The Influence of Graphical User Interface on Motion Onset Brain-computer Interface Performance and the Effect of Data Augmentation on Motor Imagery Brain-computer Interface

Tao Lyu

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Engineering
College of Science and Engineering
University of Glasgow

Jan 2023

Abstract

Motor Imagery Brain Computer Interface (MI BCI) is one of the most frequently used BCI modalities, due to the versatility of its applications. However, it still has unresolved issues like time-consuming calibration, low information transfer rate, and inconsistent performance across individuals. Combining MI BCI with Motion Onset Visual Evoked Potential (mVEP) BCI in a hybrid structure may solve some of these problems. Combining MI BCI with more robust mVEP BCI, would increase the degrees of freedom thereby increasing the information transfer rate, and would also indirectly improve intrasubject consistency in performance by replacing some MI-based tasks with mVEP. Unfortunately, due to Covid -19 pandemic experimental research on hybrid BCI was not possible, therefore this thesis focuses on two BCI separately.

Chapter 1 provides an overview of different BCIs modalities and the underlying neurophysiological principles, followed by the objectives of the thesis. The research contributions are also highlighted. Finally, the thesis outlines are presented at the end of this chapter. Chapter 2 presents a comprehensive state of the art to the thesis, drawing on a wide range of literature in relevant fields. Specifically, it delves into MI BCI, mVEP BCI, Deep Learning, Transfer Learning (TL), Data Augmentation (DA) and Generative Adversarial Networks (GANs). Chapter 3 investigates the effect of graphical elements, in online and offline experiments. In the offline experiment, graphical elements such as the color, size, position, and layout were explored. Replacing a default red moving bar with a green and blue bar, changing the background color from white to gray, and using smaller visual angles did not lead to statistically significant improvement in accuracy. However, the effect size of η^2 (0.085) indicated a moderate effect for these changes of graphical factors. Similarly, no statistically significant difference was found for the two different layouts in online experiments. Overall, the mVEP BCI has achieved a classification accuracy of approximately 80%, and it is relatively impervious to changes in graphical interface parameters. This suggests that mVEP is a promising candidate for a hybrid BCI system combined with MI, that requires dynamic, versatile graphical design features. In Chapter 4, various DA methods are explored, including Segmentation and Recombination in Time Domain, Segmentation and Recombination in Time-Frequency Domain, and

Spatial Analogy. These methods are evaluated based on three feature extraction approaches: Common Spatial Patterns, Time Domain Parameters (TDP), and Band Power. The evaluation was conducted using a validated BCI set, namely the BCI Competition IV dataset 2a, as well as a dataset obtained from our research group. The methods are effective when a small dataset of single subject are available. All three DA methods significantly affect the performance of the TDP feature extraction method. Chapter 5 explored the use of GANs for DA in combination with TL and cropped training strategies using ShallowFBCSP classifier. It also used the same validated dataset (BCI competition IV dataset 2a) as in Chapter 4. In contrast to DA method explored in Chapter 4, this DA is suitable for larger datasets and for generalizing training based on other people's data. Applying GAN-based DA to the dataset resulted on average in a 2% improvement in average accuracy (from 68.2% to 70.7%). This study provides a novel method to enable MI GAN training with only 40 trials per participant with the rest 8 people's data for TL, addressing the data insufficiency issue for GANs. The evaluation of generated artificial trials revealed the importance of inter-class differences in MI patterns, which can be easily identified by GANs.

Overall the thesis addressed the main practical issues of both mVEP and MI BCI paving the way for their successful combination in future experiments.

University of Glasgow

College of Science & Engineering

Statement of Originality

Name: Tao Lyu

Registration Number:

I certify that the thesis presented here for examination for a Ph.D. degree of the University of Glasgow is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it) and that the thesis has not been edited by a third party beyond what is permitted by the University's PGR Code of Practice.

The copyright of this thesis rests with the author. No quotation from it is permitted without full acknowledgment.

I declare that the thesis does not include work forming part of a thesis presented successfully for another degree.

I declare that this thesis has been produced in accordance with the University of Glasgow's Code of Good Practice in Research.

I acknowledge that if any issues are raised regarding good research practice based on the review of the thesis, the examination may be postponed pending the outcome of any investigation of the issues.

Signature: .....Tao Lyu.....

Date: .....30/Jun/2022.....

Contents

List of Tables

List of Figures

List of Acronyms

ALS	Amyotrophic Lateral Sclerosis
AT	Artificial Trials
BA	Brodmann Area
BAR	Brain-Area-Recombination
BCI	Brain Computer Interface
BLDA	Bayesian Linear Discriminant Analysis
BN	Batch Normalization
BNCI	Brain Neural Computer Interface
BP	Band Power
CDAN	Conditional Domain Adaption Neural Network
CNN	Convolutional Neural Network
CNS	Central Nervous System
CSP	Common Spatial Pattern
CV	Computer Vision
CWGAN-GP	Conditional Wasserstein Generative Adversarial Network with Gradient Penalty
CWT	Continuous Wavelet Transform
DA	Data Augmentation
DCGAN	Deep Convolutional Generative Adversarial Networks
DFBCSP	Discriminative Filter Bank Common Spatial Pattern
DFT	Discrete Fourier Transform
DL	Deep Learning
EEG	Electroencephalography
ERD	Event Related Desynchronization
ERS	Even Related Synchronization
fMRI	functional Magnetic Resonance Imaging
FBCSP	Filter Bank Common Spatial Pattern
GAN	General adversarial networks
GPU	Graphics Processing Unit

GUI	Graphical User Interface
IC	Independent Components
IMFs	Intrinsic Mode Functions
LDA	Linear Discriminant Analysis
LGN	Lateral Geniculate Nuclei
LOO	Leave One Out
LR	Linear Logistic Regression
LSTM	Long-term, Short-term Memory Network
MEG	Magnetoencephalography
MI	Motion Imagery
MI-BCI	Motor Imagery Brain Computer Interface
ML	Machine learning
MSE	Mean Square Error
MTS	Medial Superior Temporal
mVEP	motion-onset Visual Evoked Potential
nIRS	near Infrared Spectroscopy
Natural Language Processing	NLP
Original Trials	OT
PC	Principal Components
PET	Positron Emission Tomography
PMA	Premotor Area
PNS	Peripheral Nervous System
PSD	Power Spectrum Density
QTFD	Quadratic Time-Frequency Distribution
ReLU	Rectified Linear Units
REM	Rapid Eye Movement
RGB	Red, Green, Blue
RF	Receptive Fields
RFB	Red Face with a Blue rectangle
RFR	Red Face with a Red rectangle
RFW	Red Face with a White rectangle
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SA	Spatial Analogy
SAM	Stacked Autoencoders
SMA	Supplementary Motor Area

SRTD	Segmentation and Recombination in Time Domain
SRTFD	Segmentation and Recombination in Time-Frequency Domain
SSVEP	Steady State Visual Evoked Potential
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
TDP	Time Domain Parameter
T.L.	Transfer Learning
TP	True Positive
VEP	Visual Evoked Potential
WNN	Wavelet Neural Network

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Aleksandra Vuckovic, for her professional guidance and unconditional support throughout my entire Ph.D. study. When our experiments were forbidden due to the lockdown, she gave me suggestions on new research topics; when I suffered from illness, she always comforted me via emails; when I encountered research questions, she gave me suggestions and provided me with references. I am grateful to her for always believing in me and giving me the support to explore my research. She provided indispensable help whenever I needed it, both as a supervisor and friend. Her wide knowledge, strong research enthusiasm, and hard-working attitude have inspired me during my Ph.D. and will profoundly affect my future career. It is extremely lucky to have Dr. Aleksandra Vuckovic as my supervisor.

Secondly, I want to thank my motherland, China, since China Scholarship Council gave me the opportunity and funding to study aboard. I appreciate my motherland, the government, and the people. I also want to thank the United Kingdom, which provided me with funding during my extension caused by the lockdown.

Lastly, I want to express my appreciation to my boyfriend, family, friends, colleagues, and the working staff at the University of Glasgow for each time they helped me and supported me.

I really appreciate the experience in the UK. I will never forget the passionate Scottish people who made me feel at home.

Chapter 1

Introduction

1.1 Introduction of Brain-computer Interface

Brain-Computer interface is a human-computer interface system, which is most commonly based on EEG signals and does not depend on the muscular pathways [1]. A typical BCI system consists of the following components: EEG acquisition and preprocessing system, signal feature extraction, awareness task classification, and control output application (Fig. 1.1) The concept of BCI was first introduced in the 1970s, when Vidal [2] proposed a system that aimed to provide physically disabled people with an unconventional non-muscular way of communication. Since then, BCI research has progressed in many fields over the last 50 years. A document produced by the International BCI society, "The future of brain/neural computer interaction: horizon 2020" defines six main applications of BCI to replace, restore, enhance, supplement, or improve the natural CNS output; BCI can also be used as a tool to investigate functions of NCO in the clinical and non-clinical population [3]. People with Amyotrophic Lateral Sclerosis (ALS) [4, 5], Cerebral palsy [6, 7], and spinal cord injury [8, 9] can rely on an assistive BCI device to interact with the environment. MI-BCI and Visual Evoked Potential (VEP)-based BCI have commonly used BCI modalities and are sometimes used together in a hybrid BCI [10]. Very often VEP based BCI is used to initiate BCI (switch it on or off), and MI BCI is used for communication and control [10]. Alternatively, they can be used together, e.g., control of a BCI game to increase the number of possible commands and to increase the dynamic of a game.

1.1.1 Application of Brain-computer Interface

According to the Brain Neural Computer Interface (BNCI) 2020 document [11], recognized by the International Brain-computer Interface society BCI:

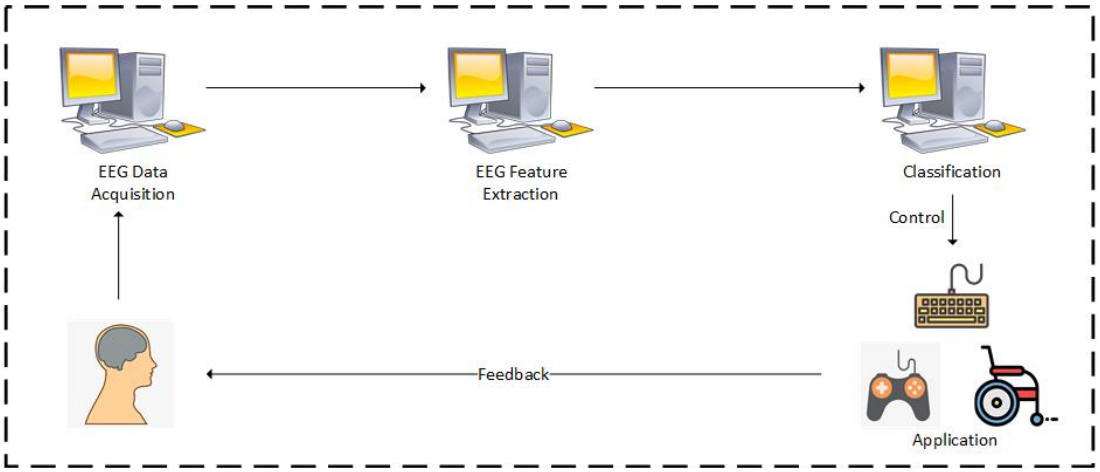


Figure 1.1: A structure of a typical Brain-computer Interface system [1]

- BCIs can restore lost functions, for example, stimulation of muscles in a paralyzed person.
- BCIs can be used to improve functions for example, in stroke rehabilitation.
- BCIs can enhance functions, for example, detection of stress levels or lapses of attention during demanding tasks).
- BCIs can replace functions for example by controlling a robotic arm, exoskeleton, or wheelchair in a paralyzed person.
- BCIs can also be used as a research tool to study brain functions.

1.1.2 Brain-computer Interface Classification

BCI can be classified based on electrode location and recording modality, based on the source of stimuli and its working mode [12].

(1) According to the electrode position, it can be divided into implantable and non-implantable. Implantable BCI uses either electrode placed directly on the cortex (electrocorticography ECOG) or an intracranial microelectrode array, which can accurately record the electrical activity of individual neurons It has the advantages of high signal-to-noise ratio, obvious features, simple post-processing, and high classification accuracy. The problems are related to high surgical risks, the issues of the best location of electrode implantation, the continuous working time of the implanted electrode, and the selection of the suitable population of the implanted electrodes.

Non-implantable BCI (EEG, MEG, nIRS, fNIRS) places electrodes on the skull or close to the skull to measure either neuronal activity or some related

measures, such as blood oxygenation. The most widely used method for recording non-invasive brain activity is EEG due to its relatively affordable price, portability, and ability to record brain activity in freely moving people. However, the EEG signal-to-noise ratio is relatively low, and the surrounding environment interference and human subjective mental state have a great impact on the EEG signal, especially for BCI applications that use spontaneous EEG signals as input.

(2) According to the BCI input signal generation method it can be classified as exogenous (evoked) and spontaneous i.e. endogenous. Users' sensory pathways generate stable and consistent evoked EEG signals through specific stimulation devices, which serve as the input to the BCI system. These signals are then processed and analyzed to extract meaningful information and enable communication or control of external devices. Since this type of system requires an additional stimulation device and depends on a certain sensory pathway of the user, the scope of use is limited. However, the regularity of evoked potentials is relatively strong, users generally do not need to be trained, and the classification accuracy is high. Examples of exogenous BCI are Steady State Visual Evoked Potential (SSVEP), motion Visual Evoked Potential (mVEP), and P300. Although MI is generated internally cue-based, MI BCI also requires an external cue to initiate MI.

