A DEEP LEARNING APPROACH FOR CLASSIFICATION OF EEG MOTOR IMAGERY SIGNALS

Author

MEMOONA IFTIKHAR

00000170972

Supervisor

DR.SHOAB AHMAD KHAN

DEPARTMENT OF COMPUTER AND SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

JANUARY, 2019

A DEEP LEARNING APPROACH FOR CLASSIFICATION OF EEG MOTOR IMAGERY SIGNALS

Author

MEMOONA IFTIKHAR

00000170972

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

MS Computer Engineering

Thesis Supervisor:

DR.SHOAB AHMAD KHAN

Thesis Supervisor's Signature:_______________________________________

DEPARTMENT OF COMPUTER AND SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,

ISLAMABAD

JANUARY, 2019

Declaration

I certify that this research work titled “A deep learning approach for classification of motor imagery signals’’ is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources is properly acknowledged / referred.

Signature of Student

Memoona Iftikhar

00000170972

Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical and spelling mistakes. Thesis is also according to the format given by the university.

Signature of Student

Memoona Iftikhar

00000170972

Signature of Supervisor

Dr Shoab Ahmad Khan

Plagiarism Certificate (Turnitin Report)

This thesis has been checked for Plagiarism. Turnitin report endorsed by Supervisor is attached.

Signature of Student

Memoona Iftikhar

Registration Number

00000170972

Signature of Supervisor

Dr Shoab Ahmad Khan

Copyright Statement

Acknowledgements

All Praise and thanks be to Allah! We praise and thank Him, ask Him for His help and Forgiveness, and we seek refuge in Allah from the evils of our souls and the mischiefs of our deeds. He whom Allah guides will not be misled, and he whom Allah misleads will never have a guide. I testify that there is no deity but Allah alone, without any partners, and that Muhammad (SAW) is His 'Abd (worshiper) and Messenger.

I would like to express special thanks to my supervisor Dr. Shoab Ahmad khan for his exemplary supervision, unparalleled insight, and constant encouragement during the course of this thesis. His constant curiosity with the challenges that I faced and his helpful notions gave me the direction to successfully complete my research..

I owe deepest gratitude to my honorable co-supervisor Dr. Ali Hassan without whom it was almost impossible to complete my research work. I am highly obliged for his support and guidance throughout my research work. I wouldn’t be here today without his encouragement throughout the final stages of this research

I would also like to thank Dr. Farhan Hussain and Dr. Omer Bin Saeed for being on my thesis guidance and examination committee and for guidance and cordial support, which helped me in completing this task through various stages.

Finally, I would like to thank my family that have been a constant source of love, support and inspiration for me throughout my education

To my father who have given me the life I love today

Abstract

Electroencephalography (EEG) is one of the most clinically and scientifically exploited signals recorded from humans. Hence, its measurement plays a prominent role in brain studies. In particular, the examination of EEG signals has been recognized as the most preponderant approach to the problem of extracting knowledge of the brain dynamics.

We proposed an EEG signals measurement and analysis methods for BCI. Our purpose of this study is to recognize subject's intention when they move their arms. EEG signals are recorded during the imaginary movement of subject's arms at electrode positions C3, CZ and C4. We analyzed ERS (Event-Related Synchronization) and ERD (Event-Related Desynchronization) which are detected when people move their limbs in the mu wave and beta wave. Results of this study showed that ERD occurred in mu waves and ERS occurred in beta waves at C3 during the imaginary movement of right arm. Similarly, ERD occurred in mu waves and ERS occurred in beta waves at C4 during the imaginary movement of left arm.

Deep learning approaches have been used successfully in many recent studies to learn features and classify different types of data. However, the number of studies that employ these approaches on BCI applications is very limited. In this study we aim to use deep learning methods to improve classification performance of EEG motor imagery signals. In this dissertation we investigate residual network architecture to classify EEG Motor Imagery signals. A new form of input is introduced to combine time, frequency information extracted from EEG signal and it is used as an input to convolutional layers.

The classification performance obtained by the proposed method on BCI competition IV dataset 2b in terms of accuracy is 84.9%, which is to the best of our knowledge is the highest accuracy on underlying dataset. Our results show that deep learning methods provide better classification performance compared to other state of art approaches. These methods can be applied successfully to BCI systems where the amount of data is large due to daily recording.

Keywords: *Brain Computer Interface (BCI), Electroencephalogram (EEG), Signal Classification, Convolutional Neural Network (CNN), Residual Nets (Resnet)*

Table of Contents

List of Figures

List of Tables

Acronyms

BCI	Brain computer interface
EEG	Electroencephalography
MVPA	Multivariate pattern analysis
SVM	Support vector machine
FMRI	Functional Magnetic Resonance Imaging
BOLD	Blood oxygen level dependent
MEG	Magnetoencephalography
CSP	Common Spatial Pattern
PCA	Principal Component Analysis
CAR	Common Average Referencing
SL	Surface Laplacian
ICA	Independent Component Analysis
FFT	Fast Fourier Transform
WT	Wavelet Transform
STFT	Short Time Fourier Transform

CHAPTER 1

Introduction

1.1 Description and Motivation

Brain research has been highlighted as an area of national interest in recent years. This research has the potential to impact many important areas, from the detection, treatment, and increased understanding of diseases such as Alzheimer's and epilepsy, to the neural control of devices to aid the handicapped, to a greater understanding of how the human brain functions on a basic level.

The classification of brain signals recorded by imaging devices using machine learning approaches is a very powerful tool in many of these areas of research. For example, machine learning techniques show promise in the early detection of Alzheimer’s or giving warning before an epileptic seizure. These techniques are already being used in devices such as the P300 speller (Guan et al., 2004) to provide a communication device for the severely handicapped.

In addition, neuroscience problems present a unique set of challenges that require innovation in machine learning. The data obtained from brain activity monitoring devices are noisy, have high dimensionality, and are costly to collect, which limits the number of data samples that can be collected. The combination of these factors leads to exceedingly complex data, which are difficult to analyze or classify, even using the most sophisticated and modern techniques.

This thesis presents novel results in the broad area of brain signal classification. First, it provides a comparative evaluation of standard machine learning and data preprocessing techniques in brain signal classification. Second, the use of deep learning techniques for brain signal classification is explored in detail. While these techniques are state of the art in many other applications of machine learning, there are relatively few published results of their use in brain signal classification.

1.2 Applications

The classification of brain signals is a growing area of research, with emerging applications in both applied and theoretical neuroscience. These applications can generally be divided into a few main areas including device control, brain state detection, medical diagnosis, and basic research.

1.2.1 Basic Research

In neuroscience, the use of machine learning techniques to classify brain signals is seen as a form of multivariate pattern analysis (MVPA). MVPA is used to examine phenomena that are difficult to measure with traditional techniques. For example, support vector machines (SVMs) were used to examine to "what extent item-specific information about complex natural scenes is represented in several category-selective areas of human extra striate visual cortex during visual perception and visual mental imagery" (Johnson, McCarthy, Muller, Brudner, & Johnson, 2015). Other examples of basic research involving the classification of brain signals include topics such as affect recognition, semantic language representation in bilingual speakers, and exploring individual differences in pain tolerance [1]; [2].

1.2.2 Brain State Detection

Another application of classification involves the continuous monitoring of brain states using imaging devices like the EEG. Rather than looking to control an external device, these techniques look to determining the subject's inner state. These applications tend to consider longer periods of data and focus on frequency analysis. These techniques are being researched for use in areas such as seizure detection and prevention [3]; [4] and truth detection (Gao et al., 2013). Brain state detection is also used as part of larger applications, such as monitoring mood to allow a larger human computer interface system to adapt its display to user's current mental state [5]. There is even interest in the classification of more disparate states such whether the subject is resting quietly, remembering events from their day, performing subtraction or silently singing lyrics, [6].

1.2.3 Medical Diagnosis

Brain signal classification is likely to play an increasing role in the diagnosis of brain diseases in the future. While convergent evidence will always be necessary, classification could be useful as a screening tool or another point of reference for diagnosis. For example, efforts have been made into using SVM techniques to aid in Alzheimer's disease diagnosis [7]. Other applications in diagnosis include drug addiction [8] and diagnosis of psychiatric disorders, such as schizophrenia [9], ADHD [10]), and bipolar disorder [10].

1.2.4 Brain Computer Interfaces

Brain computer interfaces (BCIs) use monitored brain activity and computation to achieve an external activity. Classification of mental states and intentions based on patterns of brain signals is a common goal in such applications. While early methods of BCIs have tended to use manually determined features and relied on the user adapting to the machine, more modern techniques generally involve the use of machine learning and allow the machine to adapt to the user.

1.3 Brain Imaging Techniques

A variety of devices and techniques capable of measuring brain signals are available. These techniques either measure primary signals of activity such as the electrical or magnetic signal produced by neural activity, or secondary signals such as the blood flow to regions of the brain that are active.

Functional Magnetic Resonance Imaging (fMRI) measures what is known as the blood oxygen level dependent (BOLD). It is capable of high spatial resolution and imaging deep brain structures, but requires a room free of electromagnetic interference and very expensive equipment. Furthermore, the temporal resolution of the signal is quite poor since it relies on measuring blood flow rather than a direct marker of brain activity. The output of fMRI, after several statistical methods are applied, is a series of voxel based images of the BOLD signal.

Magnetoencephalography (MEG) measures the magnetic component created by the electricity moving through the brain. It has both a relatively high spatial resolution and a very high temporal resolution, since it measures a direct and quickly propagated marker of brain activity that is largely unaffected by the scalp. While the spatial resolution is fairly good, only outer portions of the brain can be accurately measured due to the drop off in strength of magnetic fields with the square of the distance. Additionally, it requires a highly magnetically shielded room and a constant supply of liquid helium to function, leading to very high costs and a lack of portability. The output of MEG is one time series per channel (generally 306 channels) representing the strength of the magnetic field at the channel.

Other brain imaging modalities such as standard magnetic resonance imaging (MRI) and positron emission tomography (PET) do not offer the temporal resolution necessary to monitor brain activity at time scales useful in most classification applications and may have other drawbacks, including exposure to ionizing radiation. Thus, due to the disadvantages of other brain imaging modalities, this thesis will focus on data collected by electroencephalography (EEG).

EEG functions by attaching electrodes to a subject's scalp in order to measure the changes in electrical potential that occur as a result of brain activity. Due to the near instantaneous propagation of these voltage changes, information acquired by the EEG can be sampled with high temporal precision. However, the human skull and scalp are insulators, which have the effect of dispersing the signal, thus limiting the spatial resolution capable of being achieved by the EEG. Furthermore, localizing the source of activity involves solving an inverse problem with an infinite number of solutions, thus further limiting the spatial resolution. However, it is comparatively inexpensive and does not require onerous protections such as magnetically shielded rooms.

Furthermore, it is, unlike the previously discussed modalities, more practical in applications that require mobility, such as BCI. The output of EEG is one time series per channel (anywhere from 32 to 256 channels is common) that represents the electrical potential on the scalp at the given channel with respect to a reference electrode, typically recorded at rates from 250 Hz to 1000 Hz.

1.4 Research Challenges and Contributions

Brain imaging data presents several challenging obstacles to machine learning, all of which are current topics of open research:

Brain signals are noisy. The information is polluted by a variety of factors including muscle movement, measurement error, brain activity that is not of interest, and electromagnetic interference from the environment.

Brain signals have high dimensionality. There are frequently hundreds of channels, sampled at up to 1000 Hz. The raw data frequently presents up to hundreds of thousands of features per trial. The data are a time series and have spatial interactions, potentially requiring investigation of temporal and frequency components in conjunction with spatial analyses.

Data collection is expensive and time consuming. Thus, it can be difficult to collect sufficient data for many of the most powerful machine learning techniques. Generally thousands or tens of thousands of samples per class are desired in modern deep learning applications, but it is often impractical to produce more than a few hundred samples per subject in brain imaging tasks, particularly if they need to be derived for clinical or medical studies. Brain activity can vary significantly between subjects and even between data acquisition sessions within a subject. Thus, classifiers that can address high levels of variability are needed. The goal of this thesis is to investigate techniques for addressing these challenges.

In this thesis we have demonstrated the effectiveness of deep neural networks in classifying brain signals.

1.5 Thesis Outline

The rest of the thesis is organized as follows: Chapter 2 will review the related works that forms the basis for this research. Chapter 3 will define the specific problems and approaches used in this thesis. Chapter 4 will include a summary of the datasets used, the implementation of the techniques used, and the results of the experiments. Chapter 5 will present a summary of the findings and a discussion on future work.

CHAPTER 2

Literature Review

The purpose of this dissertation is to develop techniques that are able to categorize different types of electroencephalography (EEG) signals in order to diagnose and treat neurological diseases. This chapter provides broader aspect of signal classification along with overview, concepts and different classification methods used to classify EEG signal. Prior to the classification concepts, this chapter introduces common ideas about background of EEG and BCI's.

2.1 EEG signals Background

The first section of this chapter provides neurophysiology structure of brain which actually generates EEG signals. In the following section the mechanism how EEG signals are generated is discussed. Different types of rhythm in EEG signals are also highlighted in the subsequent section. Further, the principle of Brain Computer Interface (BCI) along with its application has been picked up for the understanding. Finally different classification methods that have been used previously are discussed in detail.

2.1.1 Neurophysiology of Human brains

There are 100 billion neurons in human brain, and these neurons are responsible for maintaining electrical charge of brain. There exists potential between these neurons through which they communicate to each other. In order to activate a neuron, one neuron transmits a chemical to the other neuron through a gap between them which is called synapse also known as neurotransmitter. When an action taken place or in other words when a neuron is activated by this chemical, there exist a potential between neurons which causes local current. This potential can be divided into two subdivisions; which are postsynaptic potentials and action potential. It the former potential reaches to a certain threshold of conduction level, the neuron fires and later potential is activated. The postsynaptic potential is summated at the cortex and spread all over the scalp which in return recorded as EEG. While, on other hand the potential field of action potential have a smaller spread and as well as small duration. Therefore, action potential doesn't contribute in recording of EEG signals either on scalp or intracranial. The potential produced by a single nerve is typically very small to measure it accurately with today's methods so large number of active neurons can produce electrical activity which is recordable on the scalp [11].

