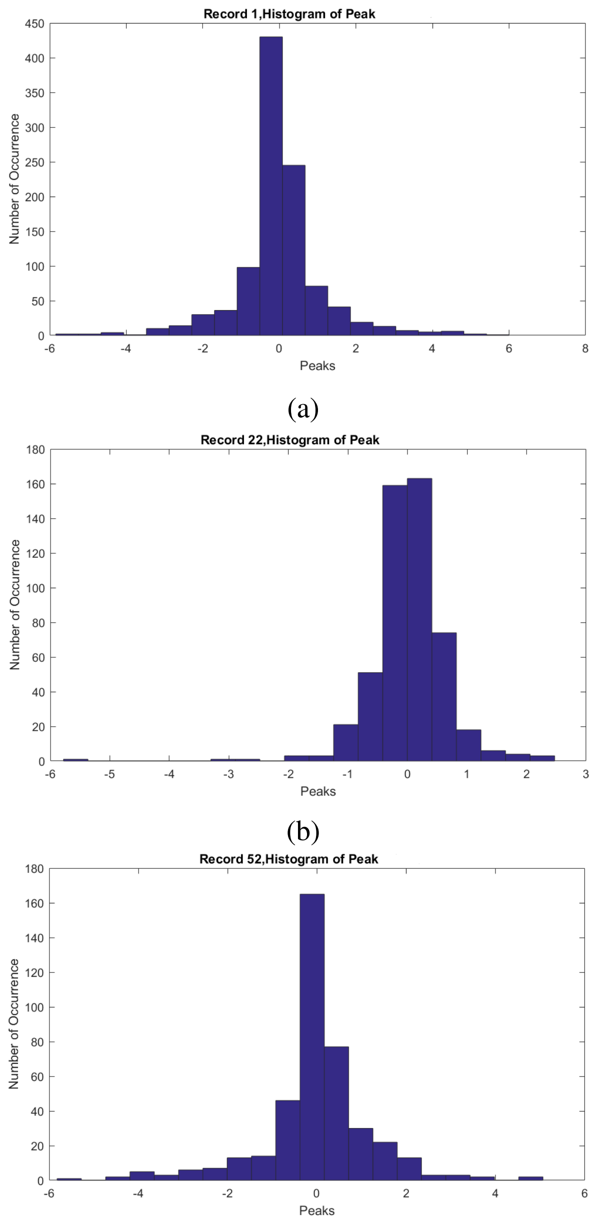
The proposed model is composed of three stages. The first stage is pre-processing EEG data, as described in Section 5.6.1. The second stage is the estimation of seizure parameters and the third stage is the fuzzy inference system. Table 5.3 shows the steps of the proposed technique.

The following procedure describes the estimation of seizure parameters for each EEG channel:

1. Find the peaks and locations for all recordings based on the following criteria:

   – Find peaks that are greater than *MinPeakHeight* MPH (locate the peaks that are higher than their neighbouring samples by at least 1 and 0.1).

   – Find peaks that have a vertical drop of more than *MinPeakProminence* MPP from the peak on both sides without encountering either the end of the signal (locate the peaks that have a prominence of at least 1 and 0.1. 1 and 0.1 were found to be good peak threshold values).

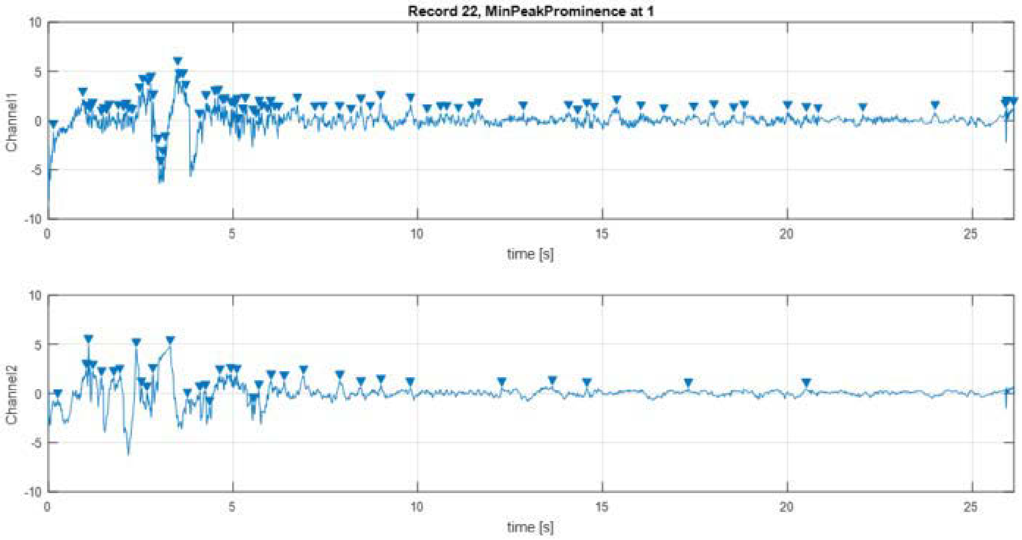


(a)

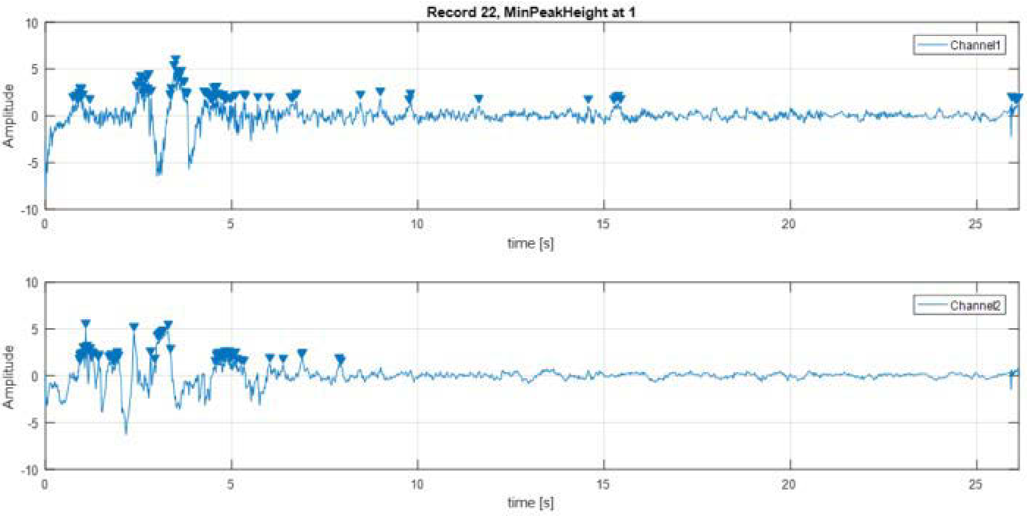
(b)

(c)

Figure 5.11: Histogram of peak amplitude: (a) EEG recording 1, (b) EEG recording 22, (c) EEG recording 52.



(a)



(b)

Figure 5.12: Peak identification for recording 22: (a) prominence peaks at 1, (b) height peaks at least 1.

Table 5.3: Algorithm of the proposed EEG-based seizure quality estimation

First stage: Pre-processing EEG signal (as described in Section 5.6.1).
Second stage: Estimation of seizure parameters using peak identification, details will be provided below
Input: EEG raw data (time series) (52 recordings (two EEG channels))
Output: Seizure indices rating
begin
Find: Peaks and locations (Test each peak and location of each sample of EEG recordings to estimate seizure parameters using the following two functions: *MinProminence at 1 Sec and 0.1 Sec*, *MinPeakHeight at 1 Sec and 0.1Sec*)
Seizure Indices
- TSLOW [2-10]
- Regularity [1-6]
- Stereotypy [0-3]
- Post-ictal Suppression [0-3]
Third stage: Fuzzy inference system (fuzzy rule-based system), details will be provided below
Fuzzy Input: TSLOW, regularity, stereotypy and suppression
Fuzzy Output: Predicting seizure quality rating [1-5]
Inference mechanism: Apply the fuzzy rule-based system to the input parameters of each recording

2. Estimate seizure variables (TSLOW, regularity, stereotypy and post-ictal suppression) based on the following steps:

   – **Time to onset of slowing (TSLOW) [2-10]**

     Each recording was divided into 10 segments, then *MinPeakProminence* was used to compute prominence of peaks at 1. To compute the score of TSLOW, two neighbouring peaks were tested to identify the time at which the signal starts slowing <=5 Hz.

   – **Regularity [1-6]**

     The location and peak values of signals are computed. *MinPeakHeight* at 1 second is used to find the peaks and locations at the minimum height of 1. The number of high peaks represents the predominant morphologic pattern, which in turn is used to estimate the regularity rating.

   – **Stereotypy [0-3]**

As mentioned in Section 2.4, the rating score of stereotypy is estimated through the summation of three variables (progression of signal, spike and wave morphology and the variability in amplitude).

  * To check the progression of signal from poly-spike phase to the slow wave phase, ranges between [0.1-0.2], [0.2-1] and [1-2] were chosen. The possible combinations of two peaks were examined then the mean value of all sets of peaks was computed. The mean value is linearly mapped and a threshold is applied to reflect the degree of progression (no progression=0; clear progression =1, while some progression = 0.5).
  * Check the transition of signal by testing each two successive peaks (No=0; some=0.5; Yes=1).
  * Test the variability in amplitude by computing the difference between two neighbouring peaks and evaluate the value, if it's near to 0, 0.5 or 1 (marked variability=0; some=0.5; No=1).

- **Post-ictal suppression [0-3]**

  The suppression score is estimated through the summation of three variables (seizure end point, transition of signal and degree of signal flatting surface).

  * Compute the peaks that have an amplitude at least 0.1, and evaluate the score to specify the end point of seizure (no end point=0; 0.5 unclear end point; end point=1).
  * Investigate the transition of signals by testing the difference between two peaks. Three ranges based on trial and error approach were selected to investigate the transition (<0.05, <0.12, <0.6) and averaged of all peaks to evaluate the transition based on this score (No=0; some=0.5; Yes=1).
  * Estimate the flatness of the surface by testing the termination of waves (the transition between samples is gradually up or down) (No=0; some=0.5; Yes=1 (very flat)).

The fuzzy inference system is the third stage of the proposed model and the design of the FIS is the same as that described in first experiment. However, in this study the range provided by the clinician for the time to onset of slowing is somewhat different (TSLOW = (2-10)). Accordingly, the fuzzy rules that are built based on the characteristics of the input variables and their effect on seizure quality rating, as outlined in the literature, slightly different from those described earlier. The first eight rules are the same as their corresponding ones in the first FIS, while rules 9,10 and 11 of the first FIS are replaced with two slightly different ones. More specifically, the ten rules are:

- *Rule 1: If TSLOW is High and Stereotypy is Low, then seizure quality is VPoor (Very Poor)*
- *Rule 2: If Regularity is Low and Stereotypy is Medium, then seizure quality is VPoor (Very Poor)*
- *Rule 3: If Regularity is Medium and Stereotypy is Low and Suppression is Low, then seizure quality is Poor*
- *Rule 4: If Regularity is Medium and Suppression is Medium, then seizure quality is Average*
- *Rule 5: If Regularity is Low and Stereotypy is Medium and Suppression is Medium, then seizure quality is Average*
- *Rule 6: If Regularity is High and Stereotypy is Medium, then seizure quality is Good)*
- *Rule 7: If Regularity is Medium and Stereotypy is Medium and Suppression is High, then seizure quality is Good*
- *Rule 8: If Regularity is High and Suppression is High, then seizure quality is VGood (Very Good)*
- *Rule 9: If Regularity is High and Stereotypy is High, then seizure quality is VGood (Very Good)*
- *Rule 10: If Regularity is High and Suppression is Medium, then seizure quality is Good*

5.7.2 Results and Discussion

The 52 EEG recordings were analysed using the proposed method. The proposed FIS considers two cases. In the first case, the constructed FIS fed with the seizure parameters that were rated by the clinicians. In the second case, the proposed EEG-based seizure parameters estimation method was applied, and the outcome was fed to the constructed FIS.