The input signal of the spontaneous BCI system comes entirely from the user's spontaneous internally generated EEG, no external stimulation device is needed. It does not occupy the user's sensory pathway. In this sense, the spontaneous BCI is completely in line with the original definition of BCI. Spontaneous EEG is particularly vulnerable to extrinsic disruption when compared to Evoked EEG due to its low-intensity and non-stationary nature. The brain's intrinsic activities are frequently disturbed by electromagnetic interference from outside sources such as environment, emotion, or cognitive load. In contrast, evoked EEG signals are elicited by specific stimuli that are often deliberately manipulated and regulated, thereby rendering them easier to regulate and interpret. Therefore, the BCI system in this way requires high signal processing methods, and the current system classification accuracy is low. Examples of spontaneous EEG are slow cortical potential BCI and BCI based on MI in asynchronous mode as explained below. Users typically need some training to learn to use this type of BCI.

(3) According to the working mode, BCI can be classified into synchronous and

asynchronous BCI. The synchronous BCI system requires users to carry out specific mental actions within a predetermined time, that is, the BCI system initializes the control time, and the communication between the user and BCI must be consistent. In this way, the system knows in advance which segment of EEG data to process, thus greatly simplifying signal processing. An example is a cue-based motor imagery task.

For asynchronous BCI systems, users can start a specific mental activity, that is, communication is not controlled by the system but is initiated by the user. However, the asynchronous BCI system must monitor the user's EEG signal in real-time, and it is necessary to accurately determine the starting point, which makes the implementation of the asynchronous BCI much more complicated than the synchronous BCI.

To overcome the problem of asynchronous BCI, hybrid BCI systems are created that combine several BCI modalities or combine control based on brain activity with some other physiological measure such as muscular activity (electromyogram EMG), eye movement (electro oculagram EOG) or heart rate (electrocorticogram ECG). In this way, one mode can be used to switch the BCI on or off, and the other to operate BCI, for example, EOG can be used to activate a speller based on P300. Alternatively, one mode can be used to serve as a cue for the other. For example, a hybrid BCI can use SSVEP as a cue for MI. Finally, two BCI modalities (e.g. mVEP and cue-based MI) can be used to increase the number of degrees of freedom of a BCI system and to increase the information transfer rate, making BCI faster and more engaging.

1.2 Motivation

Brain-Computer Interfaces (BCIs) have gained popularity in the field of research as an inclusive modality for both rehabilitation and entertainment. Several studies have confirmed the efficacy of Motor Imagery (MI) BCI in the recovery of motor function. By imagining the movement of the hand or even observing someone else's hand movement, users can perform exercises to aid in their rehabilitation. Our laboratory is dedicated to using BCI for neurorehabilitation in individuals whose motor function has been impaired. Nevertheless, MI BCI has some limitations that prevent it from being commonly used in clinical or home settings. For example, the variability of MI patterns among users results in a lengthy calibration time for a single user. Moreover, the setup and calibration of BCI typically require a professional. Furthermore, people easily become bored

and fatigued during offline data collection. In addition, a MI BCI protocol typically comprises a few seconds for the imagination of hand movement followed by a few seconds of rest. As a result, the information transfer rate is quite low for any real-time application. Lastly, MI BCI typically has 3-4 degrees of freedom and as a result, can only produce a limited number of commands for communication and control [13]. It can be concluded that MI BCI is not user-friendly at its current stage.

1.3 Objectives

This thesis proposes different approaches to increase the usability of MI BCI. Firstly, to enhance the appeal of MI BCI, the traditional simple and plain UI can be replaced with a game UI. To increase the degree of freedom in the game, MI BCI can be integrated with Visual Evoked Potential (VEP) based BCI to create a hybrid BCI gaming platform. VEP-based BCI has the advantages of a higher Information Transfer Rate (ITR), quick response, and less training demand. We have successfully built such a hybrid BCI gaming platform; however, due to the COVID-19 pandemic, experiments on participants were not possible. Consequently, the research topic of this thesis will focus on data analysis of VEP BCI and MI BCI separately. During the integration of the two types of BCI into a gaming platform, various types of VEP were considered, such as Stead State Visual Evoked Potential (VEP), motion Visual Evoked Potential (mVEP), and P300. Compared to the other two, mVEP has the advantage of causing the least fatigue and requiring the least concentration, which is crucial for long-term gaming use. Concerning mVEP runs in the context of the game, the graphic factors of VEP drew our attention. Published papers have indicated that the color of the background and targets in SSVEP and P300 causes a significant difference in user performance. However, few studies have explored how these factors affect mVEP. Thus, the first research topic in this thesis is the influence of different graphic factors of mVEP BCI. For the mVEP study, the research objective mainly focuses on the influence of mVEP graphical factors when developing 3D BCI games. Based on the related studies on P300 and SSVEP, colors, sizes, and layouts might be graphical that affect the BCI performance. To explore the graphic factors, both offline and online experiments were conducted. The offline mVEP BCI experiment adopted the UI from Tsinghua University as the baseline session (Layout A) with four boxes located at the four sides of the screen, a red moving bar as the target, and a pure white background. The other five sessions (Layouts B-F) changed only one of the graphic factors compared to Layout A.

The change in Layouts B and C is the color of the target moving bar, replacing red with green and blue. Meanwhile, Layout D has a grey background color when compared with Layout A, Layout E has a smaller size of the moving target, and Layout F has a more central view than Layout A. As such, the six layouts compare the influence of the color of targets, and the background color. The impact of box positioning on performance within the layout is also investigated. The proposed research questions for the mVEP study are as follows:

(1) Does the components of the graphical presentation of an mVEP BCI have an impact on its performance? More specific:

   a) The color of the moving target

   b) The background color

   c) The size of the moving target

   d) The visual angle of the moving target.

(2) Based on the findings, what factors should be taken into account when integrating mVEP UI into a game UI?

(3) What is the ideal combination and number of electrodes for mVEP?

(4) whether people have a preference for a specific side (up, down, left, and right of the screen) of the moving target.

The aforementioned questions aim to determine the factors that affect the performance of the mVEP BCI system and to identify the best approach to integrate mVEP UI into game UI. The study also aims to determine the optimal combination and number of electrodes for mVEP, which is critical for its practical use.

In the context of the online experiment, two distinct layouts are investigated. Unlike the offline section, where comparisons are made among different layouts. Based on the accuracy of the online experiment, the research question that arose is :

(1) whether the target's positions affect accuracy.

These findings can be used as a reference for designing the MVEP Game UI.

We also explored the possibility of generating synthetic data to reduce the calibration time of the MI BCI and make it more user-friendly. These methods are based on both traditional machine learning and deep learning. Three artificial data generation methods were proposed in the literature [14],

including segmentation and recombination of time series EEG, segmentation and recombination of frequency domain EEG, and spatial analogy. Lotte et al. used Common Spatial Patterns (CSP) as a feature extraction method to compare the performance of several types of different mental imagery trials before and after artificial trials were added. The results showed no significant impact on MI trials. In this thesis, further work is being conducted on MI trials using Lotte's artificial generation methods. The objectives of this part are to investigate three different feature extraction methods combined with three different artificial data generation techniques, proposed by Lotte et al. The Bipolar montage-based Band Power (BP) and Time Domain Parameter (TDP) feature extraction methods were evaluated in addition to CSP. Although the published paper [14] demonstrates that the CSP feature extraction method achieves higher classification accuracy than the Band power and TDP feature extraction methods, the latter two methods have the advantage of using fewer electrodes. Moreover, the three artificial data generation methods correspond to the time, frequency, and spatial characteristics of MI trials, whereas the TDP, Band power, and CSP features are extracted from the time domain, frequency domain, and spatial filters of MI trials. Consequently, in addition to assessing the performance of artificial trials based on different feature extraction methods, the study also examines the relationship between the performance of artificial trials based on different data generation methods and feature extraction methods. Therefore, the research question of this section can be summarized as follows:

(1) How does the choice of feature affect the classification for the small size of the original dataset?

(2) Is there an interaction between the DA method and feature selection?

(3) How does the ratio of real and DA affect the classification accuracy?

(4) How robust are results when tested on different datasets?

In recent years, GAN-based domain adaptation (DA) has gained popularity. However, the DCGAN architecture is highly data-hungry, and acquiring useful information from limited calibration trials in the MI EEG dataset poses a significant challenge. This requirement for a large number of training trials contradicts the purpose of DA, which is to yield more useful artificial trials from a small number of initial trials. To overcome this challenge, we combined cropped trials training and transfer learning in the third study to maximize the existing dataset while training new subjects.

Another approach to improve the performance of MI BCI in this thesis is to combine transfer learning and GANs to generate artificial MI trials. Although several published papers on GANs have demonstrated the ability to generate MI trials, these GANs typically require preprocessing steps such as applying STFT on MI trials and using the resulting time-frequency image as the input for the GAN. However, in the third study, the raw EEG data is directly used as the input to avoid feature representation. This approach not only simplifies the input processing but also allows for more accurate and reliable classification of MI tasks. By eliminating the need for feature extraction, we reduce the risk of information loss and enable more efficient training of the classifier. Overall, this approach represents a significant advancement in the field of MI BCI research and has the potential to improve the accuracy and effectiveness of BCI applications. The objectives of this part are:

(1) To propose a GAN that can take the original raw MI trials as input.

(2) To solve the problem of a small training dataset for GAN.

(3) To show the results of combining transfer learning and GAN for MI trials

(4) To use a deep learning classifier to test the performance of the artificial trials generated by the above method.

By achieving these objectives, we aim to demonstrate the effectiveness of our proposed approach and its potential to improve the accuracy and reliability of MI BCI systems. This research represents a significant contribution to the field of BCI research and has the potential to benefit patients with disabilities by enabling more accurate and efficient control of assistive devices. The last study of the thesis focused on addressing the following research questions:

(1) How can GAN be applied for DA when only a small number of trials are available?

(2) How can transfer learning (T.L.) and GAN-based artificial trials benefit the CNN classification?

1.4 Research Contributions

The aforementioned objectives indicate that the thesis aims to explore the graphical factors that affect the mVEP and generate artificial trails based on different methods. The contribution of this thesis is summarized as follows to achieve these goals:

In the mVEP experiments, disparate research contributions were made between the offline and online experiments.

Regarding the offline mVEP BCI experiments, inspiration was drawn from other VEP-based BCIs such as SSVEP and P300. In addition to testing different colors of moving targets, the effects of various background colors, viewing angles, and target sizes on BCI performance were also investigated. While these changes had already been explored in other VEP-based BCIs, they had not been previously studied in the mVEP field. Furthermore, the experiment utilized eight electrodes, and the optimal number of electrodes for classification results was explored. The suitable electrode combinations for each participant were also identified, and the frequency of electrodes used in these optimal combinations was determined. Finally, the classification accuracy of different box positions was separately calculated. The accuracy of different boxes is compared and analyzed in each layout to find if people gave preference for box positions. This information can serve as a valuable reference for subsequent mVEP experiments.

In the online experiments, data from two different mVEP layouts were recorded and analyzed. In each layout, the positions of four boxes differed. One layout had boxes positioned at the top, bottom, left, and right of the screen, while the other had four boxes arranged in a row. Unlike the offline experiments, it was concluded that the placement of boxes did not affect mVEP BCI. This research is also innovative in the field of mVEP BCI.

In regards to MI BCI, traditional and deep learning methods were used for data augmentation(DA). When using traditional methods for DA research, Lotte et al.'s DA method was further explored. The MI EEG was transformed in the time domain, time-frequency domain, and spatial projection to generate artificial data. The effect of these generated data on improving the classification accuracy of MI BCI was studied, and not only the CSP feature extraction method was used, but the Band Power and TDP methods were also employed as feature extraction methods. Ultimately, it was found that with very little original data, the three DA methods based on TDP feature extraction had a significant impact on classification accuracy. Another conclusion was reached that DA methods were not related to feature extraction methods.

In deep learning-based domain adaptation (DA) methods, a novel deep convolutional generative adversarial network (DCGAN) was introduced. This network offers the benefit of dispensing with feature representation for EEG data and instead directly inputs filtered time-series EEG to generate EEG signals, rather than their features. Additionally, the challenge of needing a significant amount of training data for the generative adversarial network (GAN)

has been addressed by utilizing a combination of transfer learning (TL) and cropped training This approach enabled GAN research to be carried out with just 40 MI EEG trials per class. While statistical analysis of the results from 9 subjects revealed that the proposed DCGAN did not significantly improve the classification accuracy of MI, one of the subjects (S3) demonstrated a significant increase in classification accuracy (30%). Consequently, three subjects were chosen to compare and evaluate the generated artificial data with the original data. The reason for this subject's exceptional performance was discovered. Compared to the other two subjects, this subject exhibited not only variations in the strength of event-related desynchronization/synchronization (ERD/ERS) on the electrodes during left and right-hand MI but also notable dissimilarities in the patterns of ERD/ERS in the time-frequency feature maps between the left and right hands.

1.5 Thesis outline

Chapter 1 of this thesis provides a detailed introduction to the BCI system, which serves as the central topic of this thesis. Specifically, The BCI category that mVEP BCI and MI BCI belong to, which are pertinent to this thesis, are expounded upon. In the motivation section, the impetus driving this research is outlined, along with a redirection of research focus that transpired during my Ph.D. tenure. Subsequently, the three studies that comprise this thesis are introduced, along with their respective objectives. Lastly, the thesis outline is presented to offer a comprehensive overview of the entire paper.

Chapter 2 commences with an introduction to the background knowledge of motor imagery and deep learning. A comprehensive literature review on VEP-based BCI and MI BCI is subsequently presented to demonstrate the advances made in these domains. The application of convolutional neural networks, generative adversarial networks, and transfer learning in EEG research is then introduced, with numerous articles cited to highlight the strides made in these areas of research.

Chapter 3 focuses on the study of mVEP BCI, in which offline and online experimental data are separately analyzed to explore the influence of graphic factors.

In Chapter 4, further research is presented based on Lotte's DA method, utilizing TDP and band power methods for feature extraction in addition to CSP.

Chapter 5 proposes a new DCGAN and employs cropped training and transfer

learning to perform DA on time-series EEG signals. The generated fake MI trials are subsequently analyzed in the time and frequency domains and compared with real data using CSP filters.