While taking the EEG measurement, the cerebral cortex recording is most important measurement because this portion of brain is mostly responsible for intellectual tasks. This portion has great influence on EEG measurements because of its position. All types of movement and processing of complex visual information, language comprehension and problem solving tasks are done by cerebral cortex. The anatomy of human brain is shown in figure 2.1, all activities of major portion of brain is highlighted with their functions.

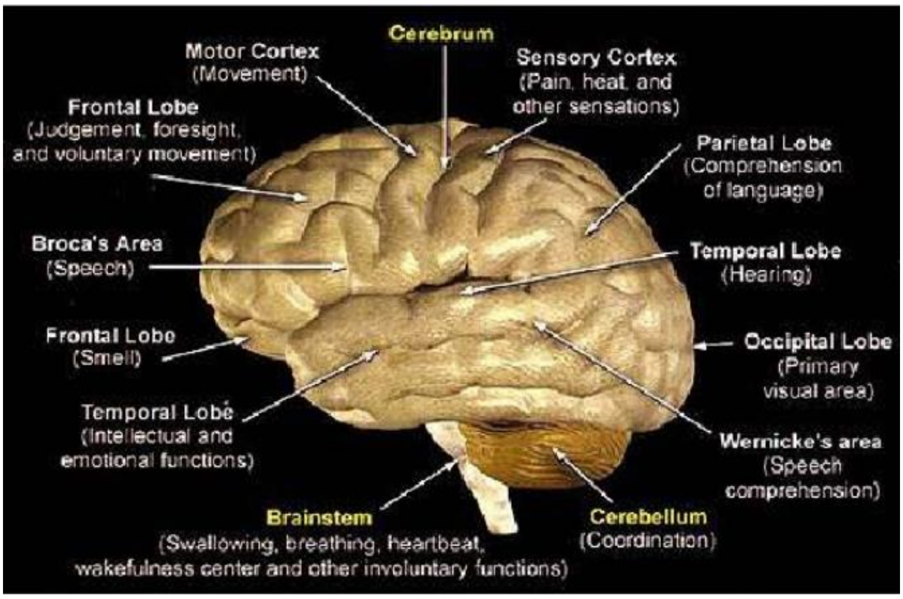

Figure 2.1: Brain's Anatomy [1]

2.1.2 Electroencephalography

The Electroencephalography (EEG) reflects the electrical activity produced by human brain. The electrical activity of human brain which is recorded by EEG is one of major tool to check whether brain is properly functioning or there exist some neurological disorder such as stroke, epilepsy, brain tumor, dementia, head injury, sleep disorder and as well as to monitor anesthesia depth given during operations. So, EEG is readily available evidence to check brain function with time. The electrical activity of brain is also suggested for the treatment of abnormalities, learning problems, change in behavior also called as Autism, gaze concentration, delay in speaking etc.

Hans Berger for the first time recorded EEG signals. He observed different behavior of brain potential produced in the form of waves. These rhythmic changes in brain waves vary with the state of mind. A number of electrodes are temporarily glued on human skull at different locations for the recording of EEG signals. An individual electrode in return is connected to amplifier and EEG recording machine. Finally, the result of recording is electrical signals in form of waves displayed on to the computer screen. According to need multiple electrodes can be placed on human skull for EEG recording. These electrodes are in range of 1 to 256 and are placed in parallel and is called multichannel EEG recording, where one channel is made by pair of electrodes placed parallel. Each channel yields a signal during an EEG recording.

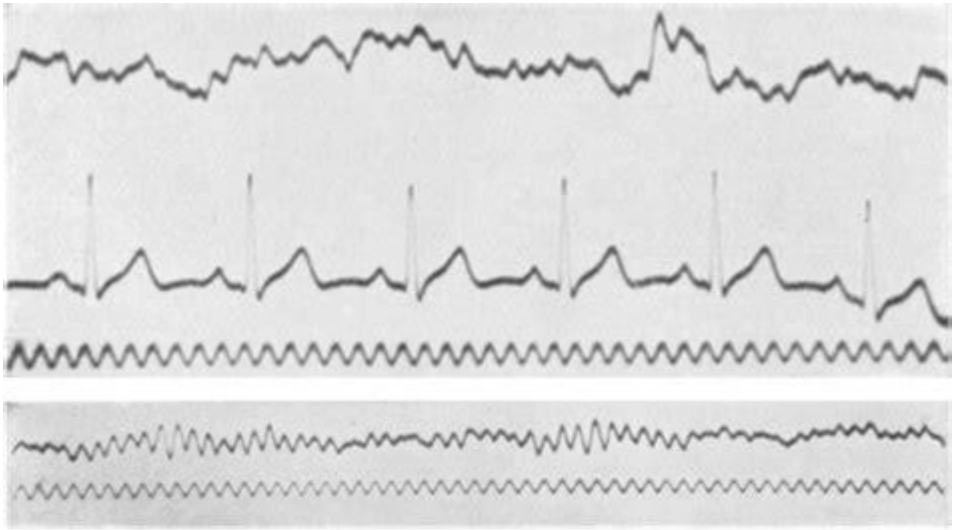

Figure 2.2: First EEG signal recorded by Hans Berger (Berger, 1929).

The EEG recording can be taken in two different ways depending on position of electrodes placed on human head: these are scalp and intracranial EEG. In Scalp EEG, the electrodes are glued on scalp temporarily with good electrical and mechanical support. In intracranial EEG special kind of sensors are placed in the brain by surgery. The EEG recorded from the cortical surface directly by subdural electrodes is named as the electrocardiogram (ECoG). In adult the amplitude of EEG signal varies from 1 to 100 µV. while measuring with subdural electrodes it usually ranges from 10 to 20mV. The EEG recordings vary with time and with location because of the non-uniform structure of scalp.

There are several techniques for electrodes localization on scalp but the standard and most renowned method is international 10-20 electrode system (Jasper, 1958a). The meaning of 10 and 20 is actually the distance between the adjacent electrodes can either be 10% or 20% of the total distance between right and left or front and back of skull. The positions of electrode are determined by; nasion and inion, in which the former is the point between forehead and the nose, and later is the skeletal prominence at the base skull at the back of the head.

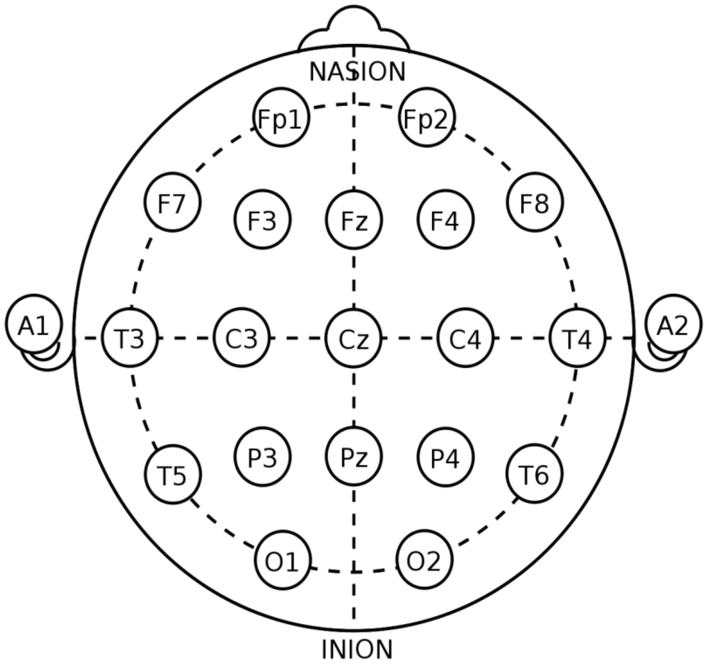

Figure 2.3: Electrode position in international 10 20 systems [3]

In figure 2.3 electrodes are placed on the skull according to the international 10-20 system. The letters used in figure represents lobe such as F for frontal lobe of brain, T for temporal lobe, similarly for central, Parietal and occipital respectively. While, the numbers associated with each letter identifies the location of hemisphere. In the following section, most common rhythm of EEG signals is discussed along with their functionality.

2.1.3 EEG signals rhythm

Frequency is the key element to study abnormalities in a given EEG pattern and find the reasons for behavior which is displayed by individuals with cognitive impairment. The electrical activity of human EEG is demonstrated as aperiodic random oscillations with irregular bursts of fluctuations which are normally characterized in definite bands. Table 2.1 demonstrates these EEG bands.

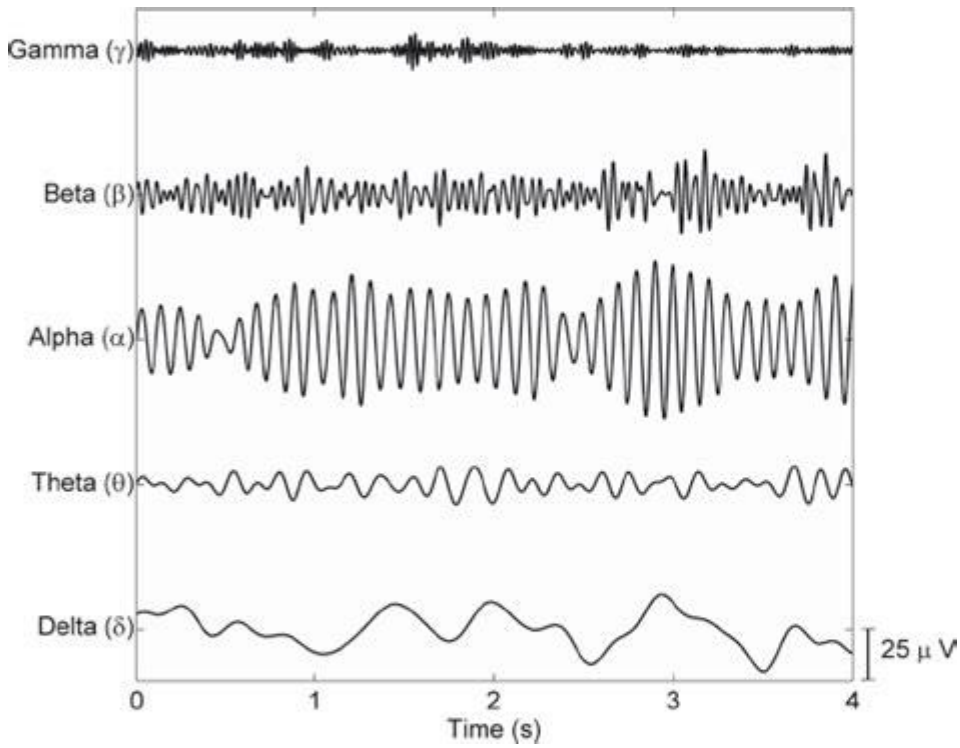

Figure 2.4: Rhythm of EEG signals

Delta wave is observed as a high amplitude wave but the slowest one. The operational frequency of delta wave exists in the range 0.5 to 4. It is primarily recorded for deep sleep and other serious brain disorders. It controls many unconscious functions and also makes us able to feel fresh when we are awake after sleep. During healing process this frequency of brain releases such hormones that the undersigned disorder healed.

Theta wave lies between 4 and 8 Hz and they are slow waves with amplitude typically greater than 20 µV. when our bran produces theta waves we fell calm and relaxed. One can feel creativity in his/her self when brain releases theta waves.

Alpha wave lies in the frequency range of 8 to 13 Hz, with 30-50m µV amplitude. Mainly, this wave produced n occipital region of brain and it is assumed that these waves give strongest signal when eyes are closed or in relaxed state. These waves are also called mu rhythm. Alpha rhythm is also called mu activity. These waves are also activated when a person is in stress or tension.

Beta wave the motor imagery tasks are done when brain produces beta wave. The operating frequency of beta wave is 13 Hz to 30 Hz. Beta waves are symmetric across both sides of frontal area. Its frequency varies and the amplitude s quit low. Beta is the brain wave mainly focuses on concrete activities and attentions.

Gamma waves have the frequency from 30 Hz and up. This rhythm is sometimes defined as having a maximal frequency around 80 Hz or 100 Hz. It is associated with various cognitive and motor functions. Electrical signals in the EEG that are originated from non-cerebral origin are called artifacts. EEG data is almost always contaminated by such artifacts.

Table 2.1 EEG Band

EEG BAND	FREQUENCY RANGE	LOCATION
DELTA	0-4 HZ	Frontal Lobe
THETA	4-7 HZ	Midline Temp
ALPHA	8-13 HZ	Frontal Occipital
BETA	13-30 HZ	Frontal Central
GAMMA	30-100HZ	Parietal Lobe

2.1.4 Concept of Brain Computer Interfaces (BCIs)

BCI technologies provide a direct interface between a brain and a computer (Vaughan et al., 2003). The ultimate object of a BCI is to provide humans an alternative communication channel, allowing direct transmission of messages from the brain by analyzing the brain's mental activities. A BCI is a technology that finds a new communicative way of using only the brain to command machines. An electrode cap is placed on the head of a user for measuring EEG signals. To command machines a user imagines a specific task such as movement of limbs, composing of words. These tasks affect the patterns of EEG signals. Computers detect and classify these patterns into different tasks in order to control a computer application (such as cursor movement), or control a machine (e.g. wheelchair). BCIs can also provide the communication and control for other user groups and goals, such as patients with less severe motor disabilities who wish to control an orthosis or wheelchair [12], and healthy users in situations where conventional means of communication are difficult, impractical or inadequate (Allison et al., 2007). BCIs could also help reduce symptoms resulting from stroke, autism, emotional and attention disorder [13]. There are two types of BCIs: invasive, which are based on signals recorded from electrodes implanted over the brain cortex (requiring surgery), and non-invasive, based on signals recorded from electrodes placed on the scalp (outside the head). In recent research, the non-invasive EEG is the most preferable technique.