For the first case, the predicted seizure quality values that were produced by the constructed FIS (after rounding) indicated that 38 recordings match the true quality scores, 13 recordings differ by one score only, while the remaining recording differ by two scores. Figure 5.13 represents the accuracy of predicting seizure quality rating. As the FIS provides real numbers, the error rate between the FIS output and the true scores was calculated and the error rate for this case was = 0.269. Figure 5.14a shows a comparison between the benchmark seizure quality values that were scored by the clinical team and the predicted scoring using the proposed fuzzy rule-based system.

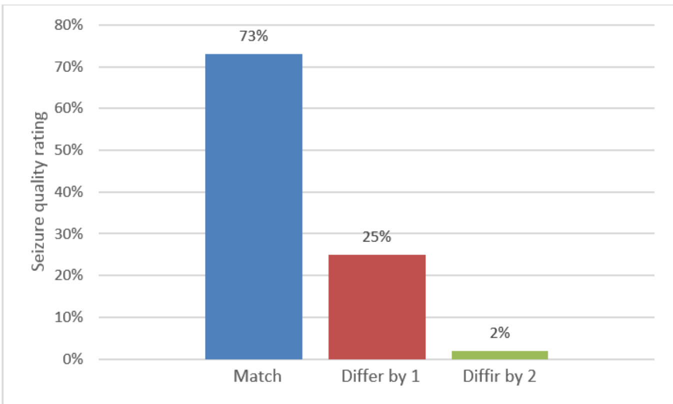
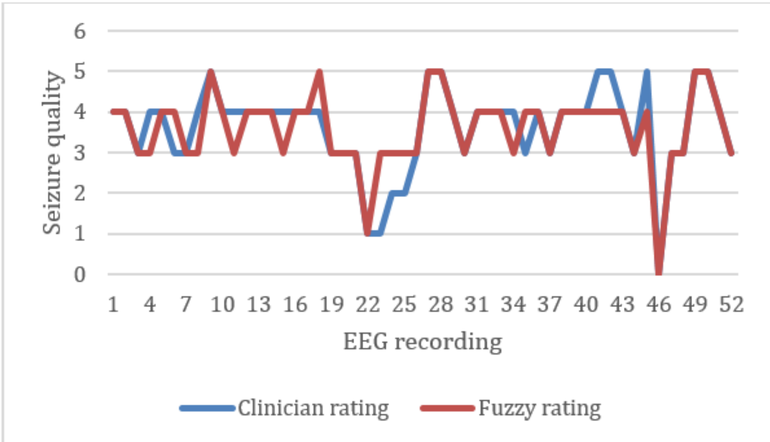
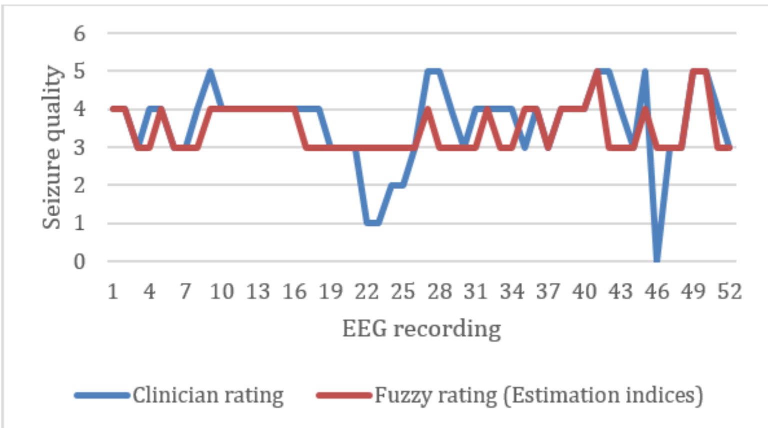


Figure 5.13: Prediction accuracy of seizure quality rating.

In the second case, the seizure parameters from the EEG data based on the algorithm described in Section 5.6.2 (stage two) are estimated, and then the predicted parameter values are fed into the FIS. The obtained scores of the seizure quality are shown in Figure 5.14b. The error rate was = 0.4038 with 31 recordings matching the true scores, 16 recordings differing from the true scores by one, and the remaining five differing by two. The obtained error rate was found to be slightly higher than that of the 52



(a)



(b)

Figure 5.14: Prediction scores of global seizure quality rating based on: (a) clinical seizure rating, (b) estimation seizure indices.

recordings, but results are still acceptable. Figure 5.15 shows the fuzzy surface between pairs of seizure parameters and the seizure quality.

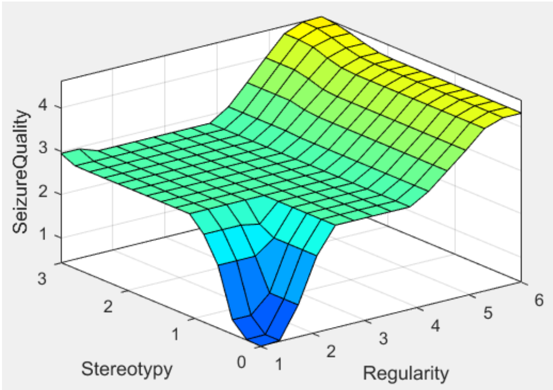
The results show that there is a monotonic relationship between the seizure parameters (regularity, stereotypy and suppression) and seizure quality, which indicates that these estimated parameters are indeed good measures for the seizure quality.

The interpretable mapping from the inputs to outputs is a significant part in many medical applications (diagnosing and identifying). By comparing the scores that obtained using proposed rule-based systems (proposed model and FIS) and clinician scores, the proposed model can be use to score seizure quality rating. Accordingly, this study represents a tool that could help clinicians to estimate the scoring of seizure quality and reduce time and effort spent on evaluating the effectiveness of ECT sessions. It should be mentioned that, this study represents the initial investigation and identification of seizure parameters and seizure quality ratings. The 750 EEG data still under cleaning and scoring with clinician team. In future studies, the results of the proposed model will be improved by increasing the number of EEG recordings that are collected through ECT treatment.

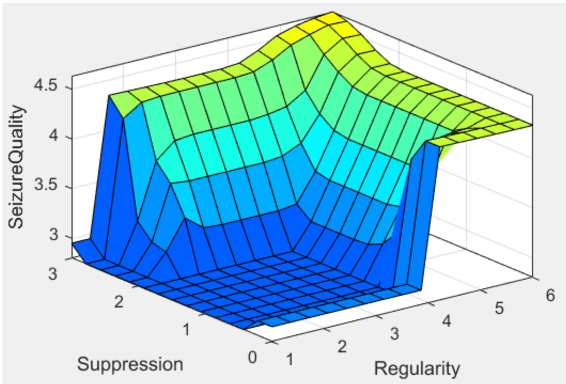
5.8 Summary

In this chapter, two experiments were presented. The first one estimated the seizure quality approach based on two classification methods: fuzzy rule-based system and decision tree. The two classifiers were each fed with four seizure parameters that were rated by a team of clinician. 750 EEG recordings were used in the first experiment. The two classifiers produced encouraging results with error rates of 0.311 and 0.249 for FIS and decision tree respectively. The classification results show that the four seizure parameters provide relevant information about the rating of seizure quality.

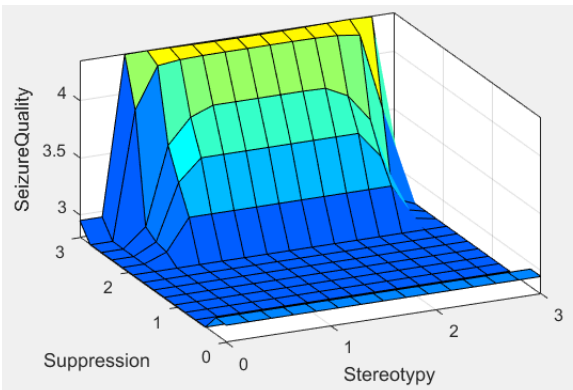
In addition to this, two sets of features were extracted based on DBN and EEG bands techniques. A regression tree method was used to examine these features.



(a)



(b)



(c)

Figure 5.15: Seizure quality based on (a) Stereotypy-Regularity (b) Suppression-Regularity, (c) Suppression-Stereotypy.

In the second experiment, 52 ECT-induced EEG recordings were used. A comparison between the obtained seizure quality scores was presented, which were produced by the fuzzy rule-based system using: the seizure parameter values that were rated by the clinicians, and the seizure parameter values that were estimated using the proposed peak analysis approach. The results of the first approach were slightly better than those for the second approach. This indicates that the second method that attempts to estimate the seizure quality from the raw EEG data has the capability of producing acceptable seizure quality scores.

Chapter 6

Conclusions and Future Directions

This chapter provides a summary of the dissertation. Major results and conclusions of the research are presented and recommendations for future research that may be related to this dissertation fields are outlined.

6.1 Summary

Depression is a major mental illness that affects people around the world. One of the most types of depression is the major depressive disorder. Transcranial direct current stimulation (tDCS) has recently been explored as a promising technique for treating depression and in particular major depressive disorder.

However, one of the issues with tDCS is that a series of tDCS sessions run over a number of weeks may be required before it is possible to determine how patients respond to the treatment. The modulatory effect of transcranial direct current stimulation (tDCS) on changes in cortical activities can be studied by analysing electroencephalogram (EEG) signal. EEG signals contain of a very large amount of relevant information about brain disorders and different types of artefacts, which are conducted through placing a number of electrodes on the scalp of the patient.

EEG plays a vital role in understanding the functional state of mental brain disease, particularly in regard to diagnosing disorders. An EEG is the neurophysiological measurement of the electrical activity generated by billions of neurons in the brain, which is recorded by multi-channel electrodes placed on the scalp [225]. EEG signals are nonstationary signals and can be contaminated by different artefacts; therefore, the analysis and classification of EEG signals are challenging. Due to the complex nature of EEG signals, special techniques for analysis, including machine learning ones that have the ability to deal with multi-channel EEG signals are needed.

In this research, three standard machine learning techniques were proposed to analyse and predict the clinical outcomes of depressive disorder based on the brain activities that were recorded by an EEG before and after tDCS treatment. In particular, deep belief networks (DBNs), which is a machine learning technique based on a restricted Boltzmann machine (RBM), was proposed for identifying and predicting the clinical outcomes of depression patients. In this research, the modifications in DBN architecture were proposed to deal with multi-channel EEG data. A DBN is a probabilistic, generative model that is constructed of multiple hidden layers. In a DBN, the initial values of network weights are selected by a greedy layer-by-layer pre-training algorithm. The undirected layers in the DBNs are formed using RBMs, and hence, RBMs can be considered as the building blocks of DBNs. RBMs have emerged in machine learning and are used as a generative model for many different types of data.

Electroconvulsive therapy (ECT), which is the second line treatment of depression disease, was also considered in this research. Two identification studies were presented in this part of research. First, the ECT seizure induced quality ratings of 750 recordings were predicted using a decision tree and fuzzy based-rule system. Those two classification systems were both fed with seizure parameters that were rated by expert clinicians. Then, in order to evaluate the relevance of the seizure parameters, two types of features are evaluated using a regression tree technique; one produced based on DBN, while the other is based on the power spectral densities of five standard EEG rhythms. A new estimation technique to identify seizure parameters was also proposed. An ECT seizure quality rating score was predicted based on the estimated values of seizure parameters.

6.2 Conclusions

The aim of this research was to modify and utilise machine learning in the treatment process of both tDCS and ECT. Particularly, this study has achieved three specific aims. First, this study has measured the brain activity of depressive disorder patients based on EEG and has proposed a special type of machine learning to predict the clinical outcomes of patients after tDCS treatment. Three classification tasks based on machine learning and intelligent systems techniques have been presented in this study. The first one involves distinguishing between three sessions of tDCS treatment (baseline EEG, active tDCS and sham tDCS). The second aim of this study is investigating the feasibility of identifying major depressive disorder patients that respond to the tDCS treatment based on resting-state EEG that is recorded prior to commencement of the treatment. Third, this study attempted to identify the quality scores of ECT-induced seizures. This study presented five key contributions (as described in Chapter 1) based on these aims. The findings in relation to each of these contributions are presented below.