Lastly, in Chapter 6, a summary and discussion of the methods used in the previous three chapters are presented. The limitations of the employed techniques in this thesis are also discussed, along with the outlook for future mVEP BCI and MI BCI DA work.

Chapter 2

Background and Literature Review

2.1 Nervous System

The human nervous system is divided into two parts: the Central Nervous System (CNS) and the Peripheral Nervous System (PNS). The CNS comprises the brain and spinal cord. The peripheral nervous system comprises the somatic and the autonomic nervous system [15]. The nerves of the somatic system that send signals from the brain are called motor nerves or descender nerves, and the nerves that send signals from various parts of the body to the central nerves are called sensory nerves or afferent nerves.

2.1.1 Central Nervous System

The Brain

The brain is the central organ of the human nervous system which consists of the cerebrum, the brain stem, and the cerebellum. These are all contained in, and protected by the skull bones of the head. The cerebrum is divided into the left and right hemispheres. Each hemisphere consists of four lobes which are the frontal, temporal, parietal, and occipital lobes. The hemisphere is interconnected by the commissural nerve tracts. The cerebral cortex has an outer layer of grey matter never cells, i.e. dendrites, which cover the white matter i.e. axons. The Brodmann area (BA) numbering system is widely used to describe the anatomical location of the cortex because Brodmann's nomenclature assigns numbers to different areas of the brain based on their cytoarchitectural organization (Fig 2.1). For the purpose of this Ph.D., we'll focus on the sensory-motor system and the visual system.

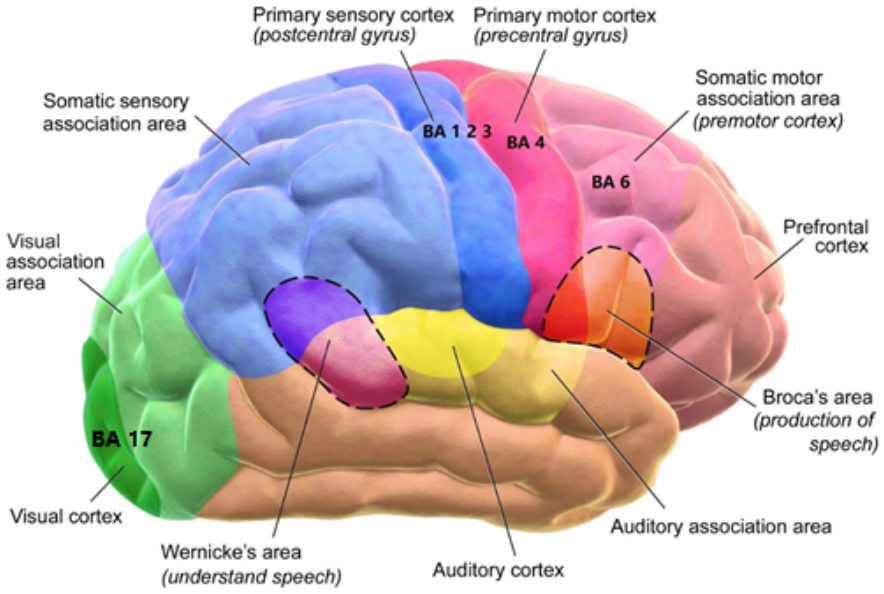


Figure 2.1: Main areas of the cortex with labeled BA corresponding to the motor cortex (BA4,6), primary sensory cortex (BA1,23) and visual cortex (BA17) [15]

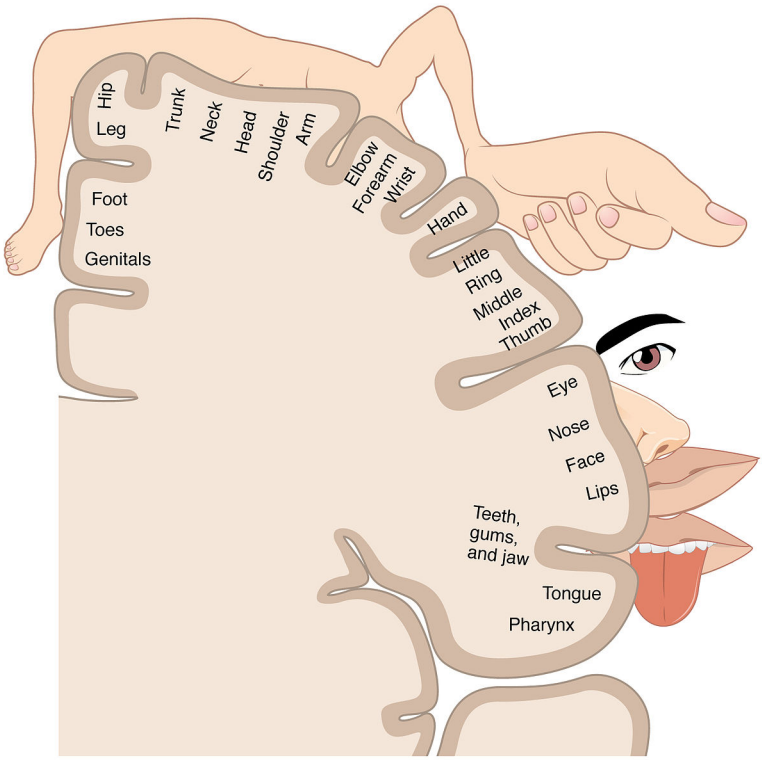


Figure 2.2: Somatotopic organization of the motor cortex [15]

2.1.2 Central Control of Movement

The Central Motor system is arranged as a hierarchy of control levels, with the forebrain at the top and the spinal cord at the bottom [15]. The central control system is divided into three levels. The highest level with the association areas of the cortex and basal ganglia of the forebrain is concerned with motor planning the goal of the movement and the movement strategy that best achieves the goal. The middle level comprises of the primary motor cortex and cerebellum and is concerned with tactics, which refers to the sequences of muscle contraction, arranged in space and time required to smoothly and accurately achieve the strategic goal. The lowest level consists of the brain stem and spinal cord. It is concerned with execution: activation of the motor neurons and interneuronal pools that generate the goal-directed movements. On the cortical level, the motor cortex comprises two BAs: BA4 corresponds to the primary motor cortex, and BA6 is located in the anterior position relative to the primary motor cortex (BA4) (Fig .2.1). BA4 projects its neurons directly to the spinal cord and to other subcortical structures and receives information from BA6 and from the primary sensory cortex (BA 1,2,3). BA6 consists of the Premotor Area (PMA) and the Supplementary Motor Area (SMA) playing an important role in the planning complex movements. Two PMAs areas are located laterally and SMA is centrally located. Neurons in the Brodman area 6 (both PMA and SMA) project to the BA4 and also directly to the spinal cord. 'Mirror neurons' found in PMA are important for motor learning and understanding actions of the other people. SMA is important for planning bilateral movements. Cortical control of movement (BA4 and BA6) has somatotopic organization, which means that certain parts of the body are controlled by certain parts of the cortex (Fig. 2.2). The presentation of the body in the cortex does not correspond to the physical size of the body, and areas of the body that require finer control (fingers, mouth) are controlled by larger areas of the cortex. In case of the paralysis or amputation of a certain part of the body, cortical areas controlling non-paralysed parts of the body tend to overtake the cortical presentation of the non-functional part. This is called disuse brain plasticity. Plasticity can also go in the reverse direction in the case of motor recovery [16]. Axons from the brain descend through the spinal cord along two major groups of pathways: the lateral pathways and the ventromedial pathways. The lateral pathways are involved in voluntary movement of the distal musculature and are under direct cortical control. The ventromedial pathways are involved in the control of posture and locomotion and are under brain stem control. There are two main spinal cord tracts in the lateral pathway, the corticospinal tract (also called the pyramidal tract) and the rubrospinal tract.

The corticospinal tract originates in the cortex and is the longest and the largest central nervous system tract. Two-thirds of the axons in the tract originate in BA areas 4 and 6 and most of the remaining axons in the tract derive from the somatosensory areas of the parietal lobe and serve to regulate the flow of the somatosensory information to the brain. The corticospinal tract's cross-section is roughly triangular when it is cut, this is the reason why it is called the pyramidal tract. At the junction of the medulla and spinal cord, the pyramidal tract cross, or decussates, thus right motor cortex directly commands the movement of the left side and the left motor cortex controls the muscles of the right side. The proper functioning of each level of the motor control hierarchy relies heavily on sensory information. In the highest level of the sensorimotor system, sensory information generates a mental image of the body and its relationship to the environment. At the middle level of the sensorimotor system, tactical decisions are based on the memory of the sensory information from past movements; while in the lowest level of the sensorimotor system, sensory feedback is used to maintain posture, muscle length, and tension before and after each voluntary movement [15].

2.1.3 Movement Imagery

Motion imagery (MI) is a mental simulation of real movement. MI is similar to physical motion execution (overt motion). The mechanism of MI is not fully understood, but studies have shown that regions of the motor cortex that are activated during overt motion are similar to areas activated during MI unlike a physical motion execution, MI is a covert process i.e. it does not result in the physical execution of motion. Covert operations also include movement observation, watching other people moving limbs, and movement attempts (which is possible in paralyzed people only). In the course of observing movements, the motor cortex is also activated in a similar way to MI [17]. Overt and covert movements have similar timing i.e. it takes a similar time to execute and imagine a movement. They also have similar biomechanics constraints, which means that one normally imagines movements that they can physically execute and finally both obey Fitts law, which states that the amount of time required for a person to move toward a target area is a function of the distance to the target divided by the size of the target [17] It is believed that MI also activates the cortico-spinal tract, which explains why MI can be used for the rehabilitation of movement. Ultrasound studies showed that in able-bodied people, MI is often accompanied by involuntary muscle twitches, i.e. it activates efferent motor pathways [18] There are two main types of MI, namely kinesthetic MI and visual MI. The kinesthetic type involves MI performed from the first-person perspective, where

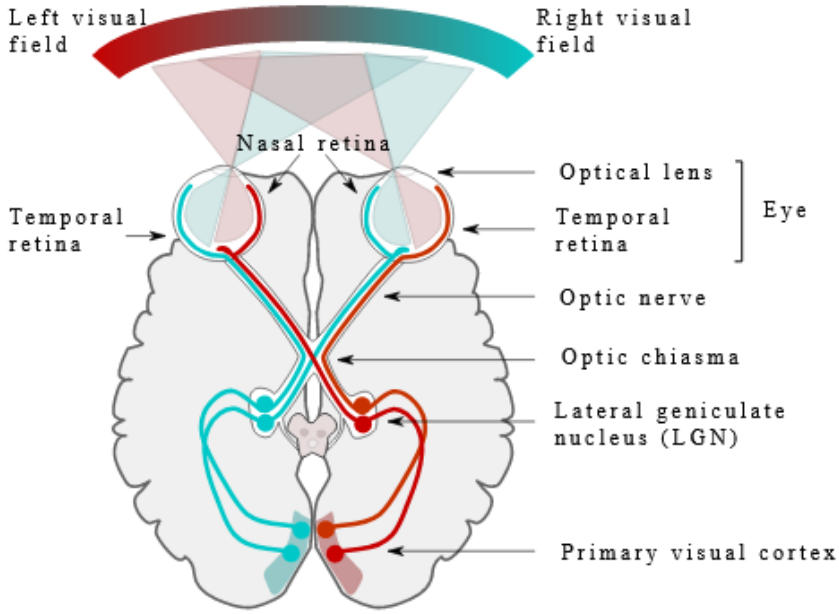


Figure 2.3: Neural pathways of the visual nervous system [20]

the person experiences kinesthetic or physical sensations related to the imagined motion. Execution of movement normally produces a stronger response than imagination/observation. The imagination of movements can be trained so that stronger and more reliable responses are achieved. On the other hand, visual MI is when the performer forms a visual image of an action or scene. It can be from a first or third-person perspective. Based on areas activated in the cortex, the former one is more similar to real movement [19]

2.1.4 The Central Visual System

The neural pathway that leaves the eye, beginning with the optic nerve, is called the *retinofugal projection* as it has the origin in the retina of the eye. The components of this system are the optic nerve, the optic chiasm, and the optic tract. The optic nerves exit from the left and the right eyes and form the optic chiasm, which lies at the base of the brain (Fig. 2.3). At the optic chiasm, the axons originating in the nasal retina cross from one side to another i.e. decussate contralaterally. Following this partial decussation, the axons form the optic tract.

The largest part of the optical tract projects to the left and right *Lateral Geniculate Nuclei* (LGN), located in the dorsal thalamus. Smaller projections innervate the hypothalamus and the superior colliculus in the midbrain. The LGN is organized into six distinctive layers, and different layers receive projections from different eyes. Cells in layers 2, 3, and 5 originate from the right eye, and

cells in layers 1,4, and 6 originate from the left eye. The axons from LGN project to the *primary visual cortex* (V1) also called the *striate cortex*, corresponding to BA17. The neocortex in general is divided into 6 layers. In V1, there are actually 9 layers but they are marked with numbers I to VI for consistency: they are layers I, II, III, IVA, IVB, IVCα, IVCβ. The largest number of axons from LGN projects to layer IVC. There are several major characteristics of the V1 receptive field:

1. **Binocularity**, meaning that neurons have two receptive fields, one for the ipsilateral eye and the other for the contralateral eye.

2. **Orientation selectivity** meaning that many neurons in V1 respond best to an elongated bar of light moving across their receptive field. The optimal orientation of the bar can be any angle 0-360o, for a particular neuron.

3. **Directional selectivity** meaning that V1 neurons respond best when a bar of light moves perpendicular to the optimal orientation in one direction but not in the other i.e. for example may respond strongly when the bar of light moves from left to right but not when it moves from right to the left.

4. **Simple and Complex receptive fields** Neurons in the LGN have antagonistic center-surrounding receptive fields and that affects how they are projected to V1. V1 neurons receive a converging input from LGN that is aligned along one axis, and that is called simple cells. Other orientation-selective cells in V1 do not have such distinct on-and-off regions and are considered complex cells.