2.1.4.1 BCI systems Architecture

BCI system typically consists of following six steps: EEG measurement, pre-processing of signal, feature extraction and reduction, classification, interpretation of command and feedback. A typical BCI system architecture is shown in figure 2.3.

1. EEG measurement: The electrical signals of brain are controlled by human intentions and to record these signals effectively is a crucial step for BCI communication. These signals are recorded through various types of electrodes and are converted to digital signals for further processing. In this dissertation, the electrical signals of brain s referred to as EEG signals.

2. Pre-processing of signal: Once the EEG is recorded accurately the next step is to put the signals for further operations desired to make the signals artifact free and improving signals to noise ratio without loss off any content rather boosting the relevant information which is embedded in the signal.

3. Feature extraction: There are certain features of each signal which is recorded. These features or values are extracted from the signal and leaving all the irrelevant features, only taking into account those features which are part of the brain activity.

4. Signal Classification: Once the relevant features are extracted from the signal, the next step is to assign particular class to these features to which they belong. this is also called feature translation as they are translated to a unique class. This classification strictly based on the type of mental condition identified

5. Command Interpretation: when the exact mental state is diagnosed such as epilepsy, stroke, dementia; a specific command is linked with the condition identified and based on command the BCI is controlled..

6. Feedback: in order to achieve good performance of BCI, here in this step user will give their feedback about their disability.

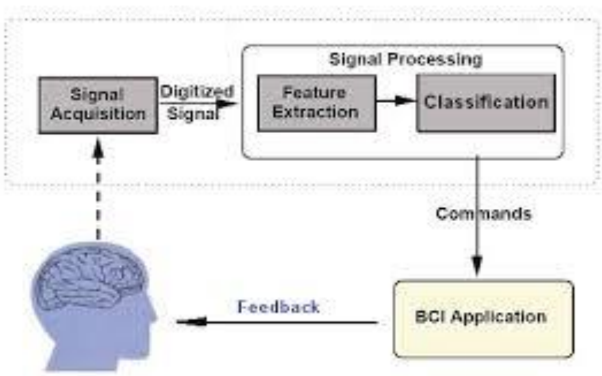

Figure 2.3: A generic BCI Architecture

The most common mental strategies are motor imagery (MI) and selective (focused) attention. Motor imagery (MI) is the imagination of a movement without actually performing the movement. On the other hand, BCIs based on selective attention require external stimuli provided by a BCI system. The stimuli can be auditory or somatosensory. In this research, we work on the MI for the BCI systems.

2.2 Pre-processing of EEG Signal

When EEG signals are collected through procurement apparatus then there are some artefacts and noises such as heart beats, eye blinks and other disturbed factors are commonly present in them. To remove such interrupts pre-processing should be done by using following well known filtering techniques.

A. CSP

To operate motor imagery based BCI Common spatial pattern technique is used. EEG signals pattern detection is done through spatial filters due to this variance of each task is either increased or decreased with respect to other task. But, multiple electrodes must be present to utilize this method in an efficient way and similarly by varying the position of electrode results of classification can be enhanced [15].

B. PCA

Principal Component Analysis (PCA) is used for analyzing and dimension reduction of data without losing information. It uses mathematical procedures that utilize orthogonal transformation. These transformations change a set of correlated vectors into a set of linearly uncorrelated vectors called principal components [15].

C. CAR

Common Average Referencing is a spatial filtering technique that removes the noise from position of interest. This noise can be considered as common activity of EEG signal [16]. The artefacts present in the EEG signal causes low signal to noise ratio which can be improved by referencing approaches. Mean of all electrodes will be eliminated in this approach to produce a noise free signal. The classification done by car has good accuracy result in all referencing methods. In common referencing methods problem can be caused in the calculation of the average of electrode due to “finite sample density” and inadequate head coverage of EEG [15].

D. SL

Surface Laplacian of the skull is the estimation of the present density ingoing or departing the scalp via skull. This approach is based on the actual spatial which is used in BCI studies. SL technique can be used for the identification of sources and for the improvement of signal to noise ration of signal [17]. EEG signal can be observed with high spatial resolution using this technique. Electrode cap generates some artefacts at the uncovered points which can be tackled by this method.SL is robust against problems and can solve the electrode referencing problems. The limitation of the method is that when the spline interpolation is carried out it is excessively sensitive to the choice of spline parameters [15].

E. Adaptive Filtering

Adaptive filtering is a filtering technique that models the relationship between two signals iteratively. Adaptive filtering consists of four parts: the input signal diluted with noise, the relationship between input and output signal, the parameter that varies with respect to the input and output of filters. A new adaptive filtering design which outperform for sensorimotor rhythms (SMR) based BCI has been introduced in [18] which are known as adaptive Laplacian (ALAP) filter. An ALAP filter works as a Gaussian filter which construct a smooth spatial gradient of channel weights. The optimal kernel radius and regularization parameters of spatial filter are also obtained in this technique. This optimization is based on gradient descent method which minimizes the leave-one-out cross-validation error. The benefit of adaptive filters is that it works well for the signals with overlapping spectra and is able to modify signal features according to signal being analyzed [15].

F. ICA

Noise from the EEG signals is split into independent components using ICA i.e. Independent Component Analysis on the basis of data characteristic without depending on the reference channel. Multi-channel EEG data decomposed into two components which are spatial- fixed and other is temporal independent using ICA algorithm in. This algorithm is computationally efficient and has capability for the particular class of signals to remove the artefacts and has ability of source extraction. ICA has better performance than PCA when the data we want to decompose is larger in size [15].

G. SVD

SVD is the more generalized form of the eigenvalue decomposition (EVD). SVD is the most stable process to divide the system into linearly independent set of modules which have their specific energy influence. SVD has countless benefits, it is capable to change the image in base of two unique subspaces which are noise subspace and data subspace that are used for the filtering of noise and it has maximum energy packing advantage too [19].

H. Digital Filters

In artefact handling, specifically for the removal of muscle artefacts digital filters like band-pass, low pass and high pass which are frequency selective are used. This has the limitation that this technique needs EEG signal and artefact that inhabit distinctive frequency bands which do not exist in reality [17]. Band pass and notch filters are the commonly used digital filters used in EEG signal [16]. Noise produced in electrical line which is caused by the grounding of EEG can be easily removed by the notch filter [20].

In Table 2.2 comparison of the pre-processing approaches are shown.

Table 2.2 list of pre-processing approaches

No.	Method	Advantages	Disadvantages
1.	CSP	Motor Imagery data can be handled.	Numerous electrodes required.
2.	PCA	Dimension of feature can be reduced.	Less efficient than ICA.
3.	CAR	Performs well as compared to other reference methods.	Needs adequate analysis.
4.	SL	Vigorous against the artefacts at bare conductor.	Artefacts sensitive. Spline pattern sensitive.
5.	ICA	Computationally proficient. Large data can be handled.	Computationally complex for decomposition.
6.	SVD	Stable and Maximum energy packing.	
7.	Adaptive filtering	Overlapped spectrum can be handled.	Require binary signals (reference signal included).
8.	Digital Filters	Easily Noise removal.	Distinct Frequency band Required for EEG signal and artefact.

2.3 Feature extraction method

Using different techniques like Independent Component Analysis (ICA), Fast Fourier Transform (FFT), Wavelet Transformations (WT), Auto Regressive (AR), Principal Component Analysis (PCA) and Wavelet Packet Decomposition (WPD) useful features can be extracted from the pre-processed EEG signal [19] . All techniques are listed in Table 2.3

A. PCA

In the preprocessing method “Principal Component Analysis” (PCA) is used, and it is also used in features extraction stage .Without loss of information data can be analyzed and dimensions of data can be reduced using this technique. The information at different time series multi- channel can be extracted using this technique. Principle components can be produced and signal dimension can be reduced by the artefacts removal.

PCA feature extraction, has advantage over the other extraction techniques that the performance can be improved [21]. The limitation of the technique is that it assume the data is linear and continuous and it is not feasible for the processing of data with complicated set of features [19].

B. ICA

Independent component analysis is one of the best techniques for pre-processing of EEG signals and for feature extraction. There are different components of ICA which are totally independent of each other and through these components features are extracted .Blind source separation is the best example of ICA.it is developed using ICA and it is also used to identify signals and to separate noise from brain signals. ICA is computationally very efficient and it is also very helpful to calculate high performance for EEG signals [18].

C. WT

WT is Wavelet Transform and it is a mathematical technique used on a very large scale for separating information from different types of continuous data like images and speech. Non-stationary signals are processed by using Wavelet transform technique because of representation of signals in time-frequency domain [21]. There are various advantages of WT e.g. analysis of signal with discontinuity using variable window size, localization of signal in frequency and time domain, Extraction of energy, cluster or distance but one of the major pitfall is the lack of suitable method to the pervasive noise application [18].

D. WPD

WPD is Wavelet Packet Decomposition (WPD), it is the advance form of WD (Wavelet decomposition).WD is different from WPD in some manners, as WD is used to get approximation and all the information in details at higher stage by breaking down approximation at each separate resolution stage. Moreover, when comparison is being done, it is observed that a precise and efficient frequency resolution is obtained by WPD technique instead of WD [22]. By using WT coefficients mean, WPD is able to get features in both domains i.e. frequency and time. WPD gives better results in extraction method in case of non-stationary signals but it can cause computation time to increase [18].

E. Parametric Approach

Parametric approaches include auto regressive (AR) method. One of the best usages of AR is to process EEG signals through signal representation of linear combination at previous levels at respective separate channel [23]. It is helpful in time domain for feature extraction technique. To increase best frequency resolution and to minimize spectral loss, AR is used with reduced data records time duration. When input parameters are AR type like in non-stationary EEG signals then this approach gives the best results. There are some disadvantages too in this approach one of them is the establishment of AR parameter model characteristic [18].

F. FFT

FFT is Fast Fourier Transform (FFT). FFT applies mathematical technique to EEG data. Power spectral density (PSD) estimation method is used to calculate characteristics of the obtained EEG signals which have to be analyzed for the purpose of EEG sample signals representation. Signal's features are extracted by converting the signal from time domain to frequency domain. Linear random process and stationary signals give better performance by using FFT. However, non-stationary signals are not processed by FFT and it doesn't compute both time and frequency [18].

G. STFT

The STFT uses FFT to analyses signal in small contents, which is achieved by dividing a signal into short consecutive segments and then computing the Fourier coefficient of each segment. But still, it is a compromise between the time resolution and frequency resolution during the procedure. Whether imprecise the time or the frequency representation, is determined by the window size. The Fourier transform can represent the frequency localization, but lost all information in time domain. While, short-time Fourier transform, as an evolved scheme of Fourier transform, has equal-length intervals in the time domain. STFT has its limit to analyze the signals with low frequencies, because of low-frequency signals take a long time interval to get the whole shape of one period, and if the length of window function cannot cover the whole period, it cannot get significance to estimate the wave-shape of whole period [24].

Table 2.3 Feature extraction Techniques

No.	Method	Advantages	Disadvantages
1.	PCA	Performs Dimensionality reduction without any loss of data.	Unable to process complicated set of data
2.	ICA	Performs well for large sized data. Low computations.	Decomposition of data requires high more computations.
3.	WT	Analyzes time and frequency aspect of signal. Particularly Fit for nonstationary signals.	No specific method to remove prevalent noise.
4.	WPD	Performs analysis of nonstationary signals.	Computational time increases
5.	Parametric approaches	Time localization. Frequency localization without any spectra loses.	Not applicable to stationary signals
6.	FFT	Performs well for stationary signals	Doesn't performs for non-stationary signals Unable to localize both time and frequency
7.	STFT	Give both time and frequency localization	Difficult to select appropriate window size

2.4 Overview of EEG signal Classification

The classification of EEG signals presents several challenges that make it a uniquely difficult problem in machine learning. The EEG signal is high dimensional, with both spatial and temporal covariance. The high dimensionality is difficult to account for in traditional machine learning techniques, leading to a desire to extract features from the signal to reduce the dimensionality. However, many time series approaches for feature extraction face issues with the non-stationary nature of the EEG signal. A stationary signal is one in which the probability distribution does not change over the course of time, and, thus, features like mean and variance will not change. While data can be transformed to account for non-stationarity, it is an imperfect solution. The signals are also very noisy, being susceptible to factors such as physical movement, mood, posture, and external noise [25].

Another issue faced in the field is a lack of comparability between experiments. Unlike in image classification, there are no standard datasets used as performance benchmarks. Not only is data collected on machines with differing specifications, and individuals with differences in physiology, but entirely different tasks are also performed during data collection leading to different target brain activities. Furthermore, some approaches use models for individuals, whereas others attempt to make a universal model, training and testing with samples from all individuals at one time. In general, the focus has been on the use of machine learning as a tool to draw conclusions in neuroscience, rather than improving the techniques of classification. Thus, while this section will explore techniques used, it will not focus on accuracy of the given techniques, since it is nearly impossible to compare them at the present

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

The extracted features are fed to some classifier so that similar extracted features can be placed in one class.

A. Traditional Approaches

The majority of published techniques currently use some form of classification techniques with a relatively simple model. These techniques require less implementation time than deep learning, and are less prone to overfitting. These traditional approaches may include following classifiers [25].

1) LDA: Linear Discriminant Analysis is a classification technique which produces models that have quite good accuracy as compared to the complex models. It has very low mathematical requirements. It creates linear combination of variables that best splits two classes and that linear combination of variables are presented in form of model of the Probability density function [25]. The complex structure of the data may not preserve if the data is non Gaussian distribution. If the discriminatory function is based on variance of the data instead of mean LDA fails it performance [26].

2) SVM: Support vector machine is a linear classifier and one of the exciting algorithms in machine learning and used by most of the BCI. It was first introduced by Vapnik and was determined by principal of statistical learning theory and structural risk minimization [25]. SVM searches a line that creates a clear gap among the dataset and split it into designated category. The line which splits the data is called hyperplane, it capitalizes the distance between the hyper plane and the points that are nearer from each class that are called as support vectors. The motivation for this method is to reduce the complexity of the learned model and improving performance [25]. Due to this, SVM provides good generalization.