- Three DBN architectures were presented that fuse the channel information at different levels. These architectures are single-stream DBN, an individual DBN for each EEG channel and multi-stream or multi-channel DBNs. The first implementation (single-stream DBN) is the traditional DBN that is fed with a concatenated feature vector of the multiple channels. The second implementation uses a separate DBN for each channel and combines the classification results using another DBN. The multi-stream DBN attempts to extract local attributes from the individual channels in the lower network layers, which are individually processed and then combined in the higher unified layer(s) to extract high level information that facilitates more accurate labelling of the data.

The EEG data used in this study consists of 62 channels was collected from MDD patients who were undergoing tDCS treatment. The data of each of the three sessions (baseline EEG, active tDCS and sham tDCS) was represented using the power spectral density of the five frequency bands. The most relevant channels for classification were identified. Each channel with its four neighbours were evaluated. The obtained results showed the frontal midline channels to be more influential.

In order to evaluate the performance of the proposed three DBNs, three standard machine learning techniques; namely support vector machine (SVM), linear discriminant analysis (LDA) and an extreme learning machine (ELM) were used. The obtained results indicated that the multi-stream DBN outperformed the other two DBN implementations as well as the SVM, LDA and ELM classifiers, particularly when processing multiple channels. These findings demonstrate the capacity of DBNs as a classification platform for multi-channel EEG data. Accordingly, the utilisation of deep learning in the analysis and classification of EEG data seems well worth investigating.

- An automated EEG analysis method that assists mental health professionals to predict the clinical outcomes of depressed participants undergoing tDCS treatment was proposed. This study considered two cases: initial prediction based on tDCS sessions and advanced prediction based on a baseline EEG. In order to study the effect of tDCS treatment on EEG, i.e., the affected regions of the brain and changes in EEG rhythms were identified. These two studies were achieved through automated EEG signals of 10 patients using three standard machine learning techniques: SVM, LDA and ELM. In order to evaluate the behaviour of patients, the labels were specified based on improvements in mood and cognition following tDCS treatment. The performance of different brain regions for each of the two modalities was evaluated based on the classification error rates of all subsets of channel pair for each region. Specifically, the two investigations are as follows:

– In the initial investigation, power spectral density features were extracted from the EEG data acquired after a session of active tDCS and sham tDCS and fed to the three classifiers (SVM, LDA and ELM). The results obtained demonstrated that the frontal brain region was more influential in predicting changes in depression levels. In addition to this, the classification results observed the cognitive improvement (9/10 participants) and for mood improvement (7/10 participants).

  Even though the data was collected from a limited number of participants, the promising results that were obtained suggest that this research direction needs to be further investigated, as it could make valuable contributions to the treatment of major depressive disorder. Based on these results, a more in-depth investigation (advanced investigation) was proposed.

– Advanced investigation: in this research the feasibility of predicting the outcome of tDCS treatment was investigated based on analysing the features of a resting-state EEG recorded at baseline. The results showed that the frontal channels performed relatively well for predicting treatment outcome based on the mood score, while the parietal-occipital channels were more influential for the cognition score. This study assessed mood and cognitive scores based on classification accuracy. The obtained results showed that the mood labels were predicted accurately in eight out of 10 participants and the cognitive labels were predicted accurately in 10 out of 10 participants. Moreover, the frontal brain regions performed well for mood labels, while, the parietal regions recorded high performance for cognitive labels.

  It should be mentioned that, in this type of highly complex EEG classification task based on subject independent approach, the training sample far more participants (not less than 50).

  In addition to these, alpha asymmetry was examined in three electrode pairs (frontal, central and parietal pairs). The results indicated that the asymmetry index was negative over the frontal and central brain region, but was positive over the parietal region. These results suggest greater right-lateralised frontal activity. Despite the limited number of participants, the encouraging results that were obtained using the proposed method can serve as a proof-of-concept that would justify recording the baseline EEG in a larger cohort of patients with major depressive disorder undergoing tDCS treatment.

These initial results could have an important influence on the adoption of tDCS treatment for depression, by enabling the early identification of participants who will respond to tDCS treatment, thus avoiding treatment delays and saving staff time and resources.

- An automated approach to predict the quality score of electroconvulsive therapy induced seizure. Two studies were presented. First, two rule-based methods were proposed to predict a seizure quality rating based on clinician's recordings of seizure parameters. Second, a method to estimate seizure parameters was proposed and incorporated with a rule-based method to predict seizure quality scores from raw EEG data. The two studies are as follows:

  – In the first study, a fuzzy rule-based inference system and a decision tree classifier were proposed to identify the rating quality of ECT-induced seizures. The rules of the fuzzy inference system were derived from the literature through identifying relationships between the input parameters and seizure quality. A decision tree was chosen to perform this classification task as it provides transparent mapping from inputs to outputs. 750 EEG recordings were used in this part of the study. Seizure parameters (TSLOW, regularity, stereotypy and post-ictal suppression) rated by a team of expert clinicians were used as inputs to the two methods.

    The two classification methods (fuzzy inference system and decision tree) are considered transparent models. This is important for a number of medical systems, as such models establish interpretable mapping from the inputs to outputs, which is not the cause with other black-box classification methods. The obtained results are quite encouraging in terms of achieving scores that are not very different from the ones produced by expert clinicians. The proposed approach indicates that the scoring of ECT induced seizures can be automated, which will be beneficial to clinicians working in this field and will provide a useful tool for the treatment of major depressive disorder using ECT.

- The second study consists of two parts which are as follows:
  - First part: two subsets of features were extracted from EEG signals recorded during ECT sessions. 52 recordings of two front-mastoid EEG channels were used in this part of the study. The signal was down sampled to 140 Hz and a band-pass filter was applied on each EEG channel to remove unwanted frequencies that mostly consist of noise. The two subsets of features were extracted based on a DBN method and the standard EEG bands. The obtained features were compared with the seizure parameters that were rated by clinicians using a regression tree method. Due to the limited number of recordings, the obtained results were found to be influenced by the dominant classes (3 and 4).
  - Second part: based on the obtained results of the two feature extraction methods described earlier, a new method was proposed to estimate the seizure parameters from the raw EEG data. The parameters of TSLOW, regularity, stereotypy and post-ictal suppression were estimated by analysing the peaks, their amplitudes and locations in the signals of two EEG channels. The estimated parameters were then fed to a fuzzy rule-based inference system to identify seizure quality rating.

  As described earlier, the rules of the fuzzy inference system (FIS) were derived from the literature through identifying the seizure quality. The proposed FIS was evaluated using two cases. In the first one, seizure parameters rated by a team of clinicians were used as input, while in the second case the ratings produced by the proposed parameters estimation algorithm were fed to the FIS.

  In both cases, the predicted scores were found to be related to the benchmark seizure scores. The results demonstrated that it is possible to automatically identify the rating of seizure quality for electroconvulsive therapy. Automatic scoring of seizure quality is expected to be beneficial to clinicians working in this field.

6.3 Directions for Future Research

The following recommendations for future research may improve the results that have been achieved by the techniques that are proposed in this dissertation:

- In terms of feature extraction, examine other methods of feature extraction that may improve the prediction results that were achieved in this research. For example, the relationship between each pairs of EEG channels could be studied using a coherence analysis technique. Special types of machine learning methods as feature extraction techniques could be used, such as DBN and RBM. Also, extract different subsets of features from 750 EEG data under ECT sessions based on DBNs, convolutional deep belief networks and other standard machine learning techniques.
- Improve multi-channel DBN techniques by considering imbalanced classes. As many real-life EEG datasets are imbalanced in nature (i.e., the ratio of samples that belong to one class is noticeably higher than that of other classes), this issue will incorporate developing a dedicated training algorithm for a multi-channel DBN that aims to achieve a better balance between sensitivity and specificity. Oversampling is one of the simple methods to address imbalanced classes. Other sophisticated approaches that address this issue should also be investigated.
- In terms of predicting the tDCS clinical outcomes, dataset collected from a large cohort of patients need to be assessed in order to further validate the proposed approach.
- Develop the other methods for ECT-induced seizure rating prediction, such as a neural mass model. Also, other seizure parameters could be considered, such as peak amplitude, seizure duration and coherence score. Coherence score represents

the relation between two frontal-EEG channels that were used to extract data. Those need to be evaluated using a larger set of ECT recordings.

Appendix A

Independent Component Analysis

A.1 Decomposing Data Using Independent Component Analysis

A description of the Independent Component Analysis (ICA) technique can be found in [51, 52].

A.1.1 Independent Component Analysis of EEG Data

The independent component analysis (ICA) is an eeglab popular tool to analyse EEG signals. This tool has two significant features:

1. Enable the removal of components that are highly influenced by artefacts.
2. Help to disentangle otherwise mixed brain signals

A.1.2 Independent Component Analysis Decomposition

1 **Running ICA Decomposition** We used EEGLAB to compute ICA components of a dataset of EEG epochs (see Figure A.1) by select [Tools > Run ICA]

Figure A.1: Independent component analysis decomposition

2 **Plotting 2-D Component Scalp Maps** We used this instruction [Plot > Component maps > In $2-D$] to plot $2-D$ scalp component maps as shown in Figure A.2.

3 **Plotting Component Head Plots** Also we plotted a 3-D head plot of component topography (see Figure A.3) by used this instruction: (Plot > Component maps > In $3-D$).

A.2 Studying and Removing ICA Components

We selected [Tools > Reject data using ICA > Reject components by map] to study component properties and label components for rejection. Also, the properties of component can be accessed directly by select [Plot > Component properties]. Figure A.4 explains the procedure of removing ICA components and Figure A.5 describe the characteristic of ICA components.

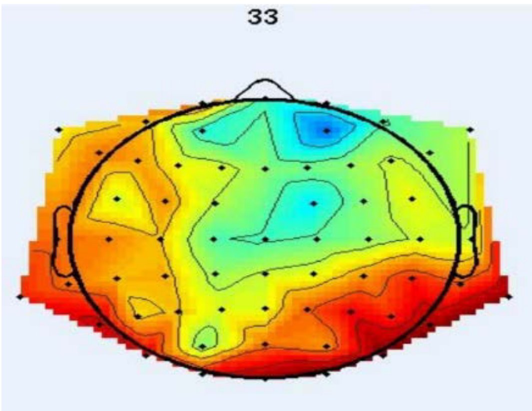
A.2.1 Rejecting Data Epochs by Inspection Using ICA

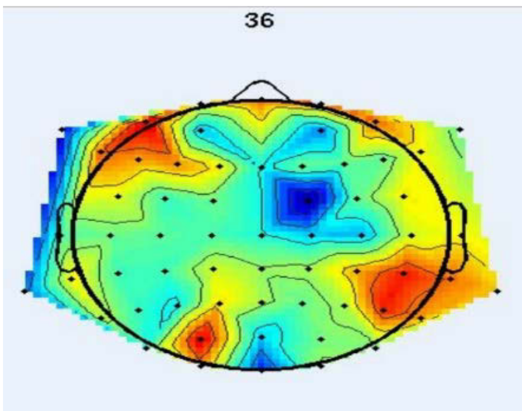
We rejected data by visual inspection of its ICA component activation (see Figure A.6) by use this instruction: (Tools > Reject data using ICA > Reject by inspection).

A.2.2 Scrolling Through Component Activation

We plotted the scroll of component activation (see Figure A.7) by select (Plot > Component activations (scroll)).