Within V1, cells are selective for different orientation, direction of motion and colors, and it is believed that there exist three distinct pathways each having one of three distinctive functions. Beyond the V1 there are two large-scale cortical streams of visual processing, one stretching dorsally from the striate cortex towards the parietal lobe and the other stretching ventrally towards the temporal lobe [21]. The function of the former is visual analysis of motion and visual control of action and the function of the later is the perception of the visual world and object recognition. Within the dorsal pathway, there are two areas responsible for visual motion, so called “area V5 (also known as MT from Middle Temporal location) and Medial Superior Temporal (MTS) area. The function of area V5 neurons is perception of action, rather than seeing real action, and they can be activated by real action and by action illusion; its cells are sensitive to linear motion. Area MST has cells sensitive to linear motion, radial motion and circular motion.

The main functions of motion processing in V1, MT and MTS are believed to be navigation, directing eye movements and motion perception.

2.2 Neuroimaging Technologies

The activity of the brain even when person is completely relaxed and mind wondering produces a variety of biochemical and electromagnetic signal changes, which can be detected by various technical means. Existing relatively mature detection methods mainly include: Positron Emission Tomography (PET), functional Magnetic Resonance Imaging (fMRI), near Infrared Spectroscopy (nIRS), Magnetoencephalography (MEG), and Electroencephalography (EEG) [22].

These measurement techniques can detect the physiological changes of the brain's mental processes from different angles, and each of the measurement techniques has its advantages and disadvantages, making them more or less suitable for brain-computer interface applications. Among them, PET / fMRI measures metabolic images of brain functional activities. Their spatial resolution can reach the order of millimeters, but the temporal resolution is low, typically more than 5s. This is partially due to the recording technique but largest delay is due to natural metabolic processes. fMRI is used mainly for experimental BCI and it works in semi real time. In addition, these two types of imaging instruments are relatively bulky, costly and do not allow a participants to move.

An alternative is nIRS which is smaller, cheaper, portable and allow person to move their head. NIRS employs the infrared light to characterise fluctuation in cerebral metabolism during neural activity. It can measure cortical activity up to 1-3 cm depth. Its disadvantage is a relative low signal to noise ratio in persons with very dark hair. Functional near-infrared spectroscopy-based brain-computer interfaces (fNIRS-BCIs) depend on the body's metabolic processes, while another principal constraining factor is the delay related to the hemodynamic response in the brain. Presently, the literature lacks a well-defined temporal boundary for the onset of a noticeable alteration in the oxygenated state of hemoglobin following a specific neural stimulus. Furthermore, this inherent temporal delay exhibits considerable inter-individual variability [23–25].

EEG and MEG measure the direct electrophysiology response of brain neurons. EEG measures the electrical activity while MEG measures the magnetic activity. These two are closely related as electrical and magnetic fields exist at the same time and are perpendicular to each other. They all have millisecond-level time resolution to meet the requirements of real-time tracking of brain activity.

However, MEG device is very bulky and costly and do not allow for movement. Main problem with MEG is a need to cancel the magnetic filed of the earth. Portable MEG are still in experimental phases. In contrast, EEG's measurement technology is mature, the instrument is portable, the operation is simple, and BCI based in EEG can be used at patients' homes. The main disadvantage of EEG over MEG is a low signal to noise ration and necessity to wash hair in case that wet EEG electrodes are used.

2.2.1 Electroencephalogram

EEG is a result of joint activity of thousands of underlying neurons. EEG measures the current flow during synaptic excitation of the dendrites of pyramidal neurons in the cerebral cortex. The amplitude of the EEG signal is proportional to the number of synchronously activated neurons. However the signal measured on the surface of the skull has to pass through several layers of nonneural tissue (e.g. meninges, fluid, skull and skin). EEG will be affected by age, sensory stimulation, disease, body physiology, biochemical conditions, to name a few [26].

Application of EEG recording developed for various purposes. For long term monitoring, it is used to monitor brain activity during the anesthesia, vigilance state or cognitive load. On the other hand, it also helps with the diagnosis of clinical neurological diseases, such as epilepsy, brain tumor, or problems with sleep. Short term EEG recording is clinically used to detect brain injury, depression, the effect of drugs, to record evoked potentials, etc, In its conventional use, outwit BCI applications, EEG requires off-line analysis i.e. analysis once the EEG recording has finished.

EEG rhythms

Oscillatory brain activity is a results of synchronous excitation of a number of cells results in a large, rhythmic waves. There are two main sources of EEG activity: central clock maker (thalamus) and a distributed sources within the cortex, which excite and inhibit each other [1,6]. In case of thalamus as source of oscillatory rhythms, voltage gated ion channels allow each thalamic cell to generate self -sustaining discharge pattern thus synaptic connection between thalamic cells synchronize of the rhythm in the group. Thalamocortical axons pass this rhythm from thalamus to the cerebral cortex which can be recorded.

In general brain rhythms are characterized by distinctive frequency bands and amplitude, and typically the amplitude decreases with increase in frequency. The most common categories used in clinical and scientific research are [26]:

Delta rhythm: Its amplitude is characterized by a low frequency and high amplitude, about 20-200 μV and frequency is 0.5-4Hz, and is best observed over the occipital and the temporal lobes. Normal adults almost have no delta waves when awake, but it can be recorded during sleep and deep anesthesia. Delta rhythms can be seen in infancy or people with brain injury.

Theta wave: It is also a rhythm with low frequency and high amplitude. It has frequency between 4-78 Hz and the amplitude about 100-150 μV. It is most prominent in the occipital and parietal lobe and midline frontal theta can be observed in healthy adults during a task with increased cognitive demand. It is more common in adolescents and is generated in adults due to stress.

Alpha wave: It has the amplitude of about 20-100 μV, but mostly below 50 μV in adults, and a frequency range of 8-12 Hz, and occurs during awaken state over the posterior regions of the head, generally with higher voltage over the occipital areas. Widespred alpha rhythm can be recorded everywhere on the cortex. It is best seen with eyes closed and under conditions of physical relaxation and relative mental inactivity while blocked or attenuated by attention, especially visual, and mental effort.

Mu Rhythm: It is also called Rolandic or central rhythm. It is closely related to the activity of the motor cortex and is not easily observed in every adult in a relaxed state unless time-frequency analysis is applied. Similar to the alpha it is in the 8-12 Hz range. During motor action (movement) of a certain part of the body energy of the mu rhythm drops as compared to the relaxed state, that is a phenomenon called Event-Related Desynchronization (ERD). At the same time, mu rhythms of the surrounding areas may increase, which is called Even Related Synchronization (ERS). The ERS can also be recorded at the same areas as ERD, upon the termination of movement [27].

Beta wave: The frequency of beta rhythm is 13-30 Hz, with amplitude 5-20 μV. Beta rhythm is usually less regular than alpha rhythm and can be observed in different areas of the cortex. Beta rhythm located in the frontal area is fairly common and increases with increased concentration. Posterior beta is reactive like alpha rhythm to opening or closing eyes. A beta rhythm observed over the central areas of the cortex serves the same function as the mu rhythm, and together these two rhythms are called sensory-motor rhythms.

Gama rhythm: The frequency is above 30Hz, and the amplitude is less than 2 μV. It is a low-amplitude high-frequency rhythm often related to perception and consciousness and Rapid Eye Movement sleep (REM) stage sleep. During working memory tasks, it is frequently observed coupled with the theta rhythm.

EEG Analysis Method

EEG is a non-stationary signal with a low signal-to-noise ratio. For that reason, it requires noise removal and subsequent analysis in time, frequency, or time-frequency domain. including [28]:

1. Time domain analysis method to extract features directly from EEG data. This EEG processing method is still widely used because of its intuitiveness and clear physical meaning. The time domain methods mainly analyze and identify the waveform characteristics of EEG, such as amplitude, mean, variance, skewness, and kurtosis. Most frequent clinical applications of time-domain analysis are detected by the spike-and-wave complex in seizures and sleep monitoring. Commonly used EEG time-domain analysis methods are zero-crossing analysis, histogram analysis, variance analysis, correlation analysis, peak detection, waveform parameter analysis, and waveform identification.

2. Frequency domain analysis method. Frequency domain analysis is the main analysis method in EEG research and clinical application. The most frequently used method is to estimate the power spectrum of EEG. By calculating the EEG signal power spectrum, we can directly observe the distribution and changes of $\delta, \theta, \alpha, \beta, \gamma$ and other rhythms in EEG. This can be further extended unto different methods to calculate connectivity between different electrode locations with methods such as coherence, phase locking values, and various methods for estimated directed coherence [29].

3. Spectral estimation is analysis in the joint time-frequency domain. Most time-frequency analysis methods are based on spectrograms utilizing Fourier transform, with a fixed resolution in time and frequency domain or wavelet transform in which time-frequency resolution depends on the frequency, i.e., time windows are shorter in higher frequencies and vice versa. The advantage of these methods is that the physical meaning is clear, and the calculation is straightforward. The disadvantage is that when analyzing segments of EEG signals using the Fast Fourier Transform (FFT), the frequency resolution is very poor in EEG because it is not periodic, resulting in energy leaking into adjacent frequency bins. This phenomenon, known as "spectral leakage", causes the spectrum to become blurred and distorted [30].

Parametric methods are used for analysis or shorter EEG recording where estimation of power density based on FFT or wavelets would give poor results

because they require averaging over a number of smaller time windows (Welsh periodogram) to obtain a smooth graph. Unlike non-parametric methods, parametric methods do not directly depend on the sampling rate of the EEG recording. An advantage of the parametric model method is that the frequency resolution is high, and the obtained spectrum is smooth, which is beneficial to extract the features of EEG. In the spectral estimation of EEG signals, the most used parameter model is the auto-regressive (AR) model. A disadvantage of these methods is that they are dependent on the properly selected model and in the case of AR on the order of the model.

EEG Electrodes Nomenclature and Referencing

EEG is based on the surface recording from the skull. The position of the electrode is usually determined by the international standard system of electrode placement (including 10/20, 10/10, and 10/5 systems [31]). The location of the 10/20 standard system and its electrode locations are shown in Fig. 2.4. The name 10/20 comes from the distance between electrodes when measured between the nasion and inion in the sagittal line, between the ears across the coronal line, and around the head circumference. Each electrode location on the left and right hemisphere is characterized by a letter, corresponding to a lobe and the number, odd on the left and even on the right side. Electrode across the sagittal line has a letter z instead of the number. Electrodes across the coronal line have the same letter (except C and T in the central line) and electrodes across the same sagittal line have the same number.

Different electrode montages are used in the EEG recording and analysis. The most frequently used montages include unipolar, bipolar, Laplacian, and common average [32]. The monopolar derivation is the default montage of EEG recording, while the other montages are derived from the unipolar montage. Three electrodes are used to measure a monopolar EEG signal. The electrodes include an active electrode, a reference electrode, and a ground electrode. Two electrodes (the active and the reference) are used to record the difference in potential between two points and the ground electrode is used for safety and to remove the common noise. The reference electrode is typically placed on some point on the head (the ear lobe, mastoid bone, tip of the nose) which records minimal brain activity but can also be placed on the skull.

Bipolar montages are obtained by subtracting two unipolar derivatives, so the common reference is canceled and EEG presents a measurement between two active electrodes, both measuring EEG activity. Because bipolar montage provides the voltage difference between two unipolar montages, it attenuates

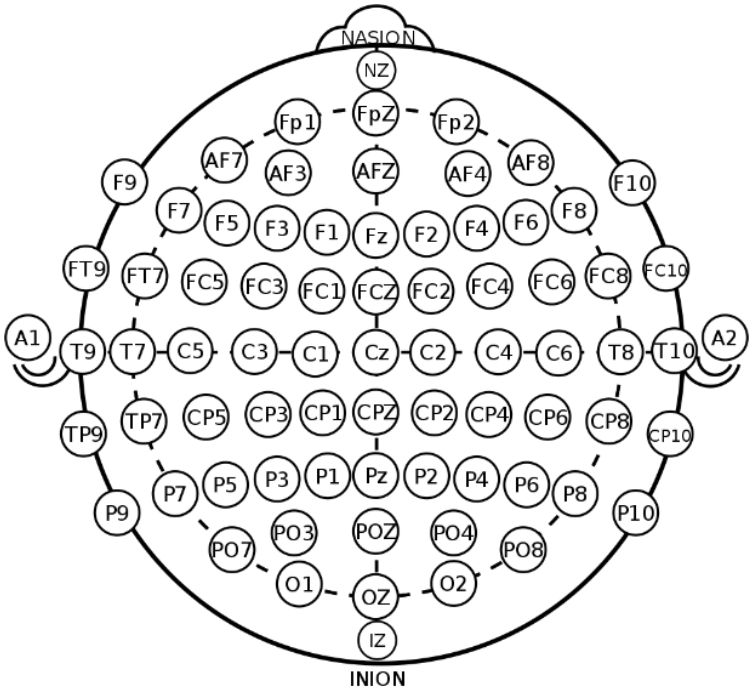


Figure 2.4: EEG electrode locations according to 10/20 system

common noise in two unipolar recordings.

The Laplacian montage is an approximation of a Laplace spatial filter [33]. To derive the Laplacian montage, the average EEG value from electrodes recorded on 4 surrounding electrodes is substrated from the central electrode. The distance of the surrounding electrode determines the size of the Laplacian filter. Usually, a small Laplacian filter is used, and the distance of the surrounding channels is 10% as defined by the 10/20 standard system. The Laplacian signal acts as a spatial filter and can greatly enhance local activity by subtracting the activity of distant sources recorded in both central and the surrounding electrodes.

Average reference(AR) is applied to reduce the common sources of electrical activity that could lead to distortion and ambiguity in the recording by subtracting the average activity of all electrodes from each individual electrode's signal.AR is also useful for removing the volume conduction effect and is used when electrodes are evenly distributed across the whole skull. For AR, volume conduction from some of the electrodes is averaged by all electrodes and subtracted from those electrodes, thus reducing the volume conduction.