3) NBC: The naïve base classifier produces nonlinear decision boundary. It is based on Bayesian theorem [27] through which posterior probabilities of feature vectors are calculated and assigns the features vector to its class with a highest posterior probability. They often outperform well as compared to the discriminative classifiers and reject uncertain samples more accurately. Hidden Markov Model

(HMM) is a Bayesian classifier and it is suitable for the classification of time series. These are not as prevalent as Linear Classifiers in the field of BCI applications . They are suitable when the dimensions are high.

4) NNC: Nearest neighbor classifiers is a non-parametric method of classification, it stores all patterns first and then search for the most nearest neighbor of the pattern and assign to it a class. k-Nearest Neighbor (k-NN) classifiers is the vibrant of the NNC. If the feature vector is from the training set then it is called as k-Nearest Neighbor classifiers. It assigns the class membership to the feature vector based on k closest patterns in the feature space. It assigns the label of a test sample with the majority label of its k-nearest neighbors from the training set. K-NN is transparent and easy to understand, and its implementation is quite easy as compared to other techniques [28].

5) ANN: Artificial Neural Network is a nonlinear classification approach and is composed of large number of connected cells called neurons. Multi-Layer Perceptron Neural Network (MLPNN) is the most frequently used neural network, there are three layers in MLPN viz., input layer, hidden layer and output layer. The input to the first layer is actually denotes the number of selected features and the output layer shows the number of labels or classes. The AAN will be complex if the hidden neurons are complex. The hidden layer also defines the accuracy of classification. If less number of hidden layers is used there might be a chance for classification error. There is no specific criterion for selection of hidden neurons it is selected by hit and trial method [25] [29].

B. Deep Learning Approaches

Deep learning is a subfield of machine learning that has evolved out of the traditional approaches to artificial neural networks. With the advent of backpropagation, neural networks began to see renewed interest and significant theoretical advancement in the form of recurrent neural networks (RNN), Convolutional neural networks (CNN), Long Short-Term Memory networks (LSTM) and Gradient decent (SGD).

1) CNN: Convolutional neural networks learn filter banks that are convolved with the original data. The filters can also be represented as a fully connected layer where the weights of the edges are tied together in way that replicates the convolution operation. This weight sharing structure allows for fewer parameters than having each weight be unique, and directly accounts for structure in the data. Each filter creates a new, processed version of the image.

2) RNN: Recurrent neural networks contributed a lot to the success of deep learning in various fields. It outperforms for time series classification because of its intra layer connections and inters layer connections. Training an RNN is a difficult task. Backpropagation must be modified to function in RNNs, since there are cycles in the graph. This is frequently handled through a technique known as backpropagation through time, wherein the network is "unrolled" for a discrete number of steps. This process creates a network with only forward connections, allowing backpropagation to work as normal, at the cost of limiting the impact of the recurrent connections.

3) LSTM: Long Short-Term Memory networks have proven successful for sequential data classification because they are specifically designed to classify for each time stamp. At each time point data is processed by individual LSTM unit. The processed result is fed to the next layer and as well as well as remains within the same layer for the processing of next time stamp. The information passed through to the next layer is passed through an activation function [30]. However, the recurrent connection within the layer is not subjected to an activation function. One of the vibrant problems in LSTM is the lack of activation function in recurrent layers to avoid the vanishing gradient problem. Due to this gradient problem it gives gradient a constant value of one. Each LSTM unit has a number of gates which control the flow of information. A gate is a combination of a sigmoidal activation unit and pointwise multiplication. The flow of information from one time stamp to another time stamp is controlled by these gates.

4) SGD: It is a derivative of traditional gradient descent, differing in that the error function is calculated using only a subsample of the available data. This is both easier to use and more efficient for training datasets that do not fit in memory. Furthermore, adding randomness to the optimization can aid in breaking through plateaus and avoiding local minima. The addition of momentum terms, which biases the gradient in the direction of recently calculated gradients, greatly improved the ability to train deep models by further increasing speed of convergence [22].

2.5 Summary

This chapter provides background of EEG signals , different types of EEG waves based upon frequencies which is considered as a main key to classify EEG signals waves. Several pre-processing EEG feature extraction techniques has been discussed here among which a best one is chosen to apply in this dissertation. This chapter also includes concept of classification and different classification techniques used in literature but without giving their accuracies. It is also shown that each technique have its limitation . hence there is a need to develop classification algorithm.

CHAPTER 3

Problem Definition and Approach

3.1 Problem Definition

Simply stated, the problem explored in this thesis is the classification of EEG data using deep learning techniques. While the classification of EEG signals can be useful in many areas, such as the detection of disease state or brain computer interfaces. Most of BCIs are based on classification of imagination of different parts of the body; these are left and right hand, left and right leg etc. In this thesis we present the classification accuracy of left and right hand imagery movements based on the evoked brain activity. These sorts of stimuli paradigms are common in psychology and neuroscience for both basic research on perception and memory, and also as datasets for technique validation. The problem of EEG classification in these types of experiments has a number of challenges that must be considered, including:

- There is high variability between subjects and within subjects,
- There is limited availability of data, and
- The data is composed of multiple channels of time series information

This thesis will address these challenges by exploring several variations of architecture selection, model search, regularization, and training paradigms within a deep learning context, with the aim of harnessing the expressive power of deep learning while avoiding overfitting.

3.1.1 High Variability

The highly variable nature of EEG data leads to difficulty in classification. Samples in the same class may be very different in nature from one another. There are multiple sources of variability both within subjects and between subjects. Sources of within subject variability are briefly summarized below:

Level of attention: If the subject is not focused on the task, the patterns of activity produced by their brain will be quite different. This not only impacts data quality, but also reduces the amount of data that can be collected.

Multiple Sessions: Many experiments call for the subject to attend multiple sessions of data collection. The subject can be more or less focused, and may be in a different mood, both of which affect the data. Another common source of variability is the physical placement of the EEG electrodes. It can be very difficult to insure they are in the exact same location. Thus, channels may be collecting data from a slightly different source from one trial to the next.

Muscle movements: A constant issue in EEG data analysis is that the electrical activity created from muscle movement has a far higher magnitude than that produced by brain activity. This is particularly noticeable and problematic when the subject blinks. There are methods to reduce the impact of certain types of movements, including eye blinks, such as using independent component analysis to regress out the artifact. However, these methods are imperfect.

Machine noise: There is a certain amount of uncertainty inherent from the machine itself. Slow drifts in the data are common and are caused by either slight movements of the electrode or sweat interfering with the sensor. Movement in wires connecting the electrodes can cause similar issues. These issues can often be mitigated through the use of band pass filters, however. Sources of between subject variability are briefly described below:

Differing physiologies: Differences in skull shape can lead to electrodes monitoring different relative portions of the brain. Thus, the brain activity collected by an electrode on one subject may be from a slightly different region than the activity collected by the same sensor on a different subject.

Differing cognitive patterns: There are individual variations in how information is processed, and thus, the same stimulus may illicit differing responses in different subjects.

Differing behavior: Some subjects will be more focused on the task than others. Some will perform better than others on the experimental task. Differing behavior is associated with differences in brain activity patterns.

3.1.2 Limitations of Dataset

As previously discussed, deep learning requires a large amount of training samples to prevent over-fitting due to the immense number of parameters in the model. Most successful applications have tens of thousands to millions of samples. However, collecting this amount of data for EEG tasks is at best difficult, and, frequently, intractable. First and foremost, there is no way to automate the data collection. Proper EEG recording requires trained experts and takes considerable setup time. The electrodes must be properly connected and the subject must be observed to make sure they are participating in the experiment correctly. Thus, it places significant time demands on the expert during data collection, limiting the number of subjects for a study. In addition, individual subject can only be expected to focus on a task for limited periods of time. As mentioned in the previous section, as attention wanes, brain activity changes significantly. Thus, data collection sessions are strongly constrained in their durations. Scheduling individual subjects for multiple sessions poses additional challenges. It may be difficult for some subjects to participate multiple times due to time constraints; it may be impossible for many subjects due to health reasons. It is also time consuming to set up the sensors multiple times. Finally, as mentioned previously, brain activity in a subject can differ from session to session, so collecting samples over multiple samples is at best an imperfect compromise

3.1.3 Multiple Channels and Time Series

EEG data is recorded as a time series of floating point numbers at 250-1000 Hz, generally using 32 to 256 channels. Each sample is usually between half a second and five seconds long. These factors lead to several important implications that make classification task harder. First and foremost is that the data is very high dimensional. In the case of an EEG with a small number of electrodes (e.g. 32) recorded for only a second, there are still 30,000 or more features in a single sample. Given that only a small number of overall samples that can be collected (hundreds per class per subject, generally), the curse of dimensionality is a real challenge. Secondly, the data is a time series in nature. This means that the value at a specific point in the signal isn't as important as the pattern of values, in general. Furthermore, there can be small variations in the onset of the pattern. So, what happens at time point 10 in one sample may not occur until time point 15 in another sample. Many models have difficulty classifying data of this nature. Finally, the information is distributed over multiple channels. The important information to distinguish between two classes may not only lie in the patterns over one channel, but rather, the joint patterns of activity over multiple channels. This is again, very difficult for many types of models to efficiently incorporate.

3.2 Approach

In order to address the challenges presented in classifying EEG data, First, we establish a baseline using traditional methods. Second, we perform a model search and hyperparameter search among deep learning architectures. The flow chart for measurements and analysis of EEG signals are shown in figure 3.1 below.

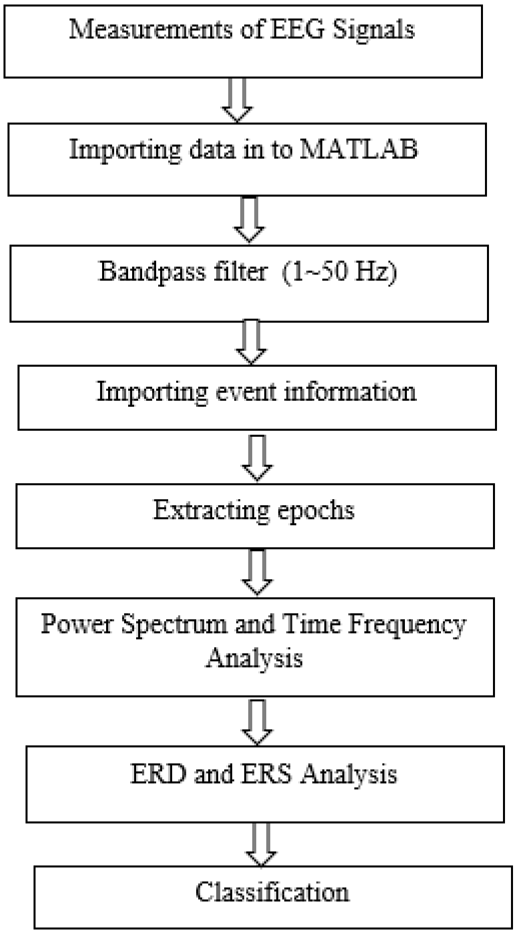

Figure 3.1: A Flow chart of measurements and Analysis

3.2.1 Traditional Baseline Approach

Given the highly diverse nature of EEG datasets, it is important to establish a baseline using traditional methods on the particular data we are classifying. A strong classification performance on one dataset, may only be a mediocre performance on another. We also explore the use of feature extraction in establishing a baseline. While wavelets and Fourier techniques are very common in literature, we consistently found them to underperform in similar datasets. Short-time Fourier Transform (STFT) analysis is one of the techniques to reveal the frequency contents of the EEG signals at each time point. This information can be used to provide control and perform several tasks using Brain Computer Interface system (BCI). This dissertation describes the STFT analysis of EEG signals obtained during left and right hand. The EEG signal was filtered using a bandpass filter with a frequency range of 8-30 Hz before the STFT was computed. The width of the Hamming window used in the STFT computation was varied to obtain the optimum width. The results of the STFT analysis showed that the frequency components of the EEG signals obtained depicts that there is a big difference between patterns of right hand movement and left hand movement. So these patterns are actually features which we extracted using Short Time Fourier transform.

As deep learning is considered a form of representation learning, learning its own features in the lower layers of the network to classified in the terminal layers, we also explore not using any explicit method of feature extraction for comparability (LeCun, Bengio, & Hinton, 2015).

3.2.2 Deep Learning Model

An immense number of diverse deep learning architectures exist. In particular, we examine convolutional neural network (CNN). A hyperparameter is a value set before the learning algorithm begins. These include everything from the activation function chosen for a given neuron to the number of neurons in a layer to the learning rate and choice of optimization function. The weight of an edge between two neurons is not a hyperparameter, since that value is learned. Since the number of network configurations grows with the product of the choices per hyperparameter, the space is combinatorically large. Grid search has been used in similar problems historically, however has proven to be ineffective in deep learning due to the size of the hyperparameter space and time required to evaluate a single configuration. Random search has proven to be more effective (Bergstra & Bengio, 2012), but is unappealing due to the loss of interpretability and the requirement to set bounds for the hyperparameters a priori. Thus, we use a manual hyperparameter search using several heuristics.

Figure 3.2 provides a flow chart of the basic logic used in the heuristic model search. After training the model with a given set hyperparameters, the training accuracy is examined first. If the training accuracy is low, then the number of parameters in the model was increased. Since classification performance is potentially limited by both the dataset itself and the basic type of model being explored (e.g., an CNN), the definition of low must be relative and determined separately for each type of model.