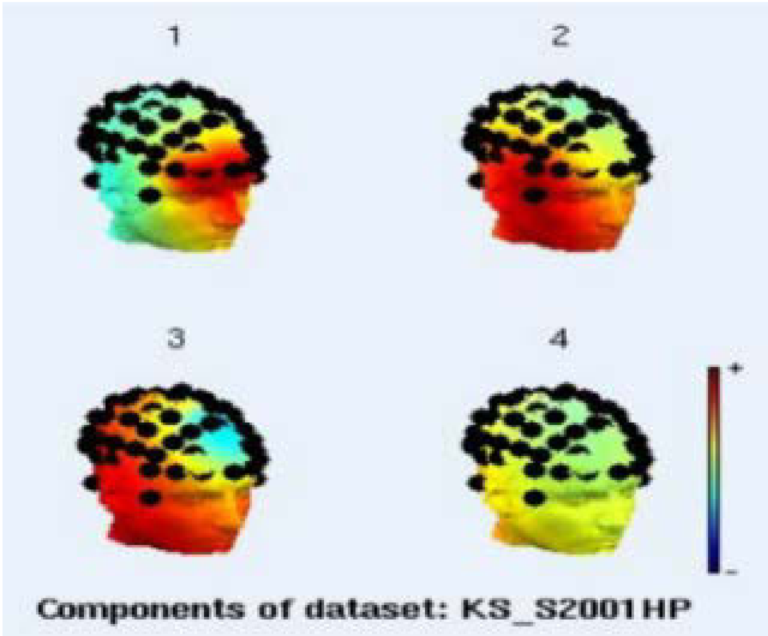
(a)



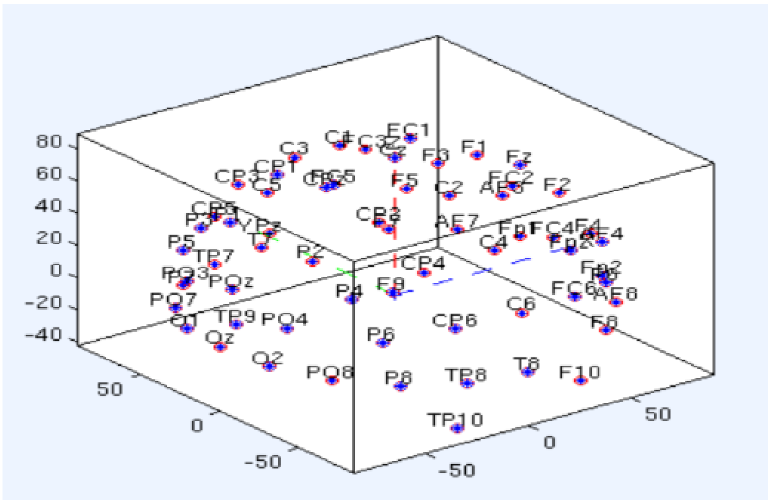


(b) (c)

Figure A.2: 2−D Component Scalp Maps



(a)



(b)

Figure A.3: 3−D Component Head Plots

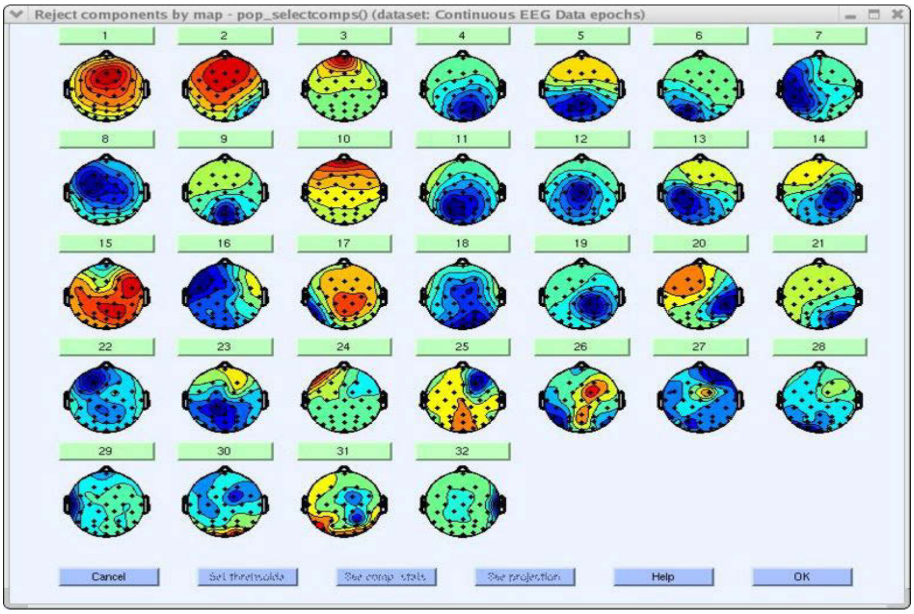


Figure A.4: Reject data using ICA

A.2.3 Plotting Component Spectra and Maps

We plotted component spectra and maps to see which components contribute most strongly to which frequencies in the data by select (Plot > Component spectra and maps)

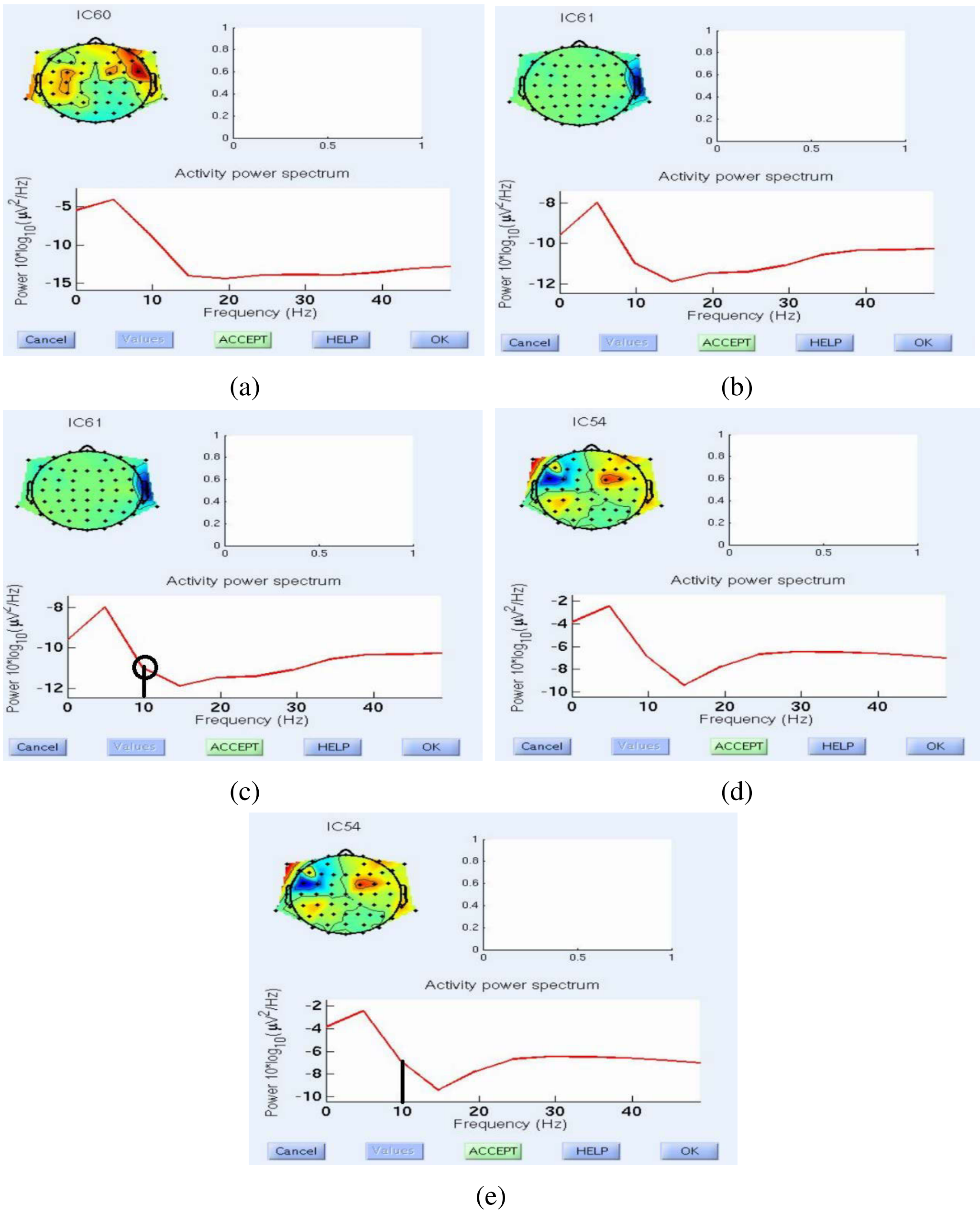
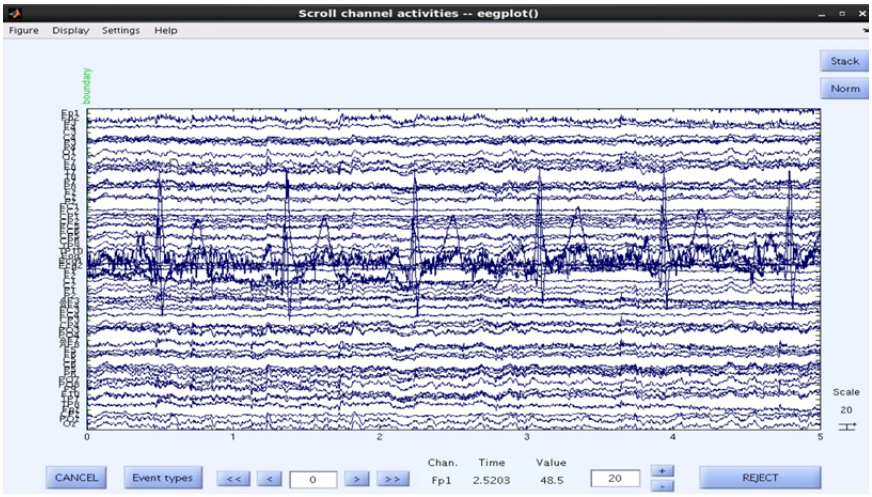
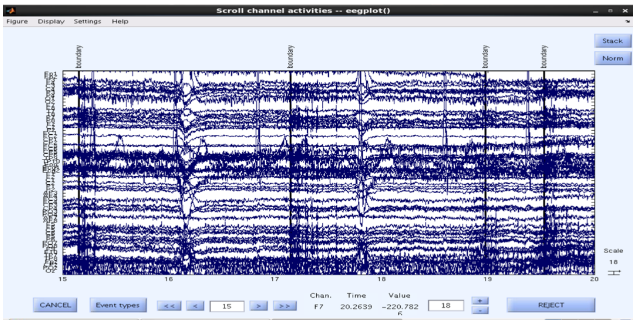


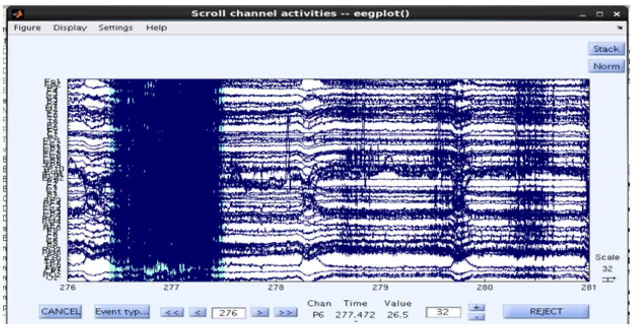
Figure A.5: Component properties



(a)



(b)



(c)