2.3 Introduction to Convolutional Neural Networks

2.3.1 Deep Learning

Machine learning (ML) has recently become very widespread in research and incorporated into various applications. Algorithms such as logistic regression and Bayesian classification used in traditional ML already show good classification capability. However, these algorithms do not learn directly from the original data but from the feature representation from the original data, thus yielding the problem of feature extraction.

Given that the EEG sample data is $\mathbf{D} = \{\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_N\}$, where $\mathbf{D}_n \in \mathbb{R}^{C \times T}$ is a 2-dimension (2D) matrix representing the n-th trial of $\mathbf{D}$ for $n \in [1, N]$. C and T are the numbers of channels and samples over time, respectively. If we define $\mathcal{F}(x)$ as a function for feature extraction, the feature of $\mathbf{D}_n$ can be obtained by: $\mathbf{f}_n = \mathcal{F}(\mathbf{D}_n)$.Finding the optimal $\mathcal{F}$ becomes the major research direction in traditional ML.

Different type of features affects the EEG-based BCIs' classification accuracy. Researchers try to propose more effective feature extraction methods for specific problems. For MI BCI, there is Time Domain Parameter (TDP) feature in the time domain, Band Power (BP) feature in the frequency domain, and Common Spatial Pattern (CSP) in the spatial transform. They can be all regarded as one-layer feature representations. Such shallow layer feature extraction methods cannot solve complex problems. DL, however, has a multi-layer data representation architecture, in which the deeper layer learns the feature from the shallower layer. The relationship between DL and ML is shown in Figure. 2.5.

With the increasing popularity of big data and the improvement of Graphics Processing Unit (GPU) computing ability, DL has attracted huge attention in pattern recognition and machine learning. DL provides fast learning on the large volume of samples and good adaptability. It also provides a whole pipeline for feature extraction and classification without separating them as in traditional ML. Learning features automatically from the data by layers called end-to-end (data to class) learning avoids extracting features manually. The multi-layered network frameworks of DL preserve and calculate more details, which enable people to discover the potential feature connections based on its architecture. Researchers have applied DL to many fields, e.g. image recognition, speech recognition, and natural language processing. Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and GAN are very popular algorithms in the DL field.

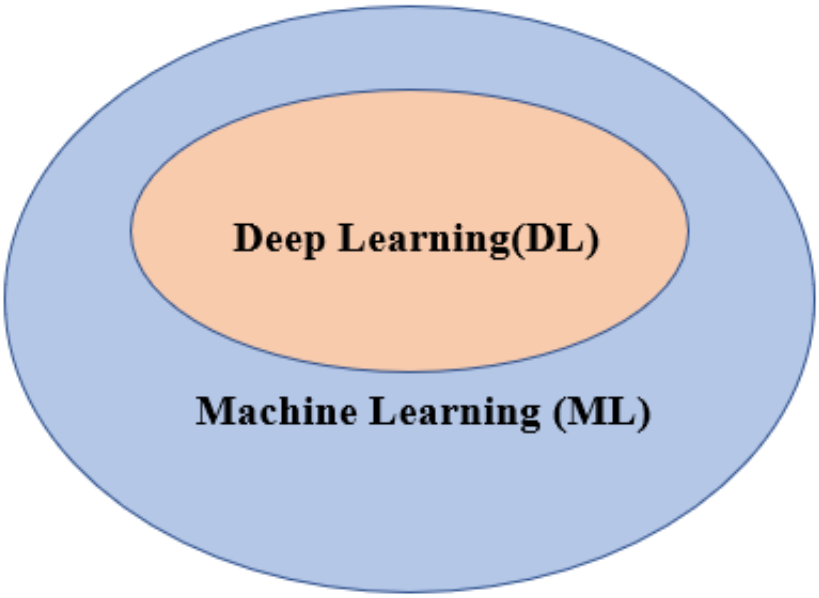


Figure 2.5: Relationship between ML and DL

CNN and GAN are the two main algorithms used in this study.

2.3.2 Structure of Convolutional Neural Network

CNN was inspired by the connection of neurons of the visual cortex of animals in 1962. Hubel and Wiesel found that cats' visual cortex cells are in an ascending hierarchy, and individual cortical neurons respond to specific stimuli called Receptive Fields (RF). Fukushima used the RF concept to develop the first CNN in 1984. After years of improvement and optimization, CNNs have become a golden standard in Image Classification, Object Detection, and many other recognition problems. In 1990, Le Cun et al. first applied the backpropagation algorithm to update the weights in CNN and achieved good classification results in the MNIST dataset [34]. In 2012, Krizhevsky et al. won the ImageNet competition using CNN. In March 2016, artificial intelligence algorithms developed from CNN and logistic regression algorithms enabled AlphaGo to beat the top human chess player. The Convolutional Neural Network is a commonly used artificial neural network in computer vision and speech recognition, known as shift invariant or space invariant. It uses convolution kernels (or filters) sliding along the input layer to generate new feature maps. The shared-weight architecture can reduce the number of weights, thus improving the network's efficiency.

Some BCI researchers have tried to apply CNN for EEG classification. Given the great success CNN achieved in the image processing field, researchers usually

use CNN analysis of image-like feature patterns. Generally, EEG signals are first transformed into time-frequency feature maps before being fed into CNNs. Piyush Kant. et al. use the Continuous Wavelet Transform (CWT) of MI EEG signals as features to feed into kinds of Deep Convolutional Generative Adversarial Networks (DCGAN) for TL and yield a validation accuracy of 95.71% [35]. People also tried to train CNN directly from raw EEG without extra transformation. The deep convert in two proposed CNNs achieved almost the same test accuracy as the Filter Bank Common Spatial Pattern (FBCSP) method with BCI Competition IV, dataset 2a, 67.8% [36]. Compared to using time-frequency feature maps as inputs, end-to-end learning can reduce the complexity of data processing.

The main advantage of CNN is its shared weight architecture, which can reduce the parameters in an artificial neural network and avoid overfitting. The structure of CNN can be divided into three main parts: an input layer, hidden layers, and an output layer. The input layer receives batches of training trails, where the Batch Size is a predefined parameter representing the number of trails fed into the CNN at once. Results from the output layer indicate the probability of each trail belonging to a certain class. The hidden layers consist of distinct layers, the Convolutional Layer, Activation layer, the Pooling layer, and the Fully Connected layer. Fig. 2.6 shows a typical CNN structure.

The Convolutional Layer

The Convolutional Layer is the most important part. It uses kernels (also called filters) sling along the input samples to calculate the convolution result and pass it to the next layer. The discrete convolution in math is explained here to understand convolution in CNN better. Assume the entry of a matrix $\mathbf{D}$ is $a_{i,j}$ for $i \in [0, m]$ and $j \in [0, n]$. The expression of $\mathbf{D}$ can be written as:

$$\mathbf{D} = \begin{pmatrix} a_{0,0} & a_{0,1} & \cdots & a_{0,n} \\ a_{1,0} & a_{1,1} & \cdots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,0} & a_{m,1} & \cdots & a_{m,n} \end{pmatrix}, \tag{2.1}$$

Define $\mathbf{F}$ as a 3×3 matrix by taking the first 3 columns and rows of $\mathbf{D}_n$, and $\mathbf{G}$ as another 3×3 matrix with each entry to be $b_{i,j}$ for $i \in [-1, 1]$ and $j \in [-1, 1]$. We can formulate f and g as follows:

$$f = \begin{pmatrix} a_{0,0} & a_{0,1} & a_{0,2} \\ a_{1,0} & a_{1,1} & a_{1,2} \\ a_{2,0} & a_{2,1} & a_{2,2} \end{pmatrix}, \tag{2.2}$$

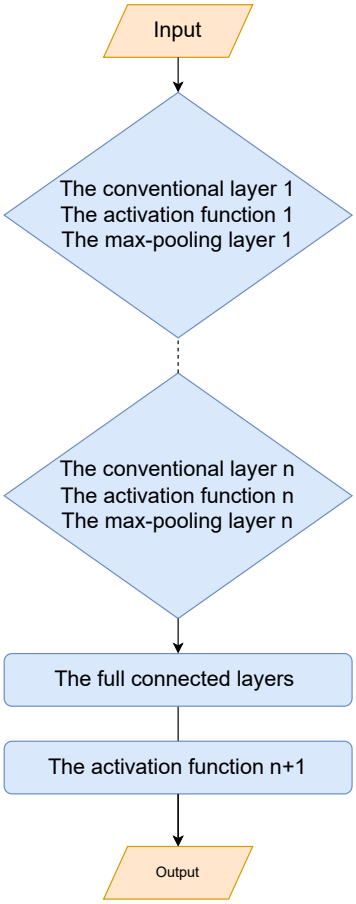


Figure 2.6: Basic structure of the CNN

$$g = \begin{pmatrix} b_{-1,-1} & b_{-1,0} & b_{-1,-1} \\ b_{0,-1} & b_{0,0} & b_{0,-1} \\ b_{1,-1} & b_{1,0} & b_{1,-1} \end{pmatrix}, \tag{2.3}$$

Discrete Convolution:

$$(f * g)(m,n) = \sum_i \sum_j f(i,j)g(m-i, n-j) \tag{2.4}$$

Taking $m = n = 1$ as an example, $f * g(1,1)$ can be calculated by:

$$\begin{aligned} (f * g)(1,1) &= \sum_{k=0}^{2} \sum_{h=0}^{2} f(h,k)g(1-h, 1-k) \\ &= a_{0,0}b_{1,1} + a_{0,1}b_{1,0} + a_{0,2}b_{1,-1} + a_{1,0}b_{0,1} \\ &+ a_{1,1}b_{0,0} + a_{1,2}b_{0,-1} + a_{2,0}b_{-1,1} \\ &+ a_{2,1}b_{-1,0} + a_{2,2}b_{-1,-1}, \end{aligned} \tag{2.5}$$

It can be seen that the product of each two elements is not coming from the corresponding position in f and g, but with g rotating 180° around the central element.

Convolution algorithm in CNN:

$$(f \otimes g)(m,n) = \sum_i \sum_j f(i,j)g(i,j). \tag{2.6}$$

Again, let us take $m = n = 1$ as an example, $f \otimes g(1,1)$ can be calculated by:

$$\begin{aligned} (f \otimes g)(1,1) &= \sum_{k=0}^{2} \sum_{h=0}^{2} f(h,k)g(h,k) \\ &= a_{0,0}b_{-1,-1} + a_{0,1}b_{-1,0} + a_{0,2}b_{-1,1} + a_{1,0}b_{0,-1} \\ &+ a_{1,1}b_{0,0} + a_{1,2}b_{0,1} + a_{2,0}b_{1,-1} \\ &+ a_{2,1}b_{1,0} + a_{2,2}b_{1,1}, \end{aligned} \tag{2.7}$$

This equation shows the difference of convolution in CNN from the math, where the two elements in the product come from the same position in f and g. The matrix g can be regarded as the kernel, and f can be regarded as the RF of the input $m * n$ matrix in CNN. $a_{i,j}$ is the neuron on the feature map $\mathbf{D}$. Since g is an unknown trainable parameter, there is no need to rotate it in CNN training. Each element in the kernel is called kernel weight. There are several methods to

initialize the kernel at the beginning of the CNN training. Generally speaking, it is to assign kernel weights random values. These values are later adjusted in training based on algorithms, e.g., BP. The $f * g(1, 1)$ can be regarded as a neuron for the next feature map. Usually, CNN contains several Convolution layers with different weights in the kernel of each layer. Such sharing weight structure in a layer significantly decreases the computing cost, thus increasing the efficiency of the complex artificial neural network.

Activation function

In a typical CNN, the value of neurons on a layer comes from the activation function. Namely, the sum of convolution adding bias from the previous layer needs to go into the activation function. The equation is shown as:

$$y_{i,j} = f(\mathbf{W}_k * \mathbf{I}(i, j) + b), \tag{2.8}$$

where $\mathbf{W}_k$ is the kernel weight of the kth convolution layer, $\mathbf{I}(i, j)$ is the RF on this layer, b is the bias value, and f is the activation function. $y_{i,j}$ is the neuron on the generated feature maps, which is also the input of the $(k+1)$th convolutional layer.

The activation function can perform a nonlinear transformation which enables the CNN to solve nonlinear problems. Commonly used activation functions in the CNN includes sigmoid, tanh, LeakyReLU, etc. The Sigmoid has an exponential shape with outputs in $(0, 1)$, represented as probability or used for input normalization:

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{2.9}$$

Tanh is also a very common activation function. Compared with sigmoid, its mean output is 0, making it converge faster than sigmoid, thus reducing the number of training iterations:

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}. \tag{2.10}$$

The Rectified Linear Units (ReLU) can keep the gradient from decaying when $x > 0$, thereby alleviating the problem of gradient disappearance:

$$ReLU(x) = \max(0, x). \tag{2.11}$$

The Leaky-ReLU is improved from ReLU. The difference is that Leaky-ReLU

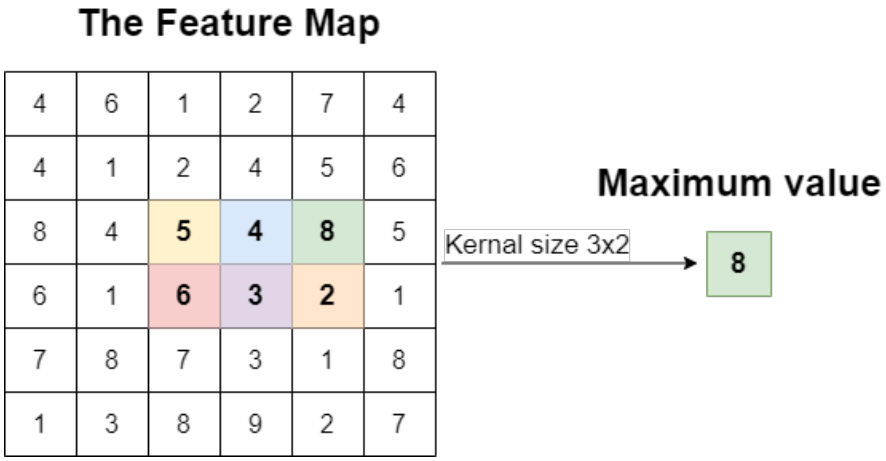


Figure 2.7: An example of max-pooling algorithm with Kernel size 3×2.

can avoid the gradient disappearance when $x < 0$:

$$f(x) = \begin{cases} x, & \text{if} \quad x \leq 0, \\ \alpha x, & \text{if} \quad x < 0, \end{cases} \tag{2.12}$$

The ELU has the same function as the Leaky-ReLU when $x \leq 0$, while the rest is nonlinear compared to Leaky-ReLU:

$$f(x) = \begin{cases} x, & \text{if} \quad x \leq 0, \\ \alpha(e^x - 1), & \text{if} \quad x < 0, \end{cases} \tag{2.13}$$

From these activation functions, it can be seen that the activation function decides whether or not to activate a neuron concerning a particular input by creating the corresponding output.