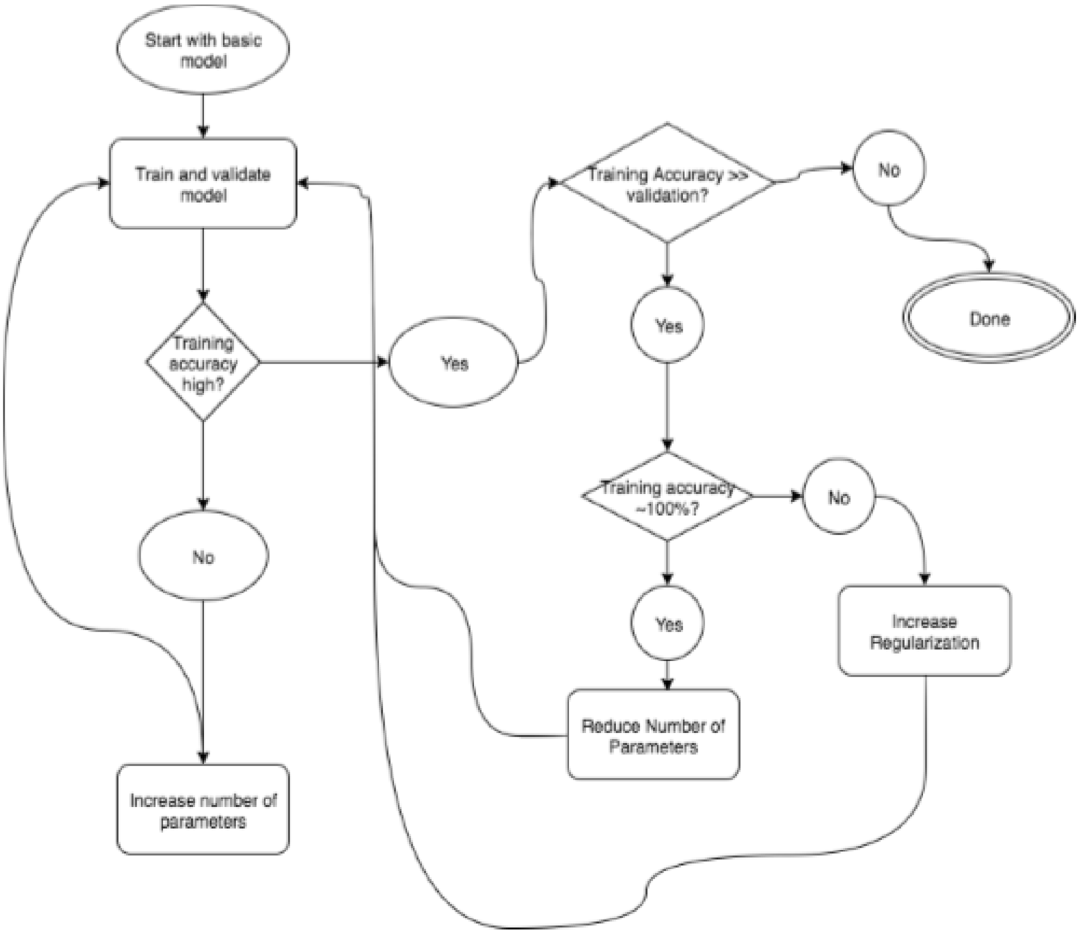

Figure 3.2: Flow chart of hyperparametric search

Generally, to increase the number of parameters in the model, either more neurons are added to layers or more layers are added to the model. Other actions are also explored at this step, including changing activation functions and adjusting the learning algorithm, including changing the learning rate or the algorithm used. Those actions are rare compared to simply increasing parameters, however. If training accuracy is high, then the validation accuracy and test accuracies are considered. If the validation and test accuracies are also high, then the selection of hyperparameters can be accepted, or more tweaks can be made to the hyperparameters to determine if the working definition of high accuracy is sufficient. If the test and validation accuracies are low when the training accuracy is high, the model is most likely overfitting. There are two main ways of handling overfitting: increasing the amount of regularization in the model and decreasing the number of parameters in the model. Increasing the amount of regularization includes changes such as adding or increasing L2 penalties on weights, adding or increasing dropout, or adding batch normalization to layers. Reducing the number of parameters is achieved by either reducing the number of neurons in the layers or reducing the number of layers. Generally, reducing the parameters of the model is reserved for when increasing regularization fails to improve validation accuracy to avoid unnecessarily reducing the expressive power of the model. However, if the model performs nearly perfectly on the training data while performing very poorly on the validation data, reducing the number of parameters can be the most effective change to the hyperparameters.

None of the models were able to converge reliably on the mentioned dataset. In CNN with the network depth increasing, accuracy gets saturated and then degrades rapidly

The advent of more powerful graphics cards and more convenient programming for those cards: In 2007 NVIDIA debuted the Compute Unified Device Architecture (CUDA) API to that enables direct and easy programming for their graphical processing units (GPUs). GPUs are specialized processors designed particularly for graphical display. Since most graphical applications rely on large amounts of calculations that can be run in parallel, these processors have a large amount of cores. Over the next several years, CUDA gained popularity for use in highly parallelizable tasks able to take advantage of the large number of cores in GPUs. Deep learning is implemented as large matrix operations, which is one such task. Furthermore, the memory capacity and processing capabilities of GPUs increased dramatically from the launch of CUDA to 2012. The hardware and software advances in GPU programming were critical for [33] ImageNet. Prior to these developments, training deep neural networks was essentially intractable- even early GPU implementations showed a 70X speed up over standard CPUs

Dropout and other regularizes: Overfitting, the phenomenon of learning patterns that happen to be present in the training data by random chance and not present in general, is a constant threat in deep learning due to the immense numbers of parameters involved [31].

Regularization techniques prevent the model from becoming too complex and specific to the training data, thus reducing the tendency to overfit. L2 regularization is a standard technique used in many forms of optimization, in which the squared sum of the weights is applied as a penalty to the optimization function. In essence, this weighs the benefit of increased classification of the training data against model complexity. By preventing the model from becoming too complex, memorization of the training data is reduced, and a more generalizable model is developed.

Dropout is a technique in which a random set of neurons from each layer (generally between 30% and 60%) is excluded from both classification and updating during a training pass through the data. This effectively allows a single model to act as an ensemble, a group of classifiers that act in union to produce a classification.

Furthermore, it prevents the direct memorization of training data, since different sets of neurons will be participating from one pass to the next.

Increasing rate of data collection: More data allows for larger, more expressive models to be developed. By having a larger sample size, more variance will exist in the data, and can be directly accounted for by the model. Similarly, variance in the training data that happens by chance will have less of an effect on the model since each sample will contribute less to the overall gradient during backpropagation, leading to better generalization to new data. Deep learning shows its greatest strength in problems of very high complexity where traditional learning techniques have struggled.

CHAPTER 4

Dataset, Implementation and Results

In this chapter we describe the dataset, the specific implementation details, and the results of the model search and transfer learning tasks.

4.1 Dataset Description

To evaluate our methodology we used dataset 2b from BCI Competition 1V. The dataset consists of EEG data from 9 subjects [32]. All subjects voluntarily participated in collection of the dataset. The participant were asked to sit on armchair relaxed, and watch monitor screen which is placed at distance of 1m from the eye level. Each subject performed 5 sessions in which first two were for collecting training dataset with no feedback i.e. without feedback, and the last three dataset were recorded with feedback.

Recordings were made at a sampling frequency of 250 Hz and only three channels were considered i.e. along C3, CZ and C4. The effect of CZ is not visible in motor imagery tasks so we also ignored it while data processing. The dynamic range for recording screening data is ±100μV while it is ±50μV for the feedback data. The electrodes were placed differently for all subjects, the difference is of the distance of anterior and posterior parts [33]. The screening data mainly consist of two classes, when cue is based on motor imagery task for left hand movement and another is when cue is based on right hand movement.

The subjects participated in screening sessions without feedback performed recordings on two different days but within two weeks. Each recording consisted of six runs, and each run consist of 10 trails for each class. So, in total there are two classes which makes 20 trails per run and 120 trails per session. Before executing the trail each subject performed and imagined other body movements such as squeezing a ball or pulling a brake, but selected the left and right movement because they performed it well.

The subjects were shown a fixation cross before actual imagination take place, and a short beep used as a warning tone. Some seconds later after this fixation cross a visual cue which shows an arrow pointing either towards right or left according to the subjected class was shown to the object for 1.25 seconds. After this cue the subject had to imagine the respective hand movement for a period of 4 seconds. After imagination the subject was asked for a short break of about 1.5 seconds. Time of up to 1 second was randomly added to the break in order to avoid adaptation. The data files are stored in General Data Format for biomedical signals. one file for per subject and session. however the first three sessions contains the annotated data. All files are listed in Table 4.1.

Table 4.1 List of all Files contained in the dataset

Sr. no	Training
1.	BCI 101T, BCI102T, BCI103T
2.	BCI201T, BCI202T, BCI203T
3.	BCI301T, BCI302T, BCI303T
4.	BCI401T, BCI402T, BCI403T
5.	BCI501T, BCI502T, BCI503T
6.	BCI601T, BCI602T, BCI603T
7.	BCI701T, BCI702T, BCI703T
8.	BCI801T, BCI802T, BCI803T
9.	BCI901T, BCI902T, BCI903T

The position of an event in samples is contained in event position file. The corresponding type can be found in event type, and the duration of that particular event is stored vent duration file. The types used in this data set are described in Table 4.2. Note that the class labels (i.e., 1 and 2, corresponding to event types 769 and 770) are only provided for the training data and not for the testing data [33].

Table 4.2 List of Event Types

Event Type	Description
276	Eyes open
277	Eyes closed
768	Start of a trail
769	Cue onset Left (class 1)
770	Cue onset Right (class 2)
781	BCI Feedback
1023	Rejected trials
1077	Horizontal eye movement
1078	Vertical eye movement
1079	Eye rotation
1081	Eye blink

Figure 4.1 shows the response on channel c3 for a single trial of the task. While the pattern of activity in response to a visual stimulus is difficult to see in a single trial, it becomes visible after averaging many trials.

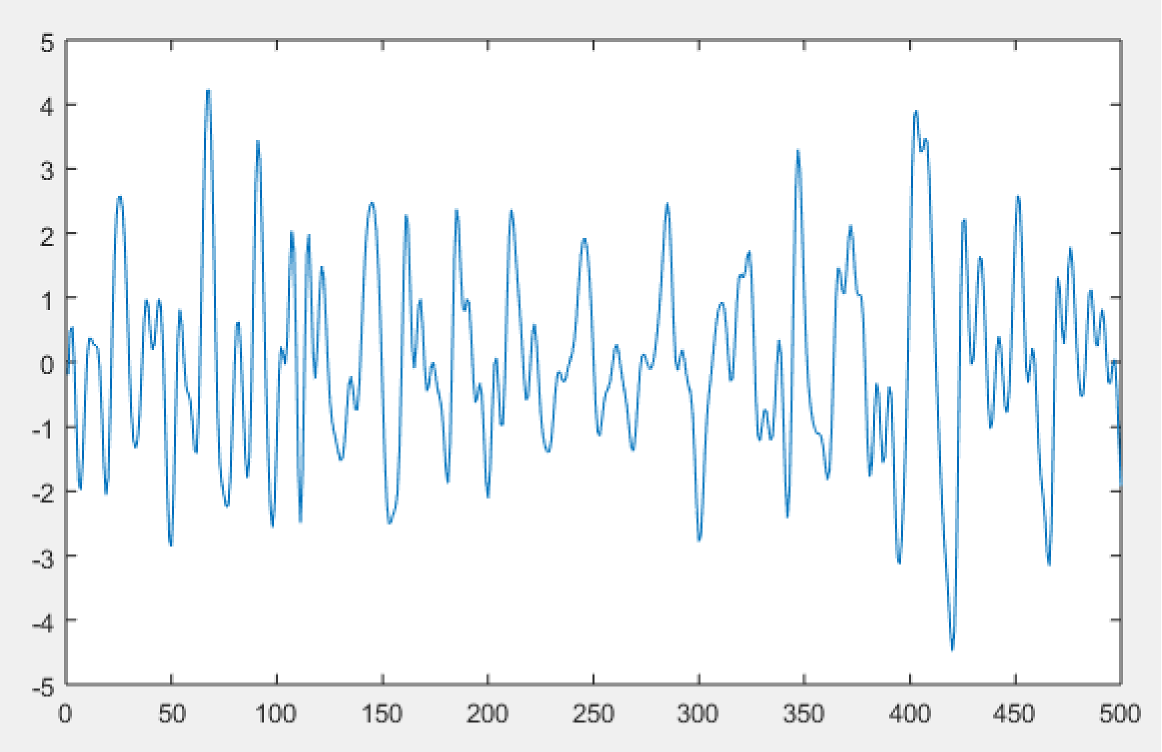

Figure4.1: Graph of single trial response on C3 channel

4.2 Implementation Details

Architecture was trained using Linux with GTX 1060 6 GB graphics card and 16 GB RAM. The primary language used was Python and MATLAB. In particular, the distribution used was the Anaconda scientific distribution of Python 3.6. NumPy and SciPy were used for the vast majority of numerical computations. Scikit learn was used for the machine learning algorithms, other than those related to deep learning. All deep learning related code was written using Keras library, with a backend of Tensorflow using CUDNN to allow for efficient GPU training.

4.2.1 Input image form

The datasets that we used in this study include recordings from three electrodes (C3, Cz and C4) during left/right hand MI task. These electrodes are located on motor area of the brain. It is shown in [5] that the energy in mu band (8–13 Hz) observed in motor cortex of the brain decreases by performing an MI task. This decrease is called event related desynchronization (ERD). An MI task also causes an energy increase in the beta band (13–30 Hz) that is called event related synchronization (ERS). Left and right hand movement MI tasks are said to cause ERD and ERS respectively in the right and left sides of the motor cortex affecting

EEG signals at C4 and C3 electrodes. Considering these facts, we designed our network input in order to take advantage of time and frequency properties of the data.

Short time Fourier transform (STFT) was applied on the time series for each 4 s long trial. In case of 250 Hz signal this is corresponding to 1000 samples. STFT was performed with window size equal to 250 and time lapses equal to 14. This leads to a 126 × 57 image where 126 and 57 are the number of samples along the frequency and time axes respectively. Then we extracted mu and beta frequency bands from the output spectrum. The frequency bands between 6–13 and 17–30 were considered to represent mu and beta bands. A sample input image constructed for a right hand , left hand MI task trial and a no movement image is also illustrated in figure . By using the proposed method, brain activations in right and left sides of the motor cortex of the brain, cause different activation patterns along the vertical cortex of the brain. The brief literature review above shows that more studies should be conducted on the changes in electroencephalographic (EEG) alpha and beta amplitudes in order to (1) ascertain predictable changes in the sensorimotor alpha and beta bands during motor imagery tasks and (2) investigate the effects of different imagined movement durations on ERD/ERS patterns.