Figure A.6: Rejecting Data Epochs by Inspection Using ICA

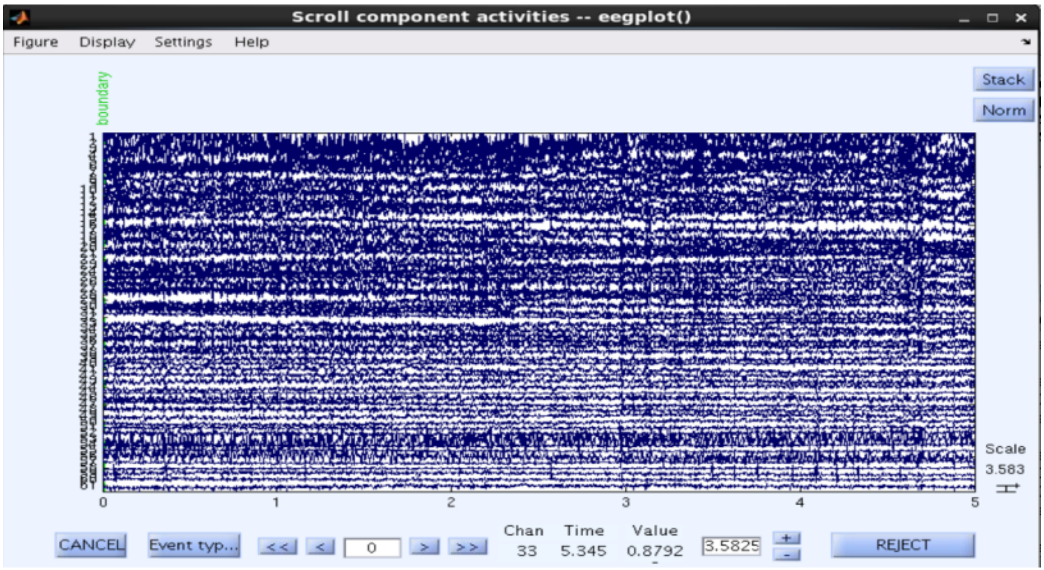


Figure A.7: Scrolling

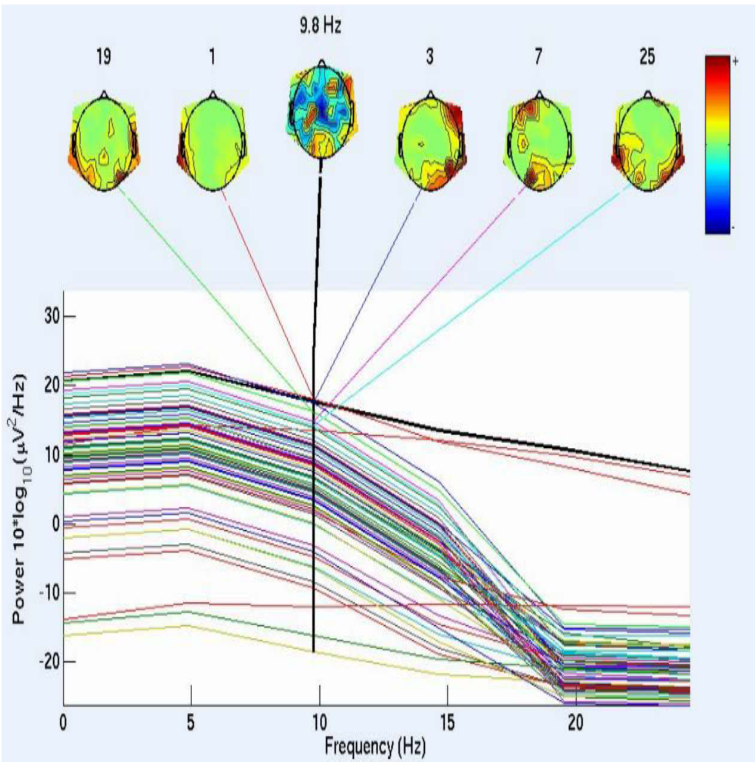


Figure A.8: Component Spectra and Maps

Bibliography

[1] Behshad Hosseinifard, Mohammad Hassan Moradi, and Reza Rostami. Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Computer methods and programs in biomedicine*, 109(3):339–345, 2013.

[2] UG Kalu, CE Sexton, CK Loo, and KP Ebmeier. Transcranial direct current stimulation in the treatment of major depression: a meta-analysis. *Psychological medicine*, 42(09):1791–1800, 2012.

[3] Colleen K Loo, Angelo Alonzo, Donel Martin, Philip B Mitchell, Veronica Galvez, and Perminder Sachdev. Transcranial direct current stimulation for depression: 3-week, randomised, sham-controlled trial. *The British Journal of Psychiatry*, 200(1):52–59, 2012.

[4] Aadil Jan Shah, Ovais Wadoo, and Javed Latoo. Electroconvulsive therapy (ect): important parameters which influence its effectiveness. *Br J Med Pract*, 6(4): a634, 2013.

[5] Ralph Meier, Heike Dittrich, Andreas Schulze-Bonhage, and Ad Aertsen. Detecting epileptic seizures in long-term human EEG: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns. *Journal of Clinical Neurophysiology*, 25(3):119–131, 2008. ISSN 0736-0258.

[6] Lihua Qiu, Xiaoqi Huang, Junran Zhang, Yuqing Wang, Weihong Kuang, Jing Li, Xiuli Wang, Lijuan Wang, Xun Yang, Su Lui, et al. Characterization of major depressive disorder using a multiparametric classification approach based on high resolution structural images. *Journal of psychiatry & neuroscience: JPN*, 39(2): 78, 2014.

[7] Tamara Y Powell, Tjeerd W Boonstra, Donel M Martin, Colleen K Loo, and Michael Breakspear. Modulation of cortical activity by transcranial direct current stimulation in patients with affective disorder. *PloS one*, 9(6):e98503, 2014.

[8] Colleen K Loo, Perminder Sachdev, Donel Martin, Melissa Pigot, Angelo Alonzo, Gin S Malhi, Jim Lagopoulos, and Philip Mitchell. A double-blind, sham-controlled trial of transcranial direct current stimulation for the treatment of depression. *International Journal of Neuropsychopharmacology*, 13(1):61–69, 2010.

[9] VH Do, X Xiao, and ES Chng. Comparison and combination of multilayer perceptrons and deep belief networks in hybrid automatic speech recognition

systems. In *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2011.

[10] Robert J DeRubeis, Greg J Siegle, and Steven D Hollon. Cognitive therapy versus medication for depression: treatment outcomes and neural mechanisms. *Nature Reviews Neuroscience*, 9(10):788, 2008.

[11] Khalid Saad Al-Harbi. Treatment-resistant depression: therapeutic trends, challenges, and future directions. *Patient preference and adherence*, 6:369, 2012.

[12] Peter M Haddad, Peter S Talbot, Ian M Anderson, and R Hamish McAllister-Williams. Managing inadequate antidepressant response in depressive illness. *British medical bulletin*, 115(1), 2015.

[13] American Psychiatric Association. *The practice of electroconvulsive therapy: recommendations for treatment, training, and privileging, 2nd edn*. American Psychiatric Press: Washington, DC, 2001.

[14] W Vaughn McCall. Concerns over antidepressant medications and suicide: what does it mean for ect?, 2005.

[15] Jan Malte Bumb, Suna Su Aksay, Christoph Janke, Laura Kranaster, Olga Geisel, Peter Gass, Rainer Hellweg, and Alexander Sartorius. Focus on ect seizure quality: serum bdnf as a peripheral biomarker in depressed patients. *European archives of psychiatry and clinical neuroscience*, 265(3):227–232, 2015.

[16] IP Collins and IF Scott. Anaesthetic technique in the practice of ect. *The British Journal of Psychiatry*, 1995.

[17] Harry JM Lemmens, David C Levi, Chuck Debattista, and John G Brock-Utne. The timing of electroconvulsive therapy and bispectral index after anesthesia induction using different drugs does not affect seizure duration. *Journal of clinical anesthesia*, 15(1):29–32, 2003.

[18] Piotr Mirowski, Deepak Madhavan, Yann LeCun, and Ruben Kuzniecky. Classification of patterns of EEG synchronization for seizure prediction. *Clinical neurophysiology*, 120(11):1927–1940, 2009.

[19] Deng-Shan Shiau, JJ Halford, KM Kelly, RT Kern, M Inman, Jui-Hong Chien, PM Pardalos, MCK Yang, and J Ch Sackellares. Signal regularity-based automated seizure detection system for scalp EEG monitoring. *Cybernetics and systems analysis*, 46(6):922–935, 2010. ISSN 1060-0396.

[20] Mario Chávez, Michel Le Van Quyen, Vincent Navarro, Michel Baulac, and Jacques Martinerie. Spatio-temporal dynamics prior to neocortical seizures: amplitude versus phase couplings. *IEEE Transactions on Biomedical Engineering*, 50(5):571–583, 2003.

[21] Michel Le Van Quyen, Jacques Martinerie, Vincent Navarro, Michel Baulac, and Francisco J Varela. Characterizing neurodynamic changes before seizures. *Journal of Clinical Neurophysiology*, 18(3):191–208, 2001.

[22] Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[23] Minoru Nakayama and Hiroshi Abe. Single-trial classification of viewed characters using single-channel EEG waveforms. *International Journal for Infonomics (IJI)*, 3(4):392–400, 2010.

[24] Hong Peng, Bin Hu, Quanying Liu, Qunxi Dong, Qinglin Zhao, and Philip Moore. User-centered depression prevention: An EEG approach to pervasive healthcare. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on*, pages 325–330. IEEE, 2011.

[25] Sasikumar Gurumurthy, Vudi Sai Mahit, and Rittwika Ghosh. Analysis and simulation of brain signal data by EEG signal processing technique using matlab. *International Journal of Engineering and Technology*, 5(3):2771–2776, 2013.

[26] Laura Frølich, Tobias S Andersen, and Morten Mørup. Classification of independent components of EEG into multiple artifact classes. *Psychophysiology*, 52(1):32–45, 2015.

[27] Christina M Krause, Lauri Sillanmäki, Mika Koivisto, Carina Saarela, Anna Häggqvist, Matti Laine, and Heikki Hämäläinen. The effects of memory load on event-related EEG desynchronization and synchronization. *Clinical neurophysiology*, 111(11):2071–2078, 2000.

[28] Nicole A Kochan, Michael Valenzuela, Melissa J Slavin, Stacey McCraw, Perminder S Sachdev, Michael Breakspear, et al. Impact of load-related neural processes on feature binding in visuospatial working memory. *PloS one*, 6(8): e23960, 2011.

[29] Hector P Martinez, Yoshua Bengio, and Georgios N Yannakakis. Learning deep physiological models of affect. *Computational Intelligence Magazine, IEEE*, 8 (2):20–33, 2013.

[30] Wei-Long Zheng, Jia-Yi Zhu, Yong Peng, and Bao-Liang Lu. EEG-based emotion classification using deep belief networks. In *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, pages 1–6. IEEE, 2014.

[31] Li Deng, Geoffrey Hinton, and Brian Kingsbury. New types of deep neural network learning for speech recognition and related applications: An overview. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 8599–8603. IEEE, 2013.

[32] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[33] DF Wulsin, JR Gupta, R Mani, JA Blanco, and B Litt. Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement. *Journal of neural engineering*, 8(3): 036015, 2011.

[34] Abdel-rahman Mohamed, George E Dahl, and Geoffrey Hinton. Acoustic modeling using deep belief networks. *Audio, Speech, and Language Processing, IEEE Transactions on*, 20(1):14–22, 2012.

[35] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 689–696, 2011.

[36] D Najumnissa and TR Rangaswamy. Detection and classification of epileptic seizures using wavelet feature extraction and adaptive neuro-fuzzy inference system. *International Journal of Computational Engineering Research*, 2(3): 755–761, 2012.

[37] Verònica Gálvez, Dusan Hadzi-Pavlovic, Harry Wark, Simon Harper, John Leyden, and Colleen K Loo. The anaesthetic-ect time interval in electroconvulsive therapy practice–is it time to time? *Brain stimulation*, 9(1):72–77, 2016. ISSN 1935-861X.

[38] World Health Organization. Depression who@ONLINE, 2015. URL http://www.who.int/topics/depression/en/.

[39] Andrew J Niemiec and Brian J Lithgow. Alpha-band characteristics in EEG spectrum indicate reliability of frontal brain asymmetry measures in diagnosis of depression. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 7517–7520. IEEE, 2006.

[40] Fourth Edition. *Diagnostic and statistical manual of mental disorders*. Am Psychiatric Assoc, 2013.

[41] Saeid Sanei and Jonathon A Chambers. *EEG signal processing*. John Wiley & Sons, 2013.

[42] Hermann Haken. *Principles of brain functioning: A synergetic approach to brain activity, behaviour and cognition*. Springer, 1996.

[43] Michal Teplan. Fundamentals of EEG measurement. *Measurement science review*, 2(2):1–11, 2002.

[44] Donald T Stuss and Robert T Knight. *Principles of frontal lobe function*. Oxford University Press, 2013.

[45] Naga Mallikarjuna Rao Dasari et al. Computational approaches to unveil the dynamics of functional brain networks during cognitive activity using EEG data. 2016.

[46] Christian Uhl. *Analysis of neurophysiological brain functioning*. Springer Science & Business Media, 2012.

[47] Walter Freeman and Rodrigo Quian Quiroga. *Imaging brain function with EEG: advanced temporal and spatial analysis of electroencephalographic signals*. Springer Science & Business Media, 2012.

[48] I NeuroSky. Brain wave signal EEG of neurosky. *NeuroSky Brain-Computer Interface Technologies.[Links]*, 2009.

[49] Mircea Steriade. Cellular substrates of brain rhythms. *Electroencephalography: Basic principles, clinical applications, and related fields*, 5:31–83, 2005.