The pooling layer

The pooling layer is also periodically inserted between successive convolutional layers like the activation layer. It can gradually reduce the number of neurons on feature maps by subsampling, thus reducing the consumption of computing. More importantly, fewer parameters can effectively avoid the over-fitting of CNN. Commonly used pooling methods include average-pooling, min-pooling, max-pooling, etc. The pooling method used in this study is max-pooling which only retains the max value in each kernel-covered field. Defining the size of the kernel in the max-pooling layer is $k_1 \times k_2$, the equation can be written as:

$$f(x) = \max(a_{[i:i+k_1],[j:j+k_2]}). \tag{2.14}$$

Fig. 2.7 gives an example of a max-pooling algorithm, where $k1$=3, $k2$=2. Only the maximum value in this kernel, i.e., 8, can pass this max-pooling layer and become the new neuron for the next feature map.

The fully connected layer

Generally, there are often one or two fully connected layers at the end of the CNN. Fully connected layers connect every neuro point in one layer to every neuron in the next layer by calculating the dot product with weight plus bias. The equation can be presented as:

$$y_l = f(w * y_{l-1} + b), \tag{2.15}$$

where $f(x)$ is the activation function of the neuron, b is the bias,w is the weights and l is the order of layers.

2.3.3 Training Process of Convolutional Neural Network

The signal is propagated forward through the layers in the network until the last layer gives the classification results. Then the error between the results and the expectation is calculated, called loss. The error BP algorithm corrects the weights to minimize the loss by periodically repeating the signals forward and error backward. The loss function used in this thesis is the Mean Square Error (MSE).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \check{y}_i)^2. \tag{2.16}$$

where n is the number of data points,y_i is the observed values and $\check{y}_i$ is the predict values.

During training, the principle is to keep the loss function decreasing. The gradient descent algorithm is used to find the optimal weights that minimize the loss function. Equation 2.17-2.18 shows the adjustment process.

$$\mathbf{W}_i = \tilde{\mathbf{W}}_i - \eta\frac{\partial E(\mathbf{W}, b)}{\partial \mathbf{W}_i}, \tag{2.17}$$

$$b_i = \tilde{b}_i - \eta\frac{\partial E(\mathbf{W}, b)}{\partial b_i}, \tag{2.18}$$

where η is the predefined learning rate,E is the prediction errors, $\frac{\partial E(\mathbf{W},b)}{\partial \mathbf{W}_i}$ and $\frac{\partial E(\mathbf{W},b)}{\partial b_i}$ are the partial derivative of errors.

2.4 Visual Evoked Potential

The VEPs are evoked electrophysiological potentials that can be extracted from EEG activity recorded at the parieto-occipital cortex using signal averaging. VEP may be caused by various stimuli, including colored, spatially localized, or rapidly moving stimuli [37]. Motion VEP (mVEP) is a subtype of VEP that can be used for BCI, similar to Steady State VEP and P300. SSVEP and mVEP are both based on detecting short-latency VEP peaks. SSVEP is more frequently used because it is technically easier to create as it does not require precise time synchronization between a stimulus and the EEG response [38].

When the human eye pays attention to periodic visual stimuli with a frequency greater than 4Hz, it will produce in the brain a periodic response of the same frequency called the steady-state visual evoked potential [39, 40]. It has a stable and repeatable response in FFT spectra on the same frequency as the visual stimulus frequency. Although it requires averaging several responses, it has a relatively high signal-to-noise ratio and requires a relatively small number of repetitions to produce reliable BCI features [41,42]. An SSVEP-based BCI system was designed by Middendorf et al. [43] where two flashing buttons are presented on the computer screen, and the user only needs to look at one of the buttons to select the target one. SSVEP is the physical response of the primary visual cortex to visual stimuli, and it is mainly concentrated in the occipital region of the brain.

The motion-onset VEP is induced in BCI applications by a fast-moving bar. A typical mVEP contains three main components, P1 around 130ms, N2 around 160ms-200ms, and P2 around 240ms. Among them, N2 is the most important component for BCI feature extraction. Researchers from Tsinghua University first designed a stimulus paradigm to induce mVEP, thus proposing mVEP-BCI [44]. Since no flickering light is required to induce mVEP, it is more suitable for long-term use than SSVEP, avoiding visual fatigue.

The P300 VEP signal is an endogenous evoked potential related to cognitive function, discovered by Sutton et al. in 1965. The most classic P300-BCI was proposed by Farwell and Donchin in 1988 [45]. Classical P300 waveforms can be obtained in the Oddball stimulation paradigm, a series of stimuli consisting of two types of stimuli: standard stimuli with a high probability of occurrence (85%) and target stimuli with a low probability of occurrence (15%), and order of appearance of the two stimuli is random. When the subject looks at the target stimulus, the sudden stimulus will induce an observable positive wave in the subject's EEG about 300ms later. Visual, auditory, and tactile stimuli can

induce P300. In a typical P300 BCI speller, characters are arranged in a matrix. It flashes randomly by row or column, and P300 is evoked when subjects fixate on the highlighted row or column that contains a selected letter. The intersection of the evoked column and row is the target character that the subject focuses on. P300 BCI has been used by ALS patients with stroke and paralyzed patients for communication [46–50]. The biggest advantage of P300-BCI (as well as of all BCI based on VEP) is that it does not require training, and even first-time users can achieve good classification accuracy. However, one of the major flaws is that only novel stimuli induce the P300 potential, and once the user is very familiar with it, it will lead to the decline of BCI control performance.

Cano et al. run two experiments to explore the P300's amplitude variation using different pictures as the stimuli [51]. In experiment one, they compared the results between normal pictures in color and black/white. The second experiment compared the picture with small, medium and large scrambled conditions. The results suggested that stimulus color contributes to ERP valence effects. In another study performed by Bekdash et al., they used Red, Green, Blue (RGB), and yellow to elicit P300 under high/low-level intensity [52]. They discovered that the blue color always led to a low response in the visual occipital region which was a weak source of attention compared to green and yellow. The color intensity level also remains a major factor affecting the P300 latency, considered a novel and important observation.

Zhang et al. also performed color modulation research on P300, where they used a Red Face with a White rectangle (RFW), a Red Face with a Blue rectangle (RFB), and a red face with a red rectangle (RFR) to train the Bayesian linear discriminant analysis (BLDA) classifier [53]. The results suggested that RFW had the highest classification accuracy($p < 0.05$). Researchers also found that the Graphic structures of the humanoid robots affect P300 Potential [54]. Another research suggested that even the Alphabet Matrix Layout in P300 Speller may affect its performance [55]. They revealed that the shorter the inter-character distance, the larger the P300 amplitude, while no better system performance was found. A commercially available P300 BCI has letters overlaid with the faces of famous people to increase the P300 amplitude and increase the system classification accuracy [56].

After comparing P300- and SSVEP-BCIs for patients with locked-in syndrome, Combaz et al. (2013) determined that SSVEP-BCIs are faster, more accurate, less mentally demanding, and more satisfying than P300-BCIs [57]. Zhao et al. have found the P300 model outputs fewer control commands than the SSVEP model; however, the P300 model(8.6s) requires a longer execution

time than the SSVEP model(5.4s) in the telepresence control of humanoid robots [58]. However, because the flickering stimulus is an uncomfortable stimulation method, the subjects will inevitably experience visual fatigue [59] and a high mental load [60] under long-term gazing. To increase the signal-to-noise ratio, thereby decreasing the number of averaging repetitions (and increasing the transfer rate) and improving the classification accuracy, previous studies looking at the influence of color on VEP-BCI. Research has shown the effect of color on P300 [61–63]. Some studies reported a correlation between stimulus color and SSVEP amplitude. In 1966 Regan discovered that red light-induced SSVEP has the most significant response at 11 Hz while SSVEP intensity drops rapidly at neighboring frequencies [64]. He also found that blue light evoked the strongest SSVEP peaked at 13 Hz while yellow light-evoked SSVEP was weakest and insensitive to stimulation frequency. The study by Arakawa et al. showed that the second and fourth harmonics of SSVEP have a different response between monochrome and color background [65].

Some research groups investigated which stimulus color corresponds to the best system performance. For example, Bieger et al. found that white light stimuli had the highest information transfer rate compared to green, red, and blue light stimuli [66]. In a study by Cao et al., they found that BCI system performance corresponding to blue, green, red, gray, and white light stimuli increased sequentially corresponding to colors [67]. The study by Singla et al. showed that SSVEP evoked by violet light has a higher classification accuracy when compared to green, red, and blue stimuli [68]. Four different stimulation frequencies (8, 11, 13, and 15 Hz) of SSVEP BCI were investigated by Tello et al. under the effects of red, green, blue, and yellow light stimuli. The results showed that the red stimulus yielded the best classification accuracy and information transfer rate.

In the case of mVEP, researchers focused on the influence of GUI graphical components on VEP. Beveridge et al. explored the effect of graphic complexity on the performance of mVEP-based BCI overlaid on 3D computer games [69]. They used five different backgrounds for five targets mVEP BCI indicating the five levels of graphic complexity. Ten subjects' mVEP patterns were extracted from 12 Channels Cz, TP7, CPz, TP8, P7, P3, Pz, P4, P8, O1, Oz, and O2. The average Leave One Out (LOO) offline testing results across ten subjects indicated that the BCI performance of the most complex level of the background had a very slight decrease compared with the other four levels. However, the ANOVA analysis suggested no significant difference. This result proved that when the complexity of mVEP's background increases, it had no significant impact on mVEP BCI's

performance, making mVEP BCI a promising application in complex 2D and 3D games.

In another study by Beveridge et al., they replaced the simple 2D and 3D game graphic background with a commercial-grade graphic with seven levels of complexity [70]. The electrode locations and mVEP targets used were the same as in the previous study. The LOO offline testing results for 7 levels were 76.1%, 74.5, 74.5%, 71.6%, 78.7%, 73.6%, and 72.2% respectively. The ANOVA test performed between the best (level 5) and the worse (level 4) was statistically significant ($p < 0.01$). However, a further ANOVA taking into account all game levels returned a value of $p = 0.18$, suggesting that the difference in accuracy for all levels was not statistically significant. After testing different game genres, they concluded that graphical complexity alone does not degrade the mVEP accuracy. Other basic properties of video games, such as primary colors, dynamic character movement, flashing imagery, and pace, might influence results. Marshall et al. applied a five-target mVEP on five different games, indicating that the players could control the games with an average online accuracy of 71% [71].

2.5 Classifiers and Features for Motor Imagery-based Brain-computer Interface

Brain-computer interfaces can enable users to communicate with the environment without depending on conventional muscle pathways. Thus researchers proposed many use scenarios based on various kinds of BCI systems. This thesis's main topic is motor imagery-based BCI which provides a promising neuro-rehabilitation application for people affected by stroke or spinal cord injury. The principle behind the application is based on EEG patterns. A short-lasting rhythmic activity that appears with a decrease in the amplitude of cortical activity during motor action(executed, attempted, or imaged) is called ERD. On the contrary, ERS is a transient increase in the power or amplitude of brain waves. The patterns of ERS and ERD can be observed by examining the EEG signal in the frequency domain, using spectral analysis techniques. Specifically, changes in power or amplitude can be identified within certain frequency bands, in response to a stimulus or event. These patterns are believed to reflect changes in cortical excitability and inhibition, bringing repeatable features as input for MI BCI systems. MI-BCI is a widely used type of BCI applied for rehabilitation of movement in patient population [72] or for control of assistive devices such as wheelchairs or exoskeletons [73, 74]. Typically data are initially collected offline,

without feedback to create a classifier. This classifier is later used online, with feedback, and some initial calibration is often required at the beginning of each online session. Offline BCI simulation with cross-validation is a valuable tool for testing a new algorithm. Offline classification algorithms have traditionally been calibrated through supervised learning using a labeled dataset.

On-line MI BCI is often implemented in a cue-based paradigm [75], to determine when a person starts imagination to establish a pre-MI baseline period. Asynchronous MI BCI, which continuously monitors brain activity, needs to solve two problems: detect the onset of movement (to classify between the baseline and resting state) and, in the case of multiclass problem, detect which type of MI has been implemented [76] (classify between different types of movement).

Classifiers used in BCI, including MI BCI, can be broadly divided into linear and non-linear, where linear typically has a simpler structure [77]. While non-linear classifiers are in general capable of creating better separation between classes, they typically have a more complex structure and therefore require more adaptation between different data sets (e.g., offline and online datasets of the same person). More complex classifiers with a larger number of input features could achieve better classification results, but they typically require a larger dataset. That is because the number of training data needed increases exponentially with the dimensionality of the feature vector [78]. A stable classifier is characterized by stable low bias and variance. For that reason, both specificity and selectivity should be presented in BCI studies, and statistical analysis should be performed when comparing two BCI classifiers to account not only for the average accuracy but also for the variance.