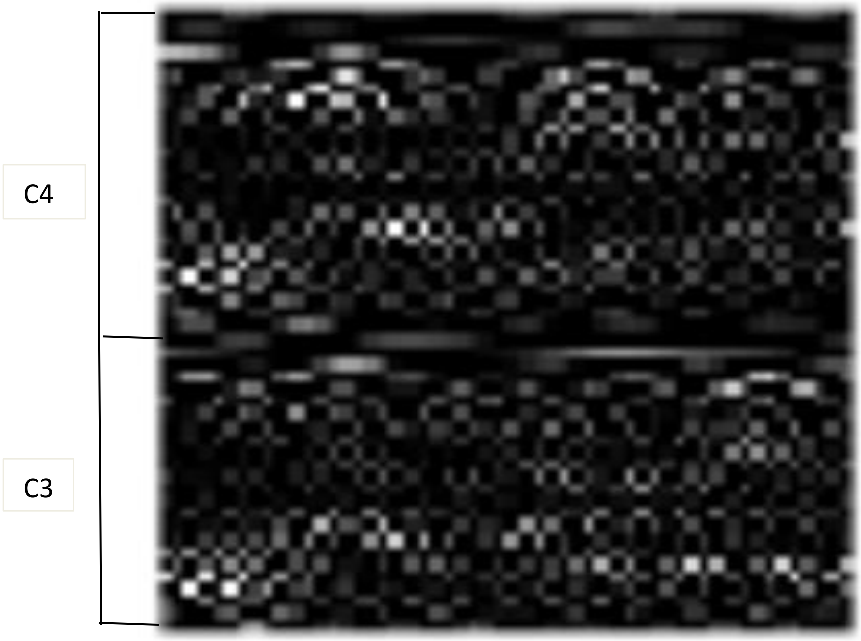

Figure4.2: ERD and ERS effects in Left hand movement

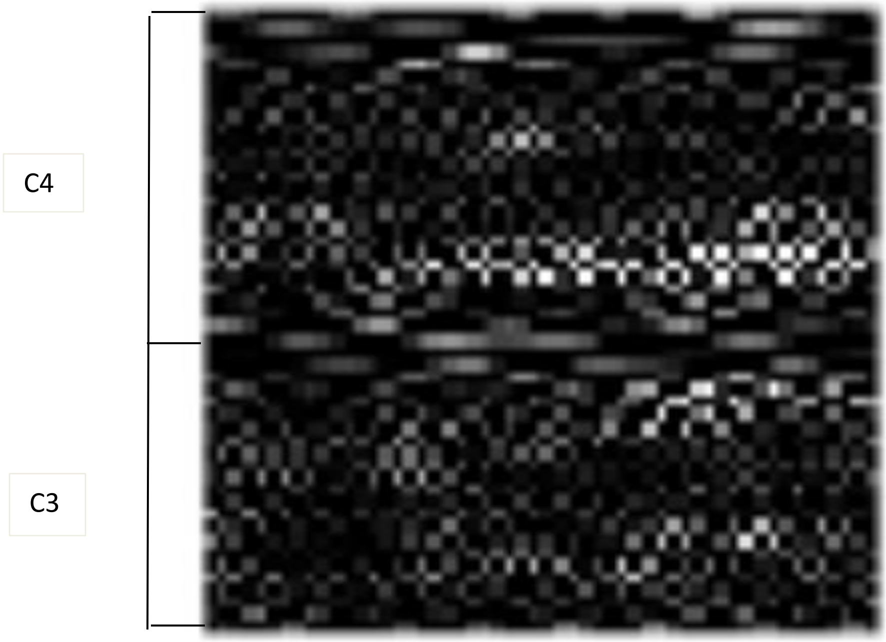

Figure 4.3: ERD and ERS effects in Right hand movement

4.2.2 Analysis of input images

EEG brain signals can be analyzed by variety of Techniques, in literature we have seen MRCP's is one of the techniques for analysis. in this dissertation we have observed Event Related Synchronization (ERS) and Event Related Desynchronization (ERD) in mu (6-13 Hz) band and in beta (17-30 Hz) band of C3 and C4 channels respectively. ERD occurs when a subjects imagine a movement while actually they are not performing the movement. ERD cannot take place alone there must be ERS effect induced in the neighboring neurons. ERD actually means a decrease in amplitude of brain signals when an even occur such as some on imagine to throw a ball. On the other hand ERS means an increase in amplitude of the brain signals when an event occurs. From the in depth knowledge of ERD and ERS it is observed that ERS occurs in mu band of the brain signals that is from 6-13 Hz and ERS occurs in beta band of the brain waves form 17-30 Hz. In this thesis we observed these effects for left and right hand movements. Another major effect which to be noted is brain always takes charges of opposite side body. So, in order to measure left motor imagery movement we need to record readings of right side electrodes which is c3 and for right hand movement the affect is opposite.

The objective of this research is to generate a new form of input which investigate event related patterns in brain signals elicited by motor imagery tasks. The electrophysiological mechanism is a useful way in understanding of ERD and ERS affects in motor imagination, but it is still not completely understood. This study gives an in depth knowledge in understanding of human cortical activity and movement imagery related rhythms. Here, we presented two motor imagery tasks Left and right hand movement and another class when there is no movement were compared to address the following question:

- Does the ERD/ERS patterns lies in mu and beta rhythms?
- Is there a noticeable change in pattern of both ERD and ERS during imagined hand movements.

The above images of left and right hand movements clearly indicate that both ERS and ERD patterns are clearly observable. We also investigate when a subject perform no movement or when his or her mind is doing no task. The resulting STFT images for no movement task is shown in figure 4.4.

Figure 4.4: Short Time Fourier Transform of no movement

4.2.2.1 Power Spectral Analysis

Power Spectrum Analysis is a technique to present the magnitude of signals at measurement points which is needed. Therefore, it is easy to know where more activated spot is at given frequency. Figure 4.5 and figure 4.6 shows power spectrum analysis of left and right hand movement in the respective c3 and c4 channels. We can see that for a right hand movement there is clear desynchronization (ERD) in alpha band of c3 channel as compared to the alpha band of c4 channel.