[50] K Srinivasan, Justin Dauwels, and M Ramasubba Reddy. Multichannel EEG compression: Wavelet-based image and volumetric coding approach. *Biomedical and Health Informatics, IEEE Journal of*, 17(1):113–120, 2013.

[51] Scott Makeig, Anthony J Bell, Tzyy-Ping Jung, Terrence J Sejnowski, et al. Independent component analysis of electroencephalographic data. *Advances in neural information processing systems*, pages 145–151, 1996.

[52] M Ungureanu, C Bigan, R Strungaru, and V Lazarescu. Independent component analysis applied in biomedical signal processing. *Measurement Science Review*, 4(2):18, 2004.

[53] Carlos Alexandre Pereira Carreiras. Interface cérebro-computador (bci) no paradigma de imagiologia motora. 2011.

[54] Amjed S Al-Fahoum and Ausilah A Al-Fraihat. Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. *ISRN neuroscience*, 2014, 2014.

[55] Abdulhamit Subasi, M Kemal Kiymik, Ahmet Alkan, and Etem Koklukaya. Neural network classification of EEG signals by using ar with mle preprocessing for epileptic seizure detection. *Mathematical and Computational Applications*, 10(1):57–70, 2005.

[56] Mohammad Ali Naderi and Homayoun Mahdavi-Nasab. Analysis and classification of EEG signals using spectral analysis and recurrent neural networks. In *Biomedical Engineering (ICBME), 2010 17th Iranian Conference of*, pages 1–4. IEEE, 2010.

[57] Brian J Roach and Daniel H Mathalon. Event-related EEG time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia bulletin*, 34(5):907–926, 2008.

[58] Dean Cvetkovic, Elif Derya Übeyli, and Irena Cosic. Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: A pilot study. *Digital signal processing*, 18(5):861–874, 2008.

[59] Subha D Puthankattil and Paul K Joseph. Analysis of EEG signals using wavelet entropy and approximate entropy: A case study on depression patients. *Int. J. Med. Health Biomed. Pharm. Eng*, 8(7):420–424, 2014.

[60] Alan Gevins and Michael E Smith. Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex*, 10(9):829–839, 2000.

[61] Ming-Chung Ho, Tsung-Ching Chen, Chin-Fei Huang, Cheng-Hsieh Yu, Jhih-Ming Chen, Ray-Ying Huang, Hsing-Chung Ho, and Chia-Ju Liu. Detect ad patients by using EEG coherence analysis. *Journal of Medical Engineering*, 2014, 2014.

[62] Robert W Thatcher. Coherence, phase differences, phase shift, and phase lock in EEG/ERP analyses. *Developmental neuropsychology*, 37(6):476–496, 2012.

[63] David Balin Chorlian, Madhavi Rangaswamy, and Bernice Porjesz. EEG coherence: topography and frequency structure. *Experimental brain research*, 198(1):59–83, 2009.

[64] Yang Li, Yingjie Li, Shanbao Tong, Yingying Tang, and Yisheng Zhu. More normal EEGs of depression patients during mental arithmetic than rest. In *Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging, 2007. NFSI-ICFBI 2007. Joint Meeting of the 6th International Symposium on*, pages 165–168. IEEE, 2007.

[65] Jonathan RT Davidson, Dana Hughes, Dana G Blazer, and Linda K George. Post-traumatic stress disorder in the community: an epidemiological study. *Psychological medicine*, 21(03):713–721, 1991.

[66] Klaus Linkenkaer-Hansen, Simo Monto, Heikki Rytsälä, Kirsi Suominen, Erkki Isometsä, and Seppo Kähkönen. Breakdown of long-range temporal correlations in theta oscillations in patients with major depressive disorder. *The Journal of Neuroscience*, 25(44):10131–10137, 2005.

[67] RA Segrave, RH Thomson, Nicholas R Cooper, Rodney J Croft, DM Sheppard, and PB Fitzgerald. Upper alpha activity during working memory processing reflects abnormal inhibition in major depression. *Journal of affective disorders*, 127(1):191–198, 2010.

[68] Eric R Kandel, James H Schwartz, Thomas M Jessell, et al. *Principles of neural science*, volume 4. McGraw-Hill New York, 2000.

[69] Richard J Davidson. EEG measures of cerebral asymmetry: Conceptual and methodological issues. *International Journal of Neuroscience*, 39(1-2):71–89, 1988.

[70] Noppadon Jatupaiboon, Setha Pan-ngum, and Pasin Israsena. Emotion classification using minimal EEG channels and frequency bands. In *Computer Science and Software Engineering (JCSSE), 2013 10th International Joint Conference on*, pages 21–24. IEEE, 2013.

[71] B-A Tabacaru. Evaluation of emotional influence on EEG activity in the frontal lobes. In *E-Health and Bioengineering Conference (EHB), 2013*, pages 1–4. IEEE, 2013.

[72] Andre Russowsky Brunoni, Roberta Ferrucci, Felipe Fregni, Paulo Sergio Boggio, and Alberto Priori. Transcranial direct current stimulation for the treatment of major depressive disorder: a summary of preclinical, clinical and translational findings. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 39 (1):9–16, 2012.

[73] Moacyr A Rosa and Sarah H Lisanby. Somatic treatments for mood disorders. *Neuropsychopharmacology*, 37(1):102–116, 2012.

[74] D Keeser, F Padberg, E Reisinger, Oliver Pogarell, Valerie Kirsch, U Palm, Susanne Karch, H-J Möller, MA Nitsche, and Christoph Mulert. Prefrontal direct current stimulation modulates resting EEG and event-related potentials in healthy subjects: a standardized low resolution tomography (sloreta) study. *Neuroimage*, 55(2):644–657, 2011.

[75] Felipe Fregni, Paulo S Boggio, Michael A Nitsche, Marco A Marcolin, Sergio P Rigonatti, and Alvaro Pascual-Leone. Treatment of major depression with transcranial direct current stimulation. *Bipolar disorders*, 8(2):203–204, 2006.

[76] AR Brunoni, R Ferrucci, M Bortolomasi, M Vergari, L Tadini, PS Boggio, M Giacopuzzi, S Barbieri, and A Priori. Transcranial direct current stimulation (tDCS) in unipolar vs. bipolar depressive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 35(1):96–101, 2011.

[77] Paulo S Boggio, Sergio P Rigonatti, Rafael B Ribeiro, Martin L Myczkowski, Michael A Nitsche, Alvaro Pascual-Leone, and Felipe Fregni. A randomized, double-blind clinical trial on the efficacy of cortical direct current stimulation for the treatment of major depression. *International Journal of Neuropsychopharmacology*, 11(2):249–254, 2008.

[78] Sergio P Rigonatti, Paulo S Boggio, Martin L Myczkowski, Emma Otta, Juliana T Fiquer, Rafael B Ribeiro, Michael A Nitsche, Alvaro Pascual-Leone, and Felipe Fregni. Transcranial direct stimulation and fluoxetine for the treatment of depression. *European Psychiatry*, 23(1):74–76, 2008.

[79] Marcelo T Berlim, Frederique Van den Eynde, and Z Jeff Daskalakis. Clinical utility of transcranial direct current stimulation (tDCS) for treating major depression: a systematic review and meta-analysis of randomized, double-blind and sham-controlled trials. *Journal of psychiatric research*, 47(1):1–7, 2013.

[80] David Liebetanz, Michael A Nitsche, Frithjof Tergau, and Walter Paulus. Pharmacological approach to the mechanisms of transcranial dc-stimulation-induced after-effects of human motor cortex excitability. *Brain*, 125(10):2238–2247, 2002.

[81] Michael A Nitsche, Paulo S Boggio, Felipe Fregni, and Alvaro Pascual-Leone. Treatment of depression with transcranial direct current stimulation (tDCS): a review. *Experimental neurology*, 219(1):14–19, 2009.

[82] Abraham P Arul-Anandam and Colleen Loo. Transcranial direct current stimulation: a new tool for the treatment of depression? *Journal of affective disorders*, 117(3):137–145, 2009.

[83] Prateek C Gandiga, Friedhelm C Hummel, and Leonardo G Cohen. Transcranial DC stimulation (tDCS): a tool for double-blind sham-controlled clinical studies in brain stimulation. *Clinical neurophysiology*, 117(4):845–850, 2006.

[84] JWT Redfearn, OCJ Lippold, and R Costain. Preliminary account of the clinical effects of polarizing the brain in certain psychiatric disorders. *The British Journal of Psychiatry*, 110(469):773–785, 1964.

[85] Jeffrey T Turner. *Time series analysis using deep feed forward neural networks*. University of Maryland, Baltimore County, 2014.

[86] N Sivasankari and K Thanushkodi. Epileptic seizure detection on EEG signal using statistical signal processing and neural networks. In *Proceedings of the 1st WSEAS international conference on Sensors and signals*, pages 98–102. World Scientific and Engineering Academy and Society (WSEAS), 2008.

[87] C Edward and M Coffey. The ictal EEG as a marker of adequate stimulus intensity with unilateral ect. *Neurosciences*, 7:295–303, 1995.

[88] Andrew D Krystal, Richard D Weiner, C Edward Coffey, and W Vaughn McCall. Effect of ect treatment number on the ictal EEG. *Psychiatry research*, 62(2): 179–189, 1996. ISSN 0165-1781.

[89] Tarique D Perera, Bruce Luber, Mitchell S Nobler, Joan Prudic, Christopher Anderson, and Harold A Sackeim. Seizure expression during electroconvulsive therapy: relationships with clinical outcome and cognitive side effects. *Neuropsychopharmacology*, 29(4):813, 2004. ISSN 0893-133X.

[90] Christian Geretsegger, Marius Nickel, Berthold Judendorfer, Erika Rochowanski, Erich Novak, and Wolfgang Aichhorn. Propofol and methohexital as anesthetic agents for electroconvulsive therapy: a randomized, double-blind comparison of electroconvulsive therapy seizure quality, therapeutic efficacy, and cognitive performance. *The journal of ECT*, 23(4):239–243, 2007. ISSN 1095-0680.

[91] Carolin Hoyer, Laura Kranaster, Christoph Janke, and Alexander Sartorius. Impact of the anesthetic agents ketamine, etomidate, thiopental, and propofol on seizure parameters and seizure quality in electroconvulsive therapy: a retrospective study. *European archives of psychiatry and clinical neuroscience*, 264(3):255–261, 2014. ISSN 0940-1334.

[92] W Michael Hooten and Keith G Rasmussen Jr. Effects of general anesthetic agents in adults receiving electroconvulsive therapy: a systematic review. *The journal of ECT*, 24(3):208–223, 2008. ISSN 1095-0680.

[93] Ardalan Aarabi and Bin He. A rule-based seizure prediction method for focal neocortical epilepsy. *Clinical Neurophysiology*, 123(6):1111–1122, 2012. ISSN 1388-2457.

[94] SS Viglione and GO Walsh. Proceedings: Epileptic seizure prediction. *Electroencephalography and clinical neurophysiology*, 39(4):435, 1975. ISSN 0013-4694.

[95] Deng Wang, Duoqian Miao, and Chen Xie. Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection. *Expert Systems with Applications*, 38(11):14314–14320, 2011.

[96] Reeda Kunhimangalam, Sujith Ovallath, and Paul K Joseph. A novel fuzzy expert system for the identification of severity of carpal tunnel syndrome. *BioMed research international*, 2013, 2013.

[97] Nina Hakacova, Elin Trägårdh-Johansson, Galen S Wagner, Charles Maynard, and Olle Pahlm. Computer-based rhythm diagnosis and its possible influence on nonexpert electrocardiogram readers. *Journal of electrocardiology*, 45(1):18–22, 2012.

[98] My Chau Tu, Dongil Shin, and Dongkyoo Shin. A comparative study of medical data classification methods based on decision tree and bagging algorithms. In *Dependable, Autonomic and Secure Computing, 2009. DASC'09. Eighth IEEE International Conference on*, pages 183–187. IEEE, 2009.

[99] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521 (7553):436–444, 2015.

[100] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[101] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification. 2003.

[102] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[103] Kwokleung Chan, Te-Won Lee, Pamela Sample, Michael H Goldbaum, Robert N Weinreb, Terrence J Sejnowski, et al. Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *Biomedical Engineering, IEEE Transactions on*, 49(9):963–974, 2002.