The most frequently used classifiers for MI BCI [79,80] are Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), k Nearest Neighbors, Naïve Bayesian, Fuzzy Classifiers, Regression trees and various structures of recurrent neural networks. Currently, most MI BCI uses LDA and SVM due to their simple structure (LDA is linear, and SVM can be linear or non-linear depending on the kernel) and ability to achieve good classification accuracy with relatively small datasets [79]. Both classifiers have been implemented in the online MI BCI paradigm, i.e., they have been successfully used in real-life applications.

Most frequently used features for MI BCI are in the time domain (Hjort parameters, AAR, Fractals), frequency features (e.g., band power), time-frequency (e.g., wavelets), and features that use spatial filters combined with the time-frequency domain (e.g., filter band Common Spatial filters)

Very often, feature reduction needs to be applied to multi-channel features such as principal component analysis or evolutional algorithms (for a review of

features and features reduction methods, see [79].

This literature review will focus on deep learning classification methods, as they have been applied in this thesis. Deep learning based on convolutional networks has been increasingly used in BCI, including MI-based BCI. For example, Xiao and Fang reported a specificity of 0.93% and sensitivity of 0.96% on the MI BCI competition dataset (the authors did not specify which competition) [81]. However, most classifiers have been tested offline and on BCI competition datasets. Li et al. [82] extracted the dependency and the temporal features and passed them to CNN (CP-MixedNet) and achieved an average accuracy of 74.6% on BCI competition IV dataset 2a. Schirrmeister et al. [83] designed a Deep CovNet and a Shallow Covent achieving the accuracy of 70.1% and 60.8%, respectively, on BCI competition IV dataset 2a time series data. For 2D image inputs, Alazrai et al. employ a Quadratic Time-Frequency Distribution (QTFD) to transform the EEG signals into 2D time-frequency images and the classification results at 73.7% and 72.8% for able-bodied and trans-radial amputated subjects' same hand MI respectively [84]. Tabar and Halici transformed the EEG signals into a 2D image using STFT and used CNN and Stacked Autoencoders (SAE) as the classifier. The testing results for BCI competition IV datasets 2b suggested a 9% increase over the baseline [85]. Lee et al. also constructed a 2D image as the input of the CNN classifier but used the CWT. The classification results of BCI competition IV dataset 2b proved that the CWT transformed images(83%) outperformed the STFT ones(74.8%). Other studies also compared the classification accuracy of CNN and traditional ML classifiers. The STFT transformed 2D image feature showed better classification performance while using CNN rather than SVM [86]. Another study used CNN-extracted spatial features and LSTM for temporal features of MI data [87]. The classification results also proved that the DL-based framework outperformed the SVM. In the study, [88], Temporal and spatial features extracted using CNN and autoencoders showed better performance over the traditional ML, which used FBCSP as features and naïve Bayes, LDA, and SVM as classifiers.

BCI competition IV dataset 2a is the main dataset used in this thesis, so it is necessary to present the classification results base on this dataset. In the study [89], researchers combine the CNN and the Long-term, Short-term Memory Network (LSTM) to classify the preprocessed FBCSP feature yielding a test classification accuracy of 83%. while in study [90–95] which also used time-series EEG as input, the test accuracy was only 75.7%, 75.9%, 74.6%, 74.5%, 67%, and 70.9% respectively. The study [96] used Discriminative Filter Bank Common Spatial Pattern (DFBCSP) as the feature fed into the CNN, yielding average

test accuracy of 85.04%. Study [97–100] also used the CSP feature as the input of the CNN, yielding classification accuracy of 83%,80.44%,73.56%, and 79.9%, respectively. Zhao et al. used 3D images to represent the EEG signals, and the classification results presented a kappa value of 64.4% [101].

2.6 Data Augmentation for Motor Imagery Brain-computer Interface

In order to achieve high accuracy, BCI needs to be trained and re-trained on a large amount of data to avoid overfitting. Unfortunately, collecting this data is time-consuming and often tiring for participants. Data Augmentation (DA) is a promising method to help improve the performance of classification accuracy in the circumstance that there are no adequate trials available. Besides, MI-BCI is highly subject-independent, which means that psychological and physiological factors result in variations in MI feature distribution. Transfer Learning (T.L.) can make use of existing data which is another solution for data Insufficiency. Since the EEG feature has a specific pattern in the time of frequency domain, the traditional DA method focuses on EEG's time domain, frequency domain, and spatial properties. Lotte et al. propose three different DA methods for MI data. The first method is to divide raw EEG into small time segments for each trail and randomly recognize them into new synthetic trails of the same class. The second method performed similar recombination in the frequency domain after the STFT. Last but not least, they project the raw EEG to the principal component space and calculate the analogy of two original trails, then project the third trail back according to the analogy. The testing results of the BCI competition IV dataset 2a with 22 EEG channels show no significant improvement in classification accuracy. However, for another dataset based on a mental task, these methods show a significant increase, especially when minimal training trials are available. Zhang et al. have proposed a DA method based on the decomposition of real EEG signals, which is called Intrinsic Mode Functions (IMFs). Then the IMFs of the specific class were mixed together, and artificial trials were created based on these IMFs. They used 12 channels from dataset 3 of BCI competition II. These artificial trails result in the Wavelet Neural Network classification accuracy improvement for Subject1(74.2% to 83.3%) [102]. Bashivan et al. transformed the EEG signals into 2D image-like features using FFT and randomly added noise to the image to the features. The results show no classification accuracy increase using CNN classifier [103]. Another study related to adding noise to EEG signals

was performed by Wang et al., where they added Gaussian white noise directly to the time-series EEG data. The classification accuracy increased with ResNet from 34.2% to 75% and with LeNet from 49.6% to 74.3% [104]. Although the DA methods used in the study [103] and [104] were not based on MI data, Pei et al. performed a channel-level DA method called Brain-Area-Recombination (BAR) and compared the BAR DA with the noised addition DA methods on MI data [104]. They divide each trail into two parts (left/right brain channels), then the generated trails were the recombination of the segmented raw EEG signals within each part. The classification accuracy of EEGnet on BCI competition III datasets 4a shows significant improvement ($p < 0.01$) when compared to the baseline results, the BAR DA method also outperforms the noise addition method. Pervan et al. also performed the adding noise DA method. They used Gaussian noise with zero mean and deviation of 0.5, which resulted in the 0.07 classification accuracy improvement in kappa coefficiency while tested on BCI competition IV datasets 2b [105].

2.7 Generative Adversarial Networks Based Data Augmentation

Generative adversarial networks (GANs) are a type of deep learning framework that incorporates a generator and a discriminator to solve the problem of generating synthetic data samples that are similar to a given dataset. GANs were first introduced by Goodfellow et al. in 2014, and have since emerged as the state-of-the-art technology for data augmentation (DA) in various applications [106].

The basic architecture of GANs consists of two neural networks that work in tandem: a generator network and a discriminator network. The generator network produces synthetic data samples, while the discriminator network tries to distinguish between real and synthetic samples. Both networks are trained together, with the aim of the generator producing data samples that are indistinguishable from real samples(Fig. 2.8).

GANs have shown remarkable results in various application domains, including image generation [107], text generation [108], video synthesis [109], and music generation [110]. In the field of computer vision, GANs have been used for image-to-image translation, such as converting grayscale images to color [111] or converting daytime images to nighttime [112]. They have also been used for generating high-quality, photorealistic images of faces [113], landscapes [114], and other objects. GANs have also been applied in the field of natural language

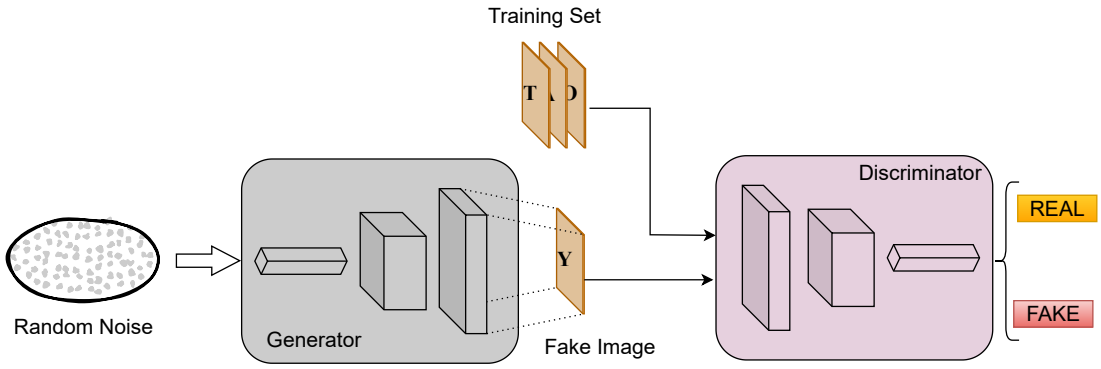


Figure 2.8: GAN Structure.

processing for text generation [115], text-to-image synthesis [116], and language translation [117].

One area where GANs have shown significant potential is in the analysis of electroencephalogram (EEG) signals. EEG signals are a measure of the electrical activity of the brain and are widely used in clinical settings for diagnosing and monitoring various neurological conditions. GANs can be used to generate synthetic EEG signals that closely resemble real EEG signals, and can be used to augment datasets or generate data for training deep learning models. GANs have also been used to generate synthetic EEG signals for brain-computer interfaces (BCIs), where they can be used to train and evaluate BCI systems in a more efficient and cost-effective manner. In the field of MI BCI, for instance, it is often difficult to obtain large amounts of training data due to the cost and time required for data collection. In this context, GAN-based DA can generate synthetic data samples that supplement the available training data and improve the model's generalization capability. In a recent EEG GAN study, Fahimi et al. used a conditional DCGAN to do cross-subject T.L. data augmentation. They then used CNN classifier to test the results of data with augmentation and found significant accuracy improvement [118]. The classification classes were 'diverted right-hand movement' and 'focused right-hand movement'. and this study aimed to find how BCI performance could be affected when people are distracted. Zhang et al. proposed the use of DCGAN for MI data augmentation. They appended the artificial data to the original data to train a CNN classifier. This study didn't involve transfer learning. Thus, for each MI class, they used a large number (up to 360) of trials [119]. They used the STFT EEG feature from C3, Cz, and C4 of BCI competition datasets 1 and 2b. The classification accuracy was up to 83.2% and 93.2%, respectively. In a conference paper, Conditional Wasserstein Generative Adversarial Network with Gradient Penalty (CWGAN-GP) was used to synthesize EEG data for MI data augmentation [120]. CNN

and TL. were used to classify BCI competition IV dataset 2b MI data. This TL method led to a 0.06 improvement in terms of the kappa coefficient compared to the baseline [121]. Dai et al. [122] proposed a hybrid-scale CNN working on time-series EEG and combined with DA and achieved the test classification accuracy at 91.57%. Ozdenizci et al. proposed an adversarial inference approach to decrease the EEG variabilities among subjects [123]. Ming et al. proposed a subject adaptation network inspired by GAN to align the distribution of different subjects [124].

Although GANs have proven to be effective in generating artificial data for various applications, they also face some limitations and challenges. One of the major challenges is the instability of the training process, which can result in mode collapse and limited variety in generated samples [125]. Additionally, evaluating the quality of generated samples remains difficult as there are no clear objective metrics to measure GANs' performance. Furthermore, GANs require large amounts of data and computational resources for training, which can be a bottleneck for some applications. For example, in MI BCI, subjects often become exhausted after around 100 calibration trials, which is far fewer than what a typical DL network requires (normally in the thousands). To address this issue, many DA methods have been proposed, with GAN being a popular approach for EEG DA. However, generating thousands of trails from just 100 can potentially cause overfitting. To mitigate this, TL can be used, which involves using the same task dataset from other people. Combining TL and GAN DA can be a promising approach to avoid overfitting and ensure the quality of artificial data. In this thesis, DCGAN is used for MI EEG DA.

2.8 Transfer Learning

Transfer Learning is another solution to reduce the calibration demand for a new BCI user. TL can reduce calibration efforts by utilizing relevant data/subjects/sessions/tasks. For the CSP feature, which is commonly used in MI BCI, researchers applied the weighted or averaged covariance matrix from existing subjects to the target subjects [126–128]. Heger et al. transformed the feature of existing datasets to the target feature space to enable BCI calibration when limited data are available [129]. Arvaneh et al. proposed a subject-to-subject adaptation algorithm that decreases the target subject's calibration time to 3-4 minutes. This approach is called feature-representation T.L. There are also other T.L. approaches that directly use the sample from the exiting data samples, which are instance-based TL [130]. Kambara randomly selected trials from the

available datasets to train the classifier. Instead of random selection, Hossain et al. developed an active T.L. algorithm that was able to find the most relevant samples to avoid trials with poor performance.

Chapter 3

Effects of Graphical factors on Motion Onset Visual Evoked Potential Brain-computer Interface

3.1 Introduction

A VEP (visual evoked potential) measures the electrical potential from the visual cortex upon the presentation of specific visual stimuli. A VEP BCI could be classified as a reactive BCI as it is based on features derived from brain activity produced as involuntary reactions to external stimuli. Commonly used visual stimuli consist of flickering or moving targets where a computer monitor or external source of light is used to evoke VEP.

Generally, a system with more distinctive sources of VEP can achieve a higher information transfer rate. However, a BCI with multiple VEP sources can be visually exhausting for users and can decrease the speed of BCI online response. SSVEP is a signal that is a natural brain response to a flashing visual stimulus in the frequency range of 4 to 100 Hz, with frequencies between 10 and 80 Hz being more commonly used in research and applications [131]. Under 6 Hz, the response is not steady-state and cannot be detected by the Fourier transform of EEG signal [132]; however, stimuli under 6 Hz with a checked pattern are called Transient Visual Evoked Potentials(tVEP) and are widely used in clinical practice to test the integrity of the visual neural system [133–135]. Existing SSVEP-BCIs mostly require eye gazing at a target for a long period of time. Traditional flash or pattern reversal VEP-based BCI is often accompanied by

visual fatigue. In addition to achieving better SSVEP response, typically, external sources of flickering lights are used (because a computer screen also has a flashing rate), which increases the complexity of setting up additional hardware for a BCI system. On the other hand, motion VEP (mVEP) BCI is elicited by the motion of a visual object and is not sensitive to contrast and luminance. mVEP can be easily implemented on a computer screen but requires more complex BCI architecture, including a time synchronization block, as explained later in the text.

The P300 is another type of VEP. Its event-related potential is a positive voltage peak of an event-related potential linked in time with a rare event that initiates sensory and mental processing [136]. By focusing the attention on the rare target stimulus (e.g., by keeping a mental count of its occurrence), the P300 amplitude can be increased and therefore classified more easily. The P300 has a delayed positive peak that occurs approximately 300ms, which is larger than the visual VEP delay because it includes higher visual and cognitive processing, i.e., unconscious image recognition.

To create either SSVEP or P300, visual stimuli involve pattern reversal, flashing lights, or symbols. On the other hand, mVEP allows for elegant stimuli that have simple visual feedback, which does not require much concentration and is not fatiguing to the eyes, thus can be used for a prolonged period of time. This is of relevance for gaming applications where users may spend hours in front of a computer screen.

BCI based on SSVEP or mVEP is often combined with MI-BCIs to increase the number of BCI commands in order to increase the transfer rate and flexibility of the game. Ultimately this makes the game more dynamic and engaging.

Motion-onset visual evoked potential (mVEP) is composed of three peaks, P1, N2, and P2, allowing visual stimuli to elicit different brain patterns depending on the motion and position of the stimuli [137,138]. The N2 peak, with a latency of 160–200ms, is predominantly motion specific. The P2 peak with a latency of 240ms increases with the increased complexity of visual moving stimuli.

The paradigm of existing mVEP-based BCI is designed as a set of moving bars in boxes as virtual targets on a screen. The brief motion of a visual bar inside the square elicits mVEP with a larger amplitude than that of peripheral screen boxes while users gaze at one of the square boxes. Thus, the EEG data segment aligned to the motion onset of the chosen target will contain time-locked mVEP, with a prominent motion-related VEP feature [139].

Among all visual motion-related VEPs tested to date [140–145], motion-onset VEPs display the largest amplitudes and the lowest inter- and intra-subject

variability, which is very important in order to achieve a high classification accuracy and reduce initial training time [146].

The amplitude and signal-to-noise ratio of mVEP depends on the neurophysiology principle of the mVEP features. In the majority of studies that used motion-onset VEPs, the stimulus consisted of a shape that started moving after being stationary for a necessary amount of time to allow a response evoked by preceding stimuli to finish [100]. This limits the stimulation frequency to around 200 ms, thus 5 or less mVEP stimulus can be delivered per second.

The motion-onset VEP is mainly characterized by occipital and occipitotemporal negativity that peaks between 150 and 200ms after motion onset. The first component of mVEP feature P1 is a pattern-off effect caused by the high contrast pattern disappearance due to its blurring at the beginning of the motion. In order to decrease a significant blur effect at the beginning of motion, a pattern of low spatial frequency or an irregular pattern with a lower number of superimposed pattern elements during motion is preferable. To minimize contamination of the motion-onset VEPs with the pattern-offset dependent positive peak, it is recommended to keep the temporal frequency below about 6 Hz while using a spatial frequency range of 0.2–1.0 c/deg and a velocity range of 5–25 deg/s to generate the desired temporal frequency of 5 Hz [142, 147]. The blur effect seems particularly effective in combination with high contrast patterns [142].

Motion-onset VEPs is also influenced by the timing of the motion stimuli. Long motion stimuli or short stationary phases between two motions (interstimulus intervals) cause adaptation of the motion-sensitive cortical areas and decrease the size of the N2 peak. The N2 peak, which probably represents motion processing system activity, seems to be generated from the extrastriate temporal-occipital and associated parietal cortical areas and is typically dominant in the right hemisphere [146]. It was reported that the N2 peak of motion-onset VEPs could be recorded at very low contrast levels of ca. 0.4%, dependent on the spatial frequency of the moving pattern [146].

The P2 peak with a latency of about 240 ms has larger inter-subject variability than the N2 peak and seems to depend on the type of motion [146]. This peak increases with more complex visual moving stimuli (expanding/contracting radial motion), perhaps because it represents a higher order of visual processing of biologically important stimuli. Its largest amplitude is usually recorded by the parietal (up to central) electrodes, and it is not sensitive to motion direction.

Rotating objects were used in some research studies instead of moving bars [147]. Some research indicates that the colors of the objects and background could affect the classification results in VEP-BCI systems because the color of

stimuli and also the contrast can influence the performance of a P300 speller [136]. Thus in this study, different combinations of colors and backgrounds layout were investigated in order to improve the performance of mVEP-BCIs [148].

3.2 Methods of Offline Study for six mVEP BCI layouts

3.2.1 Participants

Ten participants with ages ranging from 19 to 27 years old (mean 24 ± 5 years old. 3F, 7M) took part in this study. Half of them had previous experience with BCI. All of the participants had normal or corrected to normal vision. They were asked to sign an informed consent form, and the study was approved by the Ethical Committee of the College of Science and Engineering at the University of Glasgow.

3.2.2 Experimental Paradigm

Visual stimuli were presented on a 23-inch LCD computer monitor (DELL, USA) with a resolution of 1920*1080 pixels. Subjects were asked to sit in front of the computer's screen, to sit approximately 50 cm from the screen. Eight electrodes CP3, CPz, CP4, P3, Pz, P4, O1, and O2 (numbered as 1, 2, 3, 4, 5, 6, 7, 8 respectively) based on 10-20 international systems were used for data acquisition, with AFz serving as a ground and the A1(the left ear) as a reference (Fig. 3.1). The sampling frequency for data acquisition was 256 Hz. The impedance was kept under 5 KΩ.

The red moving bar and a box composed a virtual target. A virtual keyboard was formed by four numbered virtual targets, representing Up/Right/Down/Left commands, respectively. During the experiment, visual stimulation was presented as a red vertical bar moving from right to left through four boxes (the four targets) spaced 8cm apart, repeated for 160ms(Fig. 3.2). Subjects were asked to gaze at the moving red bar in the selected boxes corresponding to a number (1 to 4) provided in the center of the screen. The order of displayed numbers was semi-random, and over one run, a bar passed by each box (1,2,3, and 4) sequentially. It took 140 ms for a red bar to cross a single box, and there was an additional 60 ms interval until the bar started to move on to the next box. The red bar passed through all four boxes for 800 ms in one run with a 200 ms delay between consecutive runs. Each had 5 runs and there were 2s rest between

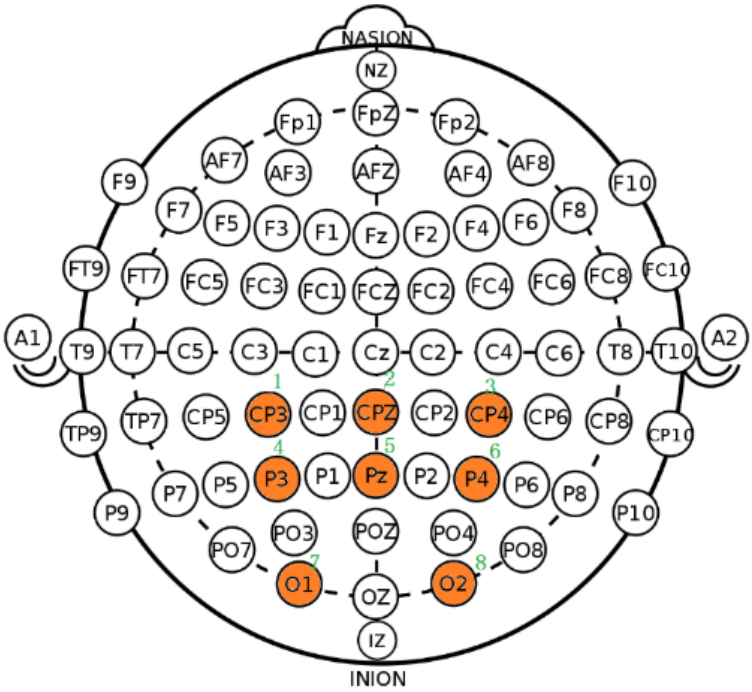


Figure 3.1: Electrodes montage according to 10/20 system. Electrodes marked in orange were used in this study

consecutive trials. 100 trials in total were presented to each subject for a 5-fold cross-validation classification. The experiment time of each layout lasted 5 minutes.

A starting point to detect mVEP was synchronized with the start of the bar moving in a box, the number of which was shown in the center of the screen. During the acquisition period of 1 trial, the participant was instructed to focus on the target boxes indicated by the number at the center of the screen. In one experimental session, for each subject, the EEG data of each target and the non-target box were recorded and epoched for offline data analysis.

There were six experimental sessions with different layouts of boxes to explore the effect of different visual presentations (Fig. 3.2). Experimental sessions were presented in semi-random order to different participants to minimize the effect of fatigue on any particular experimental layout. In order to integrate the mVEP into a game scene, it is necessary to investigate the influence of the different colors of the moving bar as well as the overall size of the boxes. The difference between Layout b and Layout c, when compared with a, is just the color of the moving bar. Layout d has a grey background to decrease the contrast of the GUI. Layout e has a smaller target box whose length and width are only 3/5 of that in Layout a. The entire size of Layout f is only 60% of the original one to provide a more central view.

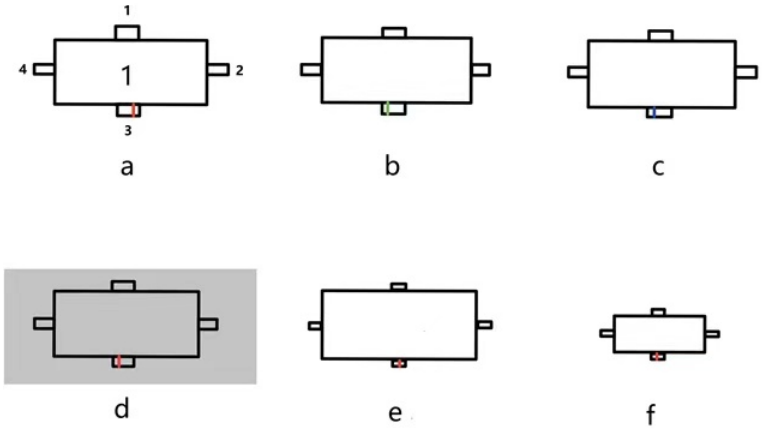


Figure 3.2: Six layouts: Layout a is the original Layout with red moving bars, layout b, and layout c have the same size as layout a, but they include green or blue moving bars. The Layout d(RGB color: R=196, G=196, B=196) has a different color of background compared with Layout a. Layout e has a smaller size of moving bars while the entire size layout f is only 60% of Layout a to give a more central view.

3.2.3 Feature Extraction and Classification

The amplitude of mVEP EEG signals is often comparable to the background EEG, which results in a low signal-to-noise ratio. The raw EEG was visually inspected in EEGLAB for removing bad epochs. Trials containing obvious artifacts (e.g., eye blinks) were removed from further analysis. Especially those EEG signals with amplitude above 100 μV were regarded as noise and removed based on visual inspection. After this step, the signal was filtered between 0.5Hz-10Hz by a 101th-order FIR filter [149]. Data epochs were extracted with respect to the corresponding motion-onset stimulus, beginning at t=0ms when the motion started and lasting until t=900ms.

According to the main component of the mVEP, the first step of feature extraction is to specify the time range of the main component of mVEP. To avoid overfitting, the data sample rate across all channels was reduced from 256 samples/s to 20 samples/s. According to the figure showing mVEP feature, three main components appear at around 200ms-500ms. This yielded 6 points per epoch after downsampling the data from 256 samples/s to 20 samples/2 [149]. The experiment consisted of one hundred trials, with each trial comprising five repeated runs, resulting in 500 single runs in total. Afterward, the data were processed by averaging every three runs among the 500, leading to a final tally of 166 (500/3) features.

It is important to use fewer electrodes to reduce the setup time without the loss of performance of BCIs.Different combinations of electrodes were tested to

investigate the influence of the number of electrodes. Initially, only the features derived from the electrode with the highest accuracy out of the 8 electrodes were considered. Subsequently, combinations of two electrodes that achieved the highest accuracy were calculated. The number of electrodes was then gradually increased by one in all combinations until all 8 electrodes were utilized. As mentioned, subjects were only asked to look at 1 out of 4 boxes at one time. Thus, the number of 'target' features was only 1/3 of the 'non-target' features. For that reason, a random 1/3 of non-target features was selected to build a classifier in order to achieve an equal number of target and non-target features. Each participant's data set was divided into a training set and a testing set to perform 5- fold cross-validation. It divided the whole 332 features (166 target features and 166 non-target features) into 5 parts and used any 4/5 for the training classifier while the rest 1/5 for testing accuracy. The final accuracy is the average of 5 tests.

Determining the presence or absence of mVEP in EEG features can be considered a binary classification problem with a linear discriminant function having a decision hyperplane defined by:

$$w \times x + b = 0, \tag{3.1}$$

where x is the m-point feature vector, w is a vector of classification weights, and b is the bias term. Fisher's linear discriminant analysis was implemented to calculate w and b from the training set. It was assumed that mVEP was elicited for one of the four boxes. Thus, the resultant target detection for the testing set was taken as the feature vector with the largest positive distance from the trained separating hyperplane.

Fig. 3.3-Fig. 3.6 show an example of features extraction steps of mVEP between target data (red line) and non-target data (blue line) in 8 channels. Fig. 3.3 is an original run of mVEP just after filtering, and it can be seen that there are only some peaks for VEPs but hard to differentiate the target data and the non-target data. After averaging across every three runs, the signal-to-noise ratio increases. Thus mVEP feature can be clearly seen in Fig. 3.4. Then Fig. 3.5 is the mVEP feature after downsampling from 256 samples to 20 samples and only 6 interested samples in Fig. 3.6 for each mVEP feature which