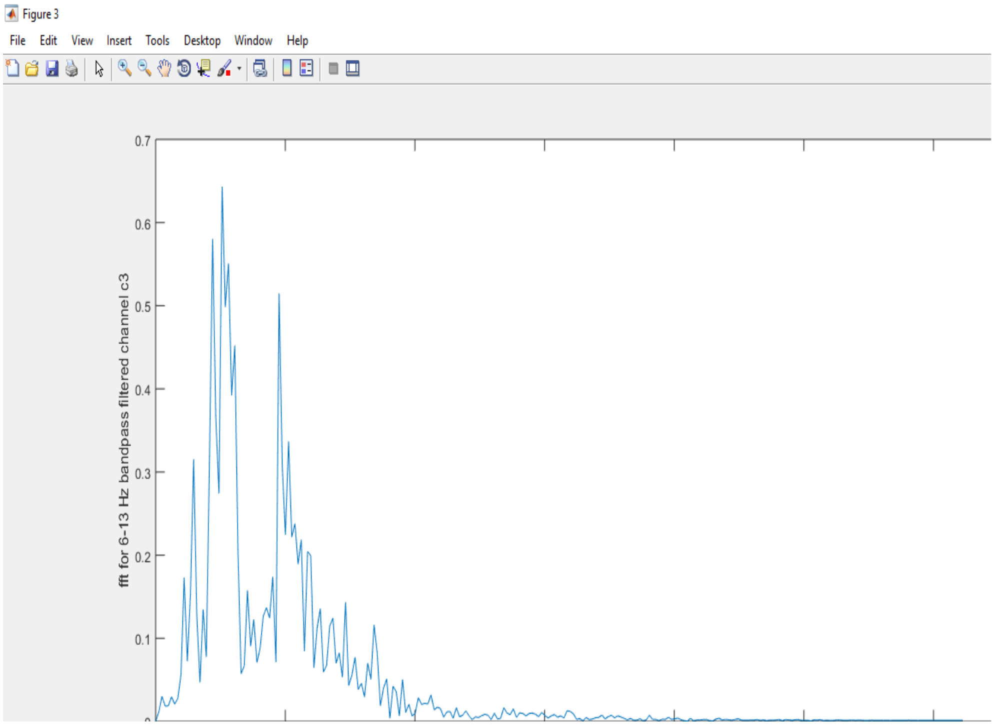

Figure 4.5: Power Spectral density of C3 channel for Right Hand Movement

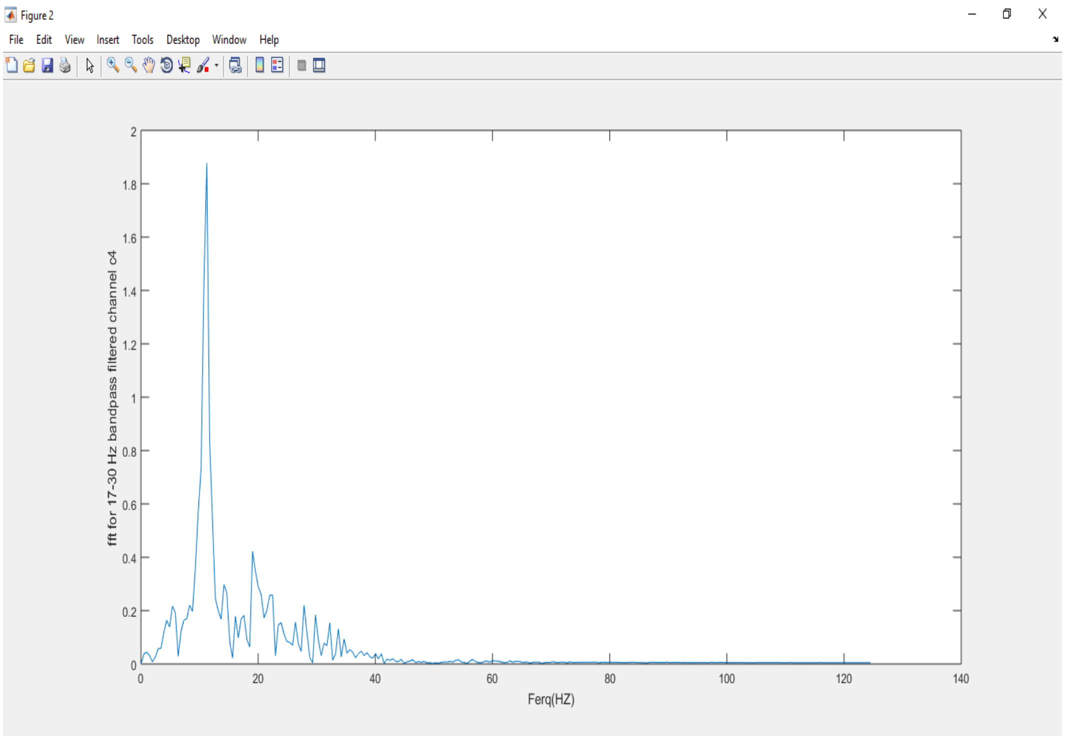

Figure 4.6: Power spectral Density of C4 channel for Right Hand Movement

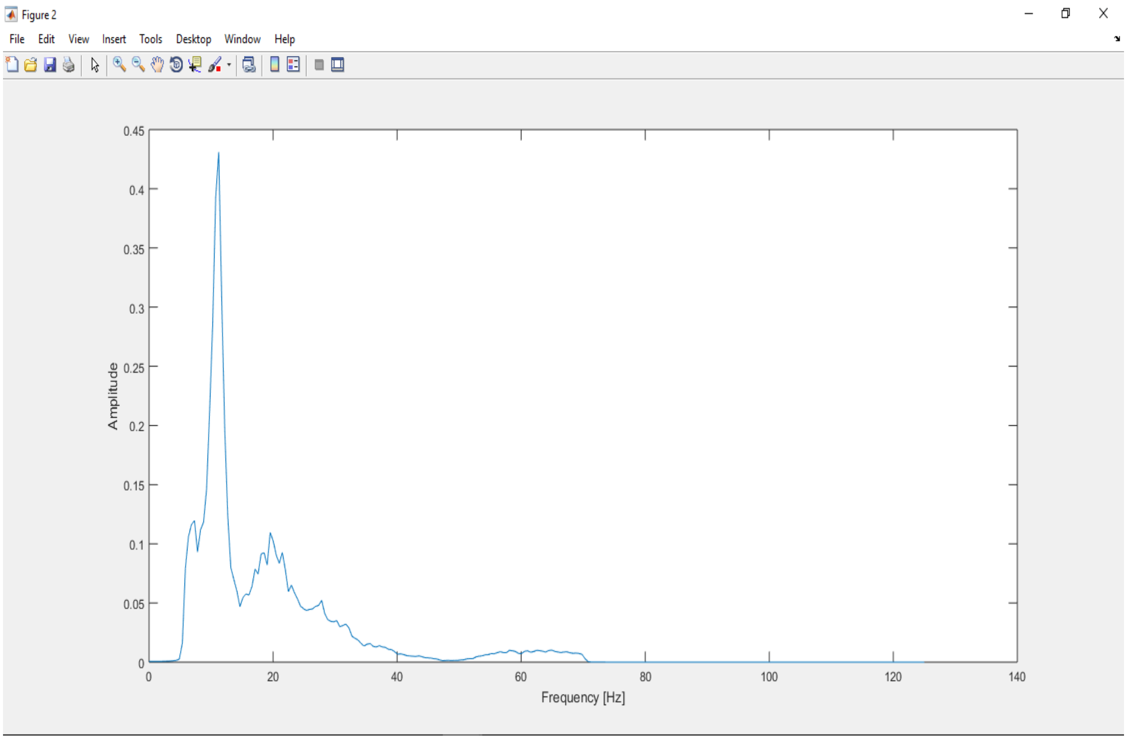

Figure 4.7: Power spectral Density of C3 channel for Left Hand Movement

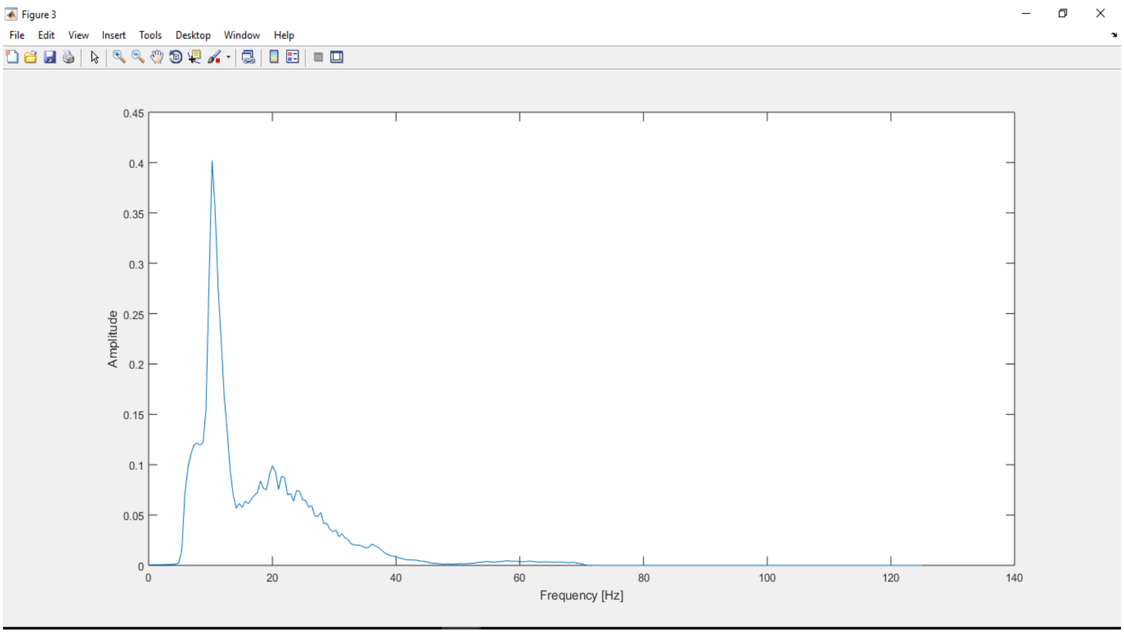

Figure 4.8: Power spectral Density of C4 channel for Left Hand Movement

The power spectral density for left hand movement is shown in figure 4.7 and figure 4.8 for respective c3 and c4 channel. Which shows the same effects as the short time Fourier transformed showed for left hand movement that is it shows ERD in alpha band of c4 channel while at the same time ERS in beta band of C4 channel.

4.2.2.2 Time Frequency Analysis

The Power Spectrum Analysis is a good approach to analyze a signal and compare two classes of tasks based upon the amplitudes at particular frequencies but the main shortcomings of this method is priori knowing of frequencies which would be actually activated. However, if we do not know at which frequency our pattern is activated and at which time span then we need to investigate the whole signal. In this research we have used Event Related information which proved to be a powerful tool to distinguish motor imagery movements in order to visualize mean event-related changes in spectral power over time in a broad frequency range. This technique generalizes event-related desynchronization and synchronization measures. The ERD and ERS effects are quite clear in above figures for left and right hand movements. In figure 4.3, which shows mu waves are relatively decreased in the C3 as compared to C4 channel during the imaginary movement of right arm. In contrast, ERS affect can be observed in the beta band of the C3 which is more activated d as compared to c4. Similarly On the other hands, in figure 4.2, the decrease in mu band in the c4 channel can be seen clearly during the imaginary movement of left arm. While beta band is highly elicited in C4 as compared to C3. In short, ERD affect occurred in the mu band and ERS is generated in the beta band at C3 during the imaginary movement of right arm. Similarly, ERD occurred in the mu band and ERS is generated in the beta band at C4 during the imaginary movement of left arm.

4.3 Deep Learning Classification

When deeper networks starts converging, a degradation problem has been exposed: with the network depth increasing, accuracy gets saturated and then degrades rapidly. Instead of learning a direct mapping of x ->y with a function H(x) (A few stacked non-linear layers). Let us define the residual function using F(x) = H(x)—x, which can be reframed into H(x) = F(x)+x, where F(x) and x represents the stacked non-linear layers and the identity function(input=output) respectively. If the identity mapping is optimal, We can easily push the residuals to zero (F(x) = 0) than to fit an identity mapping (x, input=output) by a stack of non-linear layers. In simple language it is very easy to come up with a solution like F(x) =0 rather than F(x)=x using stack of non-linear CNN layers as function . So, this function F(x) is what the authors called Residual function. A block diagram of plan block and Resnet block is shown in figure 4.9.

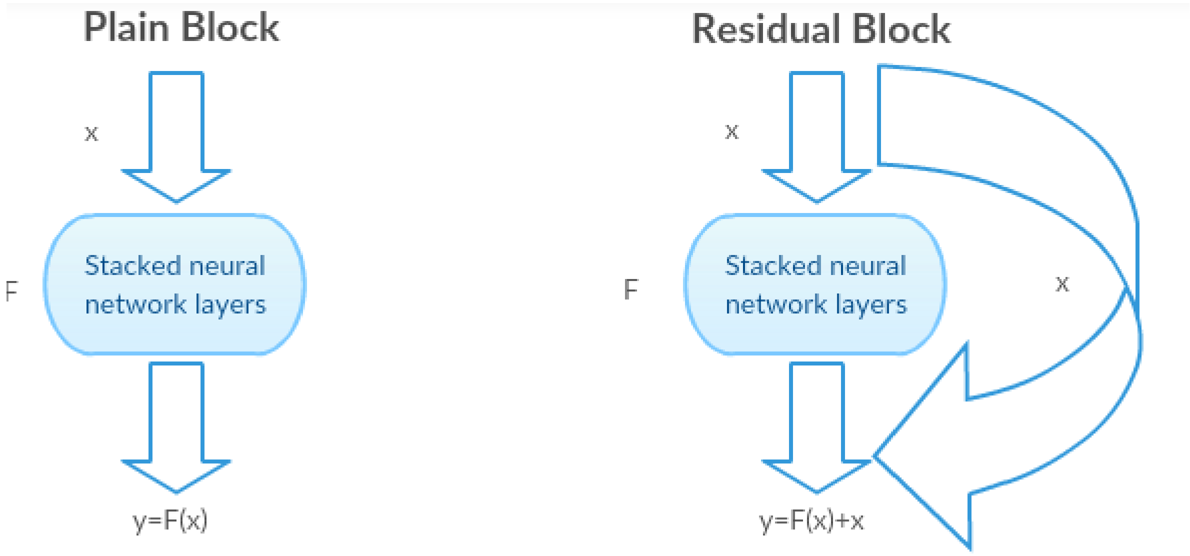

Figure 4.9: Identity mapping in Residual blocks

4.3.1 Results

The algorithms were evaluated by using BCI Competition IV dataset2b. The performance of the proposed method was evaluated by using 10 × 10 fold cross validation and accuracy metrics. In this way, in each session, 90% of dataset were selected randomly as the training set and the remaining 10% were selected as the test set. This process was repeated for 10 times. Similarly for each EEG trial, we extracted the time interval of 4 sec after the cue was displayed. Then STFT was applied to the extracted signal and the input was constructed as explained. In deep neural network method, we used the network shown in figure 16 and described in section 4.3. The network was trained by using batch training method with batch size equal to 32 for 10 epochs. To the best of our knowledge the testing accuracy we obtained is above 80% which is quite higher as compared to simple CNN and other deep neural networks. There are some hyperparameters which were tuned for optimization and is listed below.

4.3.1.1 Hyperparameters Tuning

Hyperparameters are settings that can be tuned to control the behavior of a machine learning algorithm.

Learning Rate: The mother of all hyperparameters, the learning rate quantifies the learning progress of a model in a way that can be used to optimize its capacity. We used the learning rate of 0.001

Optimizer: All of the optimizers manage to converge in a reasonable time. Adam learns the fastest. Adam is more stable than the other optimizers, it doesn't suffer any major decreases in accuracy.

Loss Function: Loss function is a key part of any machine learning model: they define an objective against which the performance of your model is measured, and the setting of weight parameters learned by the model is determined by minimizing a chosen loss function. We have used the Mean Squared Loss Function because in our research we needed pixel by pixel information so, Mean squared error function is the best candidate for the desired application.

Time constraints: Since BCI systems are usually being trained prior to operation and perform testing in real time, testing time is much more critical than the training time for these systems. One training session takes about 720s. However, testing is almost immediate with about 96ms duration. One training session with same dataset takes about 1157 s for CNN-SAE method. However, testing takes with about 400 ms duration which clearly shows that our system achieve faster performance than all previous techniques in testing state as well as the training time of our method is also small.

4.4 Comparison

Accuracy of different classification techniques are compared in the Table 4.3. Which shows our proposed method has the best performance accuracy

Table 4.3 Comparison of different Techniques

Authors and year	**Methods**	**Classifier**	**Accuracy (%)**
Proposed Method	Time Frequency Analysis	Deep Learning Approach	84.9
Indu Dokare, Naveeta Kant 2016	Wavelets decomposition	SVM	69.9
Yousef Rezaei Tabar 2017	Time Frequency Analysis	CNN	72.4
Yousef Rezaei Tabar 2017	Time Frequency Analysis	SAN	70.3
Yousef Rezaei Tabar 2017	Time Frequency Analysis	SAN+CNN	75.1

CHAPTER 5

Conclusion and Future Work

5.1 Conclusion

The use of deep learning in EEG classification is still in its early stages. Currently, it is difficult to find architectures and training paradigms that result in improvements over traditional methods. However, as research continues into the use of deep learning for classification of EEG signals, best practices will become well known. The results of our experiments show a strong case for the heavy use of regularization through both direct methods and dropout, the strength of residual models. Deep learning approach perform best during motor imagery tasks. Residual based convnets models are able to outperform other traditional methods. With the application of transfer learning, deep learning models tested were superior to traditional techniques. It was difficult to exceed the performance of traditional machine learning techniques with deep learning techniques. It took many iterations of model architectures and hyperparameters, along with weeks of computing time to find architectures that outperformed the traditional results initially. However, with the continued exploration of deep learning in the classification of EEG and better guidelines, this time drastically reduced. Furthermore, the benefits provided by transfer learning may be significant enough to allow classification of problems traditional techniques fail on.

5.2 Future Work

There are three main avenues for the extension of this research: application to more datasets, and other forms of transfer learning.

5.2.1 Feature Extraction Technique

There are wide variety of techniques which can be utilized effectively for the extraction of features for left hand and right hand motor imagery tasks. One of them is wavelet decomposition. We have performed feature extraction by using this technique which gives quite better results from visualization perspective and it is hereby expected that our deep learning architecture will give best results for the said technique as compared to Short Time Fourier Analysis. the input images of this techniques which we have extracted is shown in figure below.