[104] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: theory and applications. *Neurocomputing*, 70(1):489–501, 2006.

[105] Ping Tan, Weiping Sa, and Lingli Yu. Applying extreme learning machine to classification of EEG BCI. In *Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2016 IEEE International Conference on*, pages 228–232. IEEE, 2016.

[106] RJ Rak, Andrzej Majkowski, and M Kołodziej. Linear discriminant analysis as EEG features reduction technique for brain-computer interfaces. *Przegląd Elektrotechniczny*, 88:28–30, 2012.

[107] Vitaly Levashenko and Elena Zaitseva. Fuzzy decision trees in medical decision making support system. In *Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on*, pages 213–219. IEEE, 2012.

[108] Sushmita Mitra, Kishori M Konwar, and Sankar K. Pal. Fuzzy decision tree, linguistic rules and fuzzy knowledge-based network: generation and evaluation. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 32(4):328–339, 2002.

[109] Cezary Z Janikow. Fuzzy decision trees: issues and methods. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 28(1):1–14, 1998.

[110] Inan Güler and Elif Derya Übeyli. Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. *Journal of neuroscience methods*, 148(2):113–121, 2005.

[111] Ínan Güler and Elif Derya Übeyli. Detection of ophthalmic artery stenosis by least-mean squares backpropagation neural network. *Computers in Biology and Medicine*, 33(4):333–343, 2003.

[112] Elif Derya Übeyli and Inan Güler. Neural network analysis of internal carotid arterial doppler signals: predictions of stenosis and occlusion. *Expert Systems with Applications*, 25(1):1–13, 2003.

[113] AS Miller, BH Blott, et al. Review of neural network applications in medical imaging and signal processing. *Medical and Biological Engineering and Computing*, 30(5):449–464, 1992.

[114] William G Baxt. Use of an artificial neural network for data analysis in clinical decision-making: the diagnosis of acute coronary occlusion. *Neural computation*, 2(4):480–489, 1990.

[115] David L Poole and Alan K Mackworth. *Artificial Intelligence: foundations of computational agents*. Cambridge University Press, 2010.

[116] Detlef Nauck and Rudolf Kruse. Obtaining interpretable fuzzy classification rules from medical data. *Artificial intelligence in medicine*, 16(2):149–169, 1999.

[117] Gang Feng. A survey on analysis and design of model-based fuzzy control systems. *IEEE Transactions on Fuzzy systems*, 14(5):676–697, 2006.

[118] Piotr S Szczepaniak and Paulo JG Lisboa. *Fuzzy systems in medicine*, volume 41. Physica, 2012.

[119] Meng Joo Er and Yi Zhou. Automatic generation of fuzzy inference systems via unsupervised learning. *Neural Networks*, 21(10):1556–1566, 2008.

[120] Tomoharu Nakashima, Gerald Schaefer, Yasuyuki Yokota, and Hisao Ishibuchi. A weighted fuzzy classifier and its application to image processing tasks. *Fuzzy sets and systems*, 158(3):284–294, 2007.

[121] Hamid Mohamadi, Jafar Habibi, Mohammad Saniee Abadeh, and Hamid Saadi. Data mining with a simulated annealing based fuzzy classification system. *Pattern Recognition*, 41(5):1824–1833, 2008.

[122] Ioannis Gadaras and Ludmil Mikhailov. An interpretable fuzzy rule-based classification methodology for medical diagnosis. *Artificial intelligence in medicine*, 47(1):25–41, 2009.

[123] Mohammed Amine Chikh, Mohammed Ammar, and Radja Marouf. A neuro-fuzzy identification of ECG beats. *Journal of medical systems*, 36(2):903–914, 2012.

[124] Reza Ali Mohammadpour, Seyed Mohammad Abedi, Somayeh Bagheri, and Ali Ghaemian. Fuzzy rule-based classification system for assessing coronary artery disease. *Computational and mathematical methods in medicine*, 2015, 2015.

[125] Shinq-Jen Wu, Cheng-Tao Wu, and Jyh-Yeong Chang. Adaptive neural-based fuzzy modeling for biological systems. *Mathematical biosciences*, 242(2):153–160, 2013.

[126] Philippe Hamel and Douglas Eck. Learning features from music audio with deep belief networks. In *ISMIR*, pages 339–344. Utrecht, The Netherlands, 2010.

[127] Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E Alsaadi. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11–26, 2017.

[128] In-Jung Kim and Xiaohui Xie. Handwritten hangul recognition using deep convolutional neural networks. *International Journal on Document Analysis and Recognition (IJDAR)*, 18(1):1–13, 2015.

[129] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[130] Ran Manor and Amir B Geva. Convolutional neural network for multi-category rapid serial visual presentation bci. *Frontiers in computational neuroscience*, 9, 2015.

[131] Benigno Uría. A deep belief network for the acoustic-articulatory inversion mapping problem. 2011.

[132] Nils J Nilsson. *The quest for artificial intelligence*. Cambridge University Press, 2009.

[133] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, DTIC Document, 1985.

[134] Yoshua Bengio, Pascal Lamblin, Dan Popovici, Hugo Larochelle, et al. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19:153, 2007.

[135] Roberto Calandra, Tapani Raiko, Marc Peter Deisenroth, and Federico Montesino Pouzols. Learning deep belief networks from non-stationary streams. In *Artificial Neural Networks and Machine Learning–ICANN 2012*, pages 379–386. Springer, 2012.

[136] Li Deng and Dong Yu. Deep convex net: A scalable architecture for speech pattern classification. In *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[137] Li Deng. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing*, 3, 2014.

[138] Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1):1, 2015.

[139] Kyunghyun Cho et al. Foundations and advances in deep learning. 2014.

[140] Ruslan Salakhutdinov. *Learning deep generative models*. PhD thesis, University of Toronto, 2009.

[141] Christian Plahl, Tara N Sainath, Bhuvana Ramabhadran, and David Nahamoo. Improved pre-training of deep belief networks using sparse encoding symmetric machines. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 4165–4168. IEEE, 2012.

[142] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[143] Amin Emamzadeh Esmaeili Nejad. An application of deep belief networks for 3-dimensional image reconstruction. *ISSN*, 7(01):618–625, 2014.

[144] Erik Schmidt and Youngmoo Kim. Learning rhythm and melody features with deep belief networks. In *ISMIR*, pages 21–26, 2013.

[145] George Dahl, Abdel-rahman Mohamed, Geoffrey E Hinton, et al. Phone recognition with the mean-covariance restricted boltzmann machine. In *Advances in neural information processing systems*, pages 469–477, 2010.

[146] Xiaoyang Yang. *Modeling Natural Images Using Deep Belief Networks and Factored 3-Way Restricted Boltzmann Machine*. PhD thesis, Master Thesis in Statistical Science, University of California, Los Angeles, 2010.

[147] Roland Memisevic and Geoffrey Hinton. Unsupervised learning of image transformations. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[148] Bin Liao, Jungang Xu, Jintao Lv, and Shilong Zhou. An image retrieval method for binary images based on DBN and softmax classifier. *IETE Technical Review*, 32(4):294–303, 2015.

[149] Asja Fischer and Christian Igel. An introduction to restricted boltzmann machines. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 14–36. Springer, 2012.

[150] David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning algorithm for boltzmann machines*. *Cognitive science*, 9(1):147–169, 1985.

[151] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[152] Asja Fischer and Christian Igel. Training restricted boltzmann machines: an introduction. *Pattern Recognition*, 47(1):25–39, 2014.

[153] Sean Patrick Parker. Gpu implementation of a deep learning network for image recognition tasks. 2012.

[154] Ruslan Salakhutdinov and Geoffrey Hinton. An efficient learning procedure for deep boltzmann machines. *Neural computation*, 24(8):1967–2006, 2012.

[155] Ruhi Sarikaya, Geoffrey E Hinton, and Anoop Deoras. Application of deep belief networks for natural language understanding. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 22(4):778–784, 2014.

[156] Graham W Taylor, Geoffrey E Hinton, and Sam T Roweis. Modeling human motion using binary latent variables. In *Advances in neural information processing systems*, pages 1345–1352, 2007.

[157] Peter V Gehler, Alex D Holub, and Max Welling. The rate adapting poisson model for information retrieval and object recognition. In *Proceedings of the 23rd international conference on Machine learning*, pages 337–344. ACM, 2006.

[158] Hugo Larochelle and Yoshua Bengio. Classification using discriminative restricted boltzmann machines. In *Proceedings of the 25th international conference on Machine learning*, pages 536–543. ACM, 2008.

[159] Volodymyr Mnih, Hugo Larochelle, and Geoffrey E Hinton. Conditional restricted boltzmann machines for structured output prediction. *arXiv preprint arXiv:1202.3748*, 2012.

[160] Geoffrey Hinton. A practical guide to training restricted boltzmann machines. *Momentum*, 9(1):926, 2010.

[161] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, and Gang Wang. Recent advances in convolutional neural networks. *arXiv preprint arXiv:1512.07108*, 2015.

[162] Saleh Albelwi and Ausif Mahmood. A framework for designing the architectures of deep convolutional neural networks. *Entropy*, 19(6):242, 2017.

[163] Honglak Lee, Peter Pham, Yan Largman, and Andrew Y Ng. Unsupervised feature learning for audio classification using convolutional deep belief networks. In *Advances in neural information processing systems*, pages 1096–1104, 2009.

[164] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y Ng. Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Communications of the ACM*, 54(10):95–103, 2011.

[165] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th annual international conference on machine learning*, pages 609–616. ACM, 2009.

[166] Jingyu Gao, Jinfu Yang, Guanghui Wang, and Mingai Li. A novel feature extraction method for scene recognition based on centered convolutional restricted boltzmann machines. *Neurocomputing*, 214:708–717, 2016.

[167] Yuanfang Ren and Yan Wu. Convolutional deep belief networks for feature extraction of EEG signal. In *Neural Networks (IJCNN), 2014 International Joint Conference on*, pages 2850–2853. IEEE, 2014.

[168] Elif Derya Übeyli. Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals. *Expert Systems with Applications*, 37(1):233–239, 2010.

[169] AS Muthanantha Murugavel and S Ramakrishnan. Hierarchical multi-class svm with elm kernel for epileptic EEG signal classification. *Medical & biological engineering & computing*, 54(1):149–161, 2016.

[170] K Sercan Bayram, M Ayyüce Kızrak, and Bülent Bolat. Classification of EEG signals by using support vector machines. In *Innovations in Intelligent Systems and Applications (INISTA), 2013 IEEE International Symposium on*, pages 1–3. IEEE, 2013.

[171] Aayush Bhardwaj, Ankit Gupta, Pallav Jain, Asha Rani, and Jyoti Yadav. Classification of human emotions from EEG signals using SVM and LDA classifiers. In *Signal Processing and Integrated Networks (SPIN), 2015 2nd International Conference on*, pages 180–185. IEEE, 2015.

[172] Hyeon-min Shim and Sangmin Lee. Multi-channel electromyography pattern classification using deep belief networks for enhanced user experience. *Journal of Central South University*, 22(5):1801–1808, 2015.

[173] Nitish Srivastava and Ruslan Salakhutdinov. Learning representations for multimodal data with deep belief nets. In *International Conference on Machine Learning Workshop*, 2012.

[174] Dongyang Cheng, Tanfeng Sun, Xinghao Jiang, and Shilin Wang. Unsupervised feature learning using markov deep belief network. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 260–264. IEEE, 2013.

[175] Dan Wang and Yi Shang. Modeling physiological data with deep belief networks. *International journal of information and education technology (IJIET)*, 3(5):505, 2013.

[176] Biao Leng, Xiangyang Zhang, Ming Yao, and Zhang Xiong. 3d object classification using deep belief networks. In *MultiMedia Modeling*, pages 128–139. Springer, 2014.

[177] Takashi Kuremoto, Shinsuke Kimura, Kunikazu Kobayashi, and Masanao Obayashi. Time series forecasting using a deep belief network with restricted boltzmann machines. *Neurocomputing*, 137:47–56, 2014.

[178] Junying Hu, Jiangshe Zhang, Nannan Ji, and Chunxia Zhang. A new regularized restricted boltzmann machine based on class preserving. *Knowledge-Based Systems*, 123:1–12, 2017.