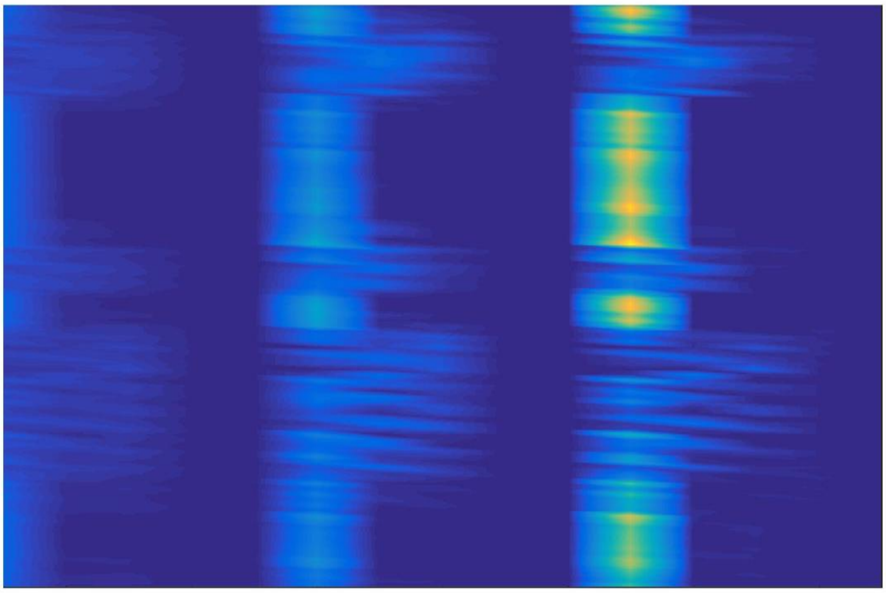

Figure 5.1: Feature extraction for Right hand movement using WT

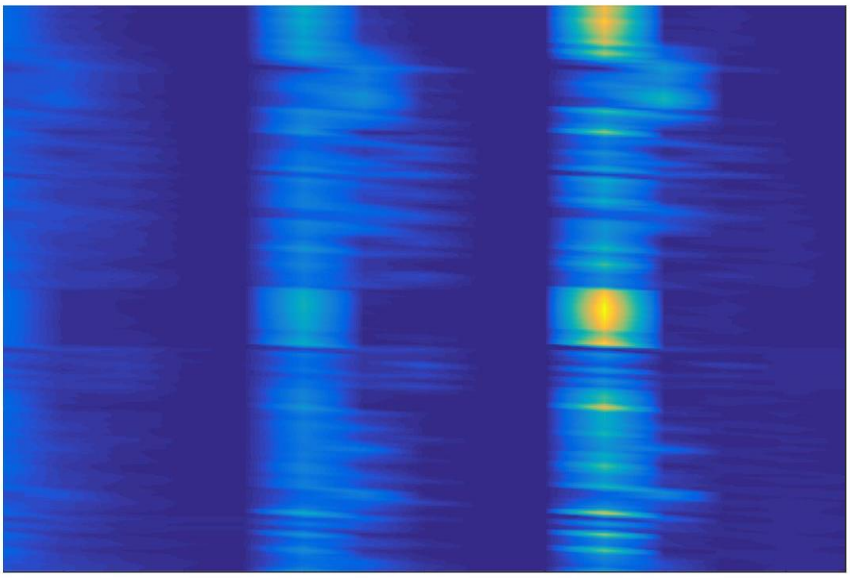

Figure 5.2: Feature extraction for Left hand movement using WT

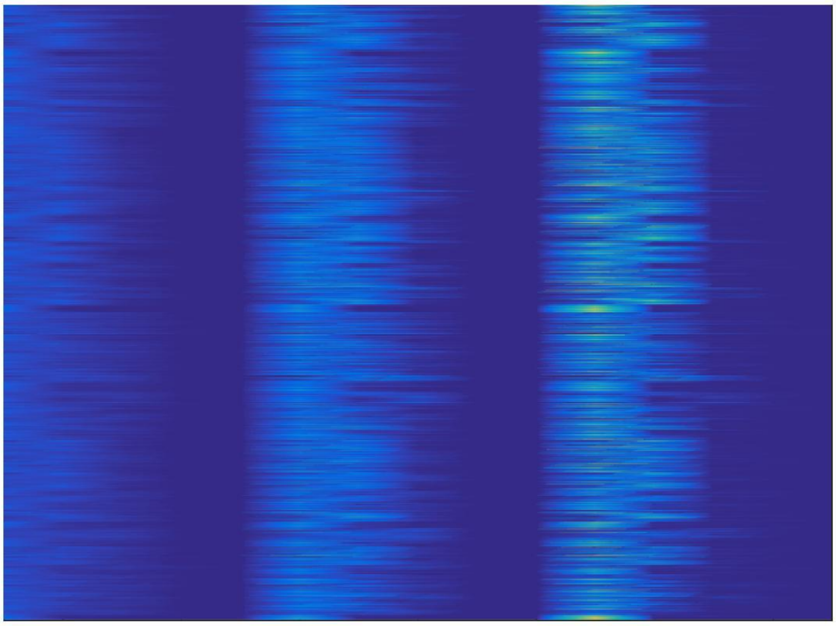

Figure 5.3: Feature extraction for No movement using WT

Applying the same ERD and ERS effects principles for the above images it is confirmed that wavelets gives good visualization for motor imagery movements individually. So, it is expected that if visually they are quite easily identifiable then our deep learning architecture will surely classify them with highest of accuracy. Even we will try to apply a simple CNN to classify them in order to avoid deep network.

5.2.2 More Datasets

Currently, these techniques were only applied to a single dataset. In order to show generalizability, it is desirable to replicate this research on further datasets with different properties. The dataset in this thesis involved the classification of motor imagery tasks. While this is a fairly common task in EEG classification, there are several other common tasks that should also be explored. Due to the ease of classification and potential upside in this task, it is a valuable extension. Similarly, experiments where the subject thinks about words from a set vocabulary have immediate potential use in communication based BCIs, and would also show a valuable extension of this research. Finally, extending this work to datasets of any sort of stimuli that have been difficult to classify using traditional techniques is important in establishing the benefit of these techniques. The dataset explored in this thesis was collected on a 3 channel EEG. The negative impact of feature reduction on deep learning found in our results suggests that having a higher number of recorded features may improve the classifiably of the signal, despite the issues with the curse of dimensionality.

5.2.3 Transfer Learning Techniques

Transfer learning is a family of techniques involving the use of one set of data to create an initialization for the classification of another set of data. Currently, we have only explored the training across all subjects in an experiment, then fine-tuning the entire network on an individual in the network. There are many other transfer learning paradigms that may show strong performance. The most straightforward change that could be made in the transfer learning paradigm is simply freezing the bottom layers of the network during fine-tuning. That is to say, only allow the classification layer and perhaps the last fully connected layer to update during the fine-tuning phase. The benefit of this approach is that it may reduce overfitting during the fine-tuning phase, which is likely a problem with our current approach given the high variance of the classification accuracies within subjects. Thus, it is reasonable to believe that applying transfer learning over several datasets to learn universal basic features for the early filters in the network could be beneficial. It may also allow networks with more parameters, and thus greater representative power to be trained than could be achieved with a single dataset. There are some barriers that would need to be considered, however, especially if the data is collected on different machines or has a different time span.

References

[1] O. C. R. a. &. S. R. H. AlZoubi, "Classification of EEG for affect recognition: An adaptive approach.," 2009.

[2] J. M. J. B. H. L. K. S. &. B. M. Correia, "EEG decoding of spoken words in bilingual listeners: From words to language invariant semantic-conceptual representations.," Frontiers in Psychology,, 2015.

[3] A. J. L. R. R. &. D. F. U. Gabor, "Automated seizure detection using a self-organizing neural network.," *Electroencephalography and Clinical Neurophysiology, ,* p. 257–266., 1996.

[4] S. T.-S. S. J. M. K. N. E. L. T. Ramgopal, "Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy.," *Epilipsy and behaviour,* p. 291– 307., 2014.

[5] G. G. N. A. &. T. U. Molina, "Emotional Brain-Computer Interfaces,," pp. 138-146, 2009.

[6] W. R. R. S. R. E. M. V. &. G. M. D. V. &. G. M. D. Shirer, "Decoding subject-driven cognitive states with whole-brain connectivity patterns.," *Cerebral cortex. ,* pp. 158-165, 2012.

[7] L. R. L. A. C. F. F. J. K. P. a. M. A. R. &. N. R. Trambaiolli, "Improving Alzheimer’s disease diagnosis with machine learning techniques.," *Clinical EEG and Neuroscience : Official Journal of the EEG and Clinical Neuroscience society,* pp. 160-165, 2011.

[8] L. S. D. T. D. V. N. &. G. R. Zhang, "Machine learning for clinical diagnosis from functional magnetic resonance imaging. Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference, 1211–12," *Computer vision and Patern Recognition,* pp. 1211-1217, 2005.

[9] N. M. E. M. B. S. R.-R. A. F. T. K. J. D. C. Koutsouleris, "Individualized differential diagnosis ofschizophrenia and mood disorders using neuroanatomical biomarkers.," pp. 2059-2073, 2015.

[10] D. A. B. D. N. M. A. &. N. J. T. Fair, "Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD.," *Proceedings of the National Academy of Sciences, 109(17), ,* p. 6769–6774. , 2012.

[11] S. a. C. J. Sanei, EEG signal processing, John Wiley & Sons, Ltd, 2007.

[12] B. A. B. M. C. L. T. V. D. a. G. A. Graimann, "‘Non-invasive brain-computer interfaces for semi-autonomous assistive devices’,," *Robust Intelligent Systems, Spring London Ltd.,* pp. 113-137., 2008.

[13] M. e. a. Kouijzer, " ‘Long-term effect of neurofeedback treatment in autosim’,," *Research in Autism Spectrum Disorders,,* pp. 469-501, 2009.

[14] B. D. G. K. M. M. K. a. C. G. Blankertz, "The non-invasive Berlin Brain-Computer interface: fast acquisition of effective performance in untrained subjects," *Neuroimage,* pp. 539-550., 2007.

[15] "Non-Invasive Brain-Machine Interaction," *Article in International Journal of Pattern Recognition and Artificial Intelligence,* August 2008.

[16] N. M. G. B. J. Baars, "Brain, and Consciousness: Introduction to Cognitive Neuroscience,," *Elsevier,* 2010.

[17] W. K. R. J. C. C. R. R. C. S. T. M. B. H. A. D. M. R. a. K. M. .. Trejo L J, "Multimodal neuroelectric interface development," *IEEE Trans. Neural Syst. Rehabil. Eng.,* p. 199–204, 2003 .

[18] T. A. P. J. S. Ebersole, Current Practice of Clinical Electroencephalography,, Lippincott Williams & Wilkins,, 2003, p. 72–99.

[19] D. T. V. P. a. D. V. C. P. M. R. Lakshmi, ""Survey on EEG Signal Processing Methods,"," *International Journal of Advanced Research in Computer Science and Software Engineering, ,* pp. 84-91, , 2014.

[20] D. J. McFarland, ""The advantages of the surface Laplacian in brain– computer interface research,"," *International Journal of Psychophysiology, ,* 2014.

[21] D. J. M. a. J. R. W. J. Lu, " "Adaptive Laplacian filtering for sensorimotor rhythm-based brain–computer interfaces,"," *Journal of neural engineering, vol. 10, ,* pp. 016002,, 2013.

[22] L. a. D. Senthilmurugan.M, "Classification in EEG-Based Brain Computer Interfaces Using Inverse Model,," *Int. Journel of Computer Theory and Engg., Vol. 3, No. 2, ,* April 2011.

[23] M. M.-T. M. H. C. M. H. a. P. R. W. S. Motamedi-Fakhr, ""Signal processing techniques applied to human sleep EEG signals—A review,"," *Biomedical Signal Processing and Control, vol. 10, ,* pp. 21-33,, 2014.

[24] F. L. F. a. F. C. M. N. Mammone, ""Automatic Artifact Rejection From Multichannel Scalp EEG by Wavelet ICA,"," *Sensors Journal, IEEE, vol. 12, ,* pp. 533-542, , 2012.

[25] P. C. a. K.-B. S. X. Yu, ""Analysis the effect of PCA for feature reduction in non-stationary EEG based motor imagery of BCI system,"," *Optik - International Journal for Light and Electron Optics, vol. 125, ,* pp. 1498-1502,, 2014..

[26] L. F. Nicolas-Alonso and J. Gomez-Gil, ""Brain computer interfaces, a review,"," *Sensors, vol. 12, ,* pp. 1211-1279, , 2012.

[27] Fabien Lotte, "Study of Electroencephalographic Signal Processing and Classification Techniques towards the use of Brain-Computer Interfaces in Virtual Reality Applications,," 2009.

[28] W. D. H. K. M. M. W. a. K. R. V. Lawhern, ""Detection and classification of subject-generated artifacts in EEG signals using autoregressive models,"," *Journal of Neuroscience Methods, vol. 208, ,* pp. 181-189, , 2012..

[29] H. C. &. Y. ZhangShuanghu., "Analysis Between Fourier Transform and Wavelet Transform.," *HeiBei: Journal of Hebei University of Technology.,* 2005 .

[30] T. A.-H. F. Abdul- Bary Raouf Suleiman, "Features Extraction Techniques of EEG Signal For BCI Applications,," Comp. and Inf. Engg. Dept. College of Electronics Engg., University of Mosul, Mosul, Iraq,, 2011..

[31] N. H. G. K. A. S. I. &. S. R. Srivastava, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research, ,* vol. http://doi.org/10.1214/12AOS1000 , no. 15, p. 1929–1958., 2014.

[32] F. L. C. K. R. S. H. B. G. P. R. Leeb, "Brain-computer communication: motivation, aim, and impact of exploring a virtual apartment.," *IEEE Transactions on Neural Systems and Rehabilitation Engineering 15, ,* pp. 473–482, , 2007.

[33] C. K. D. Z. R. S. R. L. G. P. A. Schl¨ogl, "A fully automated correction method of EOG artifacts in EEG recordings.," *Clin.Neurophys,* pp. 118(1):98-104., 2007 .

[34] E. L. a. P. N. G.E. Chatrian, " Ten percent electrode system for topographic studies of spontaneous and evoked EEG activity.," *Am J EEG Technol,* pp. 25:83-92, 1985..

[35] H. Berger, "Uber das Elektroenkephalogram des Menschen," *Archives of Psychiatry,* pp. Vol. 87, pp. 527-70..

[36] N. E. a. L. d. S. F, "Electroencephalography: basic principles, clinical applications, and related fields," Lippincott Williams & Wilkins,, 2005.

[37] T. M. ,. H. W. J. ,. T. L. J. ,. R. W. W. M. M. M. K. A. D. B. H. B. N. D. E. W. E. W. a. W. J. R. Vaughan, "Brain-computer interface technology: a review the second international meeting," *IEEE Transection on Neural systems and rehabilitation Engineering,* pp. 94-109, 2003.

[38] W. E. a. W. J. Allison. B.Z., "Brain computer interface systems: progress and prospects.," *British Review of Medical Device, ,* pp. 463-474., 2007.

[39] R. Brunelli, " Template Matching Techniques in Computer Vision:," in *Theory and Practice, ,* Wiley, New York., 2009.

[40] E. D. Ubeyli, "‘Decision support systems for time-varying biomedical signals: EEG signals classification’,," *Expert Systems with Applications, ,* p. 2275–2284, 2009a.

[41] P. &. M. R. Stoica, "Introduction to Spectral Analysis,," New Jersey:, Prentice Hall., 1997.

[42] A. R. L. A. a. R. F.-R. S. Amiri, "A Review of P300, SSVEP, and Hybrid P300/SSVEP Brain- Computer Interface Systems," 2013.

[43] A. B. R. K. W. G. E. B. M. Fatourechi, "EMG and EOG artifacts in brain computer interface systems: a survey.," *Clinical Neurophysiology 118, ,* pp. 480–494, , 2007.

[44] J. K. J. E. H. S. G. M. A. Schl¨ogl, "Evaluation criteria in BCI research," pp. 327–342, , 2007..

[45] W. Nowinski, "Human brain atlasing: past, present and future.," *The Neuroradiology Journal, ,* vol. 30(6), pp. 504-519., 2017.

[46] J. &. S. A. Kevric, "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system," *Biomedical Signal processing and control ,* pp. 398-408, 2017.

Completion Certificate

It is certified that the contents of thesis document titled “A deep learning approach for classification of motor imagery signals” submitted by Ms. Memoona Iftikhar Registration No. 00000170972 have been found satisfactory for the requirement of degree.

Thesis supervisor: _______________________

(Dr Shoab Ahmad Khan)