[179] Yan Wu and HJ Cai. A simulation study of deep belief network combined with the self-organizing mechanism of adaptive resonance theory. In *Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on*, pages 1–4. IEEE, 2010.

[180] Yoshua Bengio. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[181] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8): 2554–2558, 1982.

[182] Geoffrey E Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800, 2002.

[183] Amin Sobhani. *P300 classification using deep belief nets*. PhD thesis, Colorado State University, 2014.

[184] Caroline D Rae, Vincent H-C Lee, Roger J Ordidge, Angelo Alonzo, and Colleen Loo. Anodal transcranial direct current stimulation increases brain intracellular ph and modulates bioenergetics. *International Journal of Neuropsychopharmacology*, 16(8):1695–1706, 2013.

[185] Charlotte J Stagg and Michael A Nitsche. Physiological basis of transcranial direct current stimulation. *The Neuroscientist*, 17(1):37–53, 2011.

[186] Michael A Nitsche, Leonardo G Cohen, Eric M Wassermann, Alberto Priori, Nicolas Lang, Andrea Antal, Walter Paulus, Friedhelm Hummel, Paulo S Boggio, Felipe Fregni, et al. Transcranial direct current stimulation: state of the art 2008. *Brain stimulation*, 1(3):206–223, 2008.

[187] Janaina F Oliveira, Tamires A Zanao, Leandro Valiengo, Paulo A Lotufo, Isabela M Benseñor, Felipe Fregni, and André R Brunoni. Acute working memory improvement after tDCS in antidepressant-free patients with major depressive disorder. *Neuroscience letters*, 537:60–64, 2013.

[188] André R Brunoni, Adriano H Moffa, Felipe Fregni, Ulrich Palm, Frank Padberg, Daniel M Blumberger, Zafiris J Daskalakis, Djamila Bennabi, Emmanuel Haffen, Angelo Alonzo, et al. Transcranial direct current stimulation for acute major depressive episodes: meta-analysis of individual patient data. *The British Journal of Psychiatry*, pages bjp–bp, 2016. ISSN 0007-1250.

[189] Hannah L Filmer, Paul E Dux, and Jason B Mattingley. Applications of transcranial direct current stimulation for understanding brain function. *Trends in neurosciences*, 37(12):742–753, 2014.

[190] Mouhsin M Shafi, M Brandon Westover, Michael D Fox, and Alvaro Pascual-Leone. Exploration and modulation of brain network interactions with noninvasive brain stimulation in combination with neuroimaging. *European Journal of Neuroscience*, 35(6):805–825, 2012.

[191] Ryan Thibodeau, Randall S Jorgensen, and Sangmoon Kim. Depression, anxiety, and resting frontal EEG asymmetry: a meta-analytic review., 2006. ISSN 1939-1846.

[192] Ian H Gotlib. EEG alpha asymmetry, depression, and cognitive functioning. *Cognition & Emotion*, 12(3):449–478, 1998. ISSN 0269-9931.

[193] AH Kemp, K Griffiths, KL Felmingham, SA Shankman, WHIM Drinkenburg, M Arns, CR Clark, and RA Bryant. Disorder specificity despite comorbidity: resting EEG alpha asymmetry in major depressive disorder and post-traumatic stress disorder. *Biological psychology*, 85(2):350–354, 2010. ISSN 0301-0511.

[194] Craig E Tenke, Jürgen Kayser, Carlye G Manna, Shiva Fekri, Christopher J Kroppmann, Jennifer D Schaller, Daniel M Alschuler, Jonathan W Stewart, Patrick J McGrath, and Gerard E Bruder. Current source density measures of electroencephalographic alpha predict antidepressant treatment response. *Biological psychiatry*, 70(4):388–394, 2011. ISSN 0006-3223.

[195] Gerard E Bruder, James P Sedoruk, Jonathan W Stewart, Patrick J McGrath, Frederic M Quitkin, and Craig E Tenke. Electroencephalographic alpha measures predict therapeutic response to a selective serotonin reuptake inhibitor antidepressant: pre-and post-treatment findings. *Biological psychiatry*, 63(12): 1171–1177, 2008. ISSN 0006-3223.

[196] Alaa M Al-kaysi, Ahmed Al-Ani, and Tjeerd W Boonstra. A multichannel deep belief network for the classification of EEG data. In *Neural Information Processing*, pages 38–45. Springer, 2015.

[197] Agata Woźniak-Kwaśniewska, David Szekely, Sylvain Harquel, Thierry Bougerol, and Olivier David. Resting electroencephalographic correlates of the clinical response to repetitive transcranial magnetic stimulation: A preliminary comparison between unipolar and bipolar depression. *Journal of affective disorders*, 183:15–21, 2015.

[198] David V Sheehan, Yves Lecrubier, K Harnett Sheehan, Patricia Amorim, Juris Janavs, Emmanuelle Weiller, Thierry Hergueta, Roxy Baker, Geoffrey C Dunbar, et al. The mini-international neuropsychiatric interview (mini): the development and validation of a structured diagnostic psychiatric interview for dsm-iv and icd-10. *Journal of clinical psychiatry*, 59:22–33, 1998.

[199] Marie Åsberg, SA Montgomery, Carlo Perris, Daisy Schalling, and Göran Sedvall. A comprehensive psychopathological rating scale. *Acta Psychiatrica Scandinavica*, 57(S271):5–27, 1978.

[200] Aaron Smith. *Symbol digit modalities test*. Western Psychological Services, Los Angeles, CA, 1991.

[201] John JB Allen, James A Coan, and Maria Nazarian. Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion. *Biological psychology*, 67(1):183–218, 2004.

[202] Terrence R Oakes, Diego A Pizzagalli, Andrew M Hendrick, Katherine A Horras, Christine L Larson, Heather C Abercrombie, Stacey M Schaefer, John V Koger, and Richard J Davidson. Functional coupling of simultaneous electrical and metabolic activity in the human brain. *Human brain mapping*, 21(4):257–270, 2004.

[203] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Stanford, CA, 1995.

[204] Donel M Martin, Kevin Yeung, and Colleen K Loo. Pre-treatment letter fluency performance predicts antidepressant response to transcranial direct current stimulation. *Journal of affective disorders*, 203:130–135, 2016.

[205] Siwei Bai, Socrates Dokos, Kerrie-Anne Ho, and Colleen Loo. A computational modelling study of transcranial direct current stimulation montages used in depression. *Neuroimage*, 87:332–344, 2014. ISSN 1053-8119.

[206] B Saletu, P Anderer, and GM Saletu-Zyhlarz. EEG topography and tomography (loreta) in diagnosis and pharmacotherapy of depression. *Clinical EEG and neuroscience*, 41(4):203–210, 2010. ISSN 1550-0594.

[207] Gerard E Bruder, Jonathan W Stewart, Craig E Tenke, Patrick J McGrath, Paul Leite, Nil Bhattacharya, and Frederic M Quitkin. Electroencephalographic and perceptual asymmetry differences between responders and nonresponders to an ssri antidepressant. *Biological psychiatry*, 49(5):416–425, 2001. ISSN 0006-3223.

[208] Felipe Fregni, Paulo S Boggio, Michael Nitsche, Felix Bermpohl, Andrea Antal, Eva Feredoes, Marco A Marcolin, Sergio P Rigonatti, Maria TA Silva, Walter Paulus, et al. Anodal transcranial direct current stimulation of prefrontal cortex

enhances working memory. *Experimental brain research*, 166(1):23–30, 2005. ISSN 0014-4819.

[209] Suk Hoon Ohn, Chang-Il Park, Woo-Kyoung Yoo, Myoung-Hwan Ko, Kyung Pil Choi, Gyeong-Moon Kim, Yong Taek Lee, and Yun-Hee Kim. Time-dependent effect of transcranial direct current stimulation on the enhancement of working memory. *Neuroreport*, 19(1):43–47, 2008. ISSN 0959-4965.

[210] Lars Michels, Morteza Moazami-Goudarzi, Daniel Jeanmonod, and Johannes Sarnthein. EEG alpha distinguishes between cuneal and precuneal activation in working memory. *Neuroimage*, 40(3):1296–1310, 2008. ISSN 1053-8119.

[211] Young Sup Woo, Joshua D Rosenblat, Ron Kakar, Won-Myong Bahk, and Roger S McIntyre. Cognitive deficits as a mediator of poor occupational function in remitted major depressive disorder patients. *Clinical Psychopharmacology and Neuroscience*, 14(1):1, 2016. ISSN 1738-1088.

[212] Siamac Fazli, Florin Popescu, Márton Danóczy, Benjamin Blankertz, Klaus-Robert Müller, and Cristian Grozea. Subject-independent mental state classification in single trials. *Neural networks*, 22(9):1305–1312, 2009. ISSN 0893-6080.

[213] RD Weiner, CE Coffey, LJ Fochtmann, RM Greenberg, KE Isenberg, CH Kellner, HA Sackeim, and L Moench. The practice of electroconvulsive therapy. *Recommendations for Treatment, Training, and Privileging: A Task Force Report of the American Psychiatric Association*, 2, 2001.

[214] Andrew D Krystal, Richard D Weiner, W Vaughn McCall, Frank E Shelp, Rebecca Arias, and Pamela Smith. The effects of ect stimulus dose and electrode placement on the ictal electroencephalogram: an intraindividual crossover study. *Biological Psychiatry*, 34(11):759–767, 1993.

[215] Mitchell S Nobler, Harold A Sackeim, Maria Solomou, Bruce Luber, DP Devanand, and Joan Prudic. EEG manifestations during ect: effects of electrode placement and stimulus intensity. *Biological psychiatry*, 34(5):321–330, 1993. ISSN 0006-3223.

[216] Andrew D Krystal, C Edward Coffey, Richard D Weiner, and Tracey Holsinger. Changes in seizure threshold over the course of electroconvulsive therapy affect therapeutic response and are detected by ictal EEG ratings. *The Journal of neuropsychiatry and clinical neurosciences*, 10(2):178–186, 1998. ISSN 0895-0172.

[217] Ross D MacPherson, Jessica Lawford, Brett Simpson, Michelle Mahon, Debra Scott, and Colleen Loo. Low dose lignocaine added to propofol does not attenuate the response to electroconvulsive therapy. *Journal of affective disorders*, 126(1): 330–333, 2010. ISSN 0165-0327.

[218] Richard D Weiner and Andrew D Krystal. EEG monitoring of ect seizures. *The clinical science of electroconvulsive therapy*, pages 93–109, 1993.

[219] Hideki Azuma, Akiko Fujita, Kiyoe Sato, Keiko Arahata, Kazuyuki Otsuki, Miki Hori, Yoshihito Mochida, Megumi Uchida, Tomoko Yamada, Tatsuo Akechi, et al. Postictal suppression correlates with therapeutic efficacy for depression in

bilateral sine and pulse wave electroconvulsive therapy. *Psychiatry and clinical neurosciences*, 61(2):168–173, 2007. ISSN 1440-1819.

[220] W Vaughn McCall, G Dave Robinette, and David Hardesty. Relationship of seizure morphology to the convulsive threshold. *The Journal of ECT*, 12(3): 147–151, 1996. ISSN 1095-0680.

[221] Brian Litt and Javier Echauz. Prediction of epileptic seizures. *The Lancet Neurology*, 1(1):22–30, 2002. ISSN 1474-4422.

[222] Alex Greaves, Arushi Raghuvanshi, and Kai-Yuan Neo. Predicting seizure onset with intracranial electroencephalogram (EEG) data-project report. 2014.

[223] Sriram Ramgopal, Sigride Thome-Souza, Michele Jackson, Navah Ester Kadish, Iván Sánchez Fernández, Jacquelyn Klehm, William Bosl, Claus Reinsberger, Steven Schachter, and Tobias Loddenkemper. Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. *Epilepsy & behavior*, 37:291–307, 2014. ISSN 1525-5050.

[224] Ali Jalali, Daniel J Licht, and C Nataraj. Application of decision tree in the prediction of periventricular leukomalacia (pvl) occurrence in neonates after heart surgery. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 5931–5934. IEEE, 2012.

[225] PL Nunez. Electroencephalography. encyclopedia of human. *Brain*, 2(2):1348, 2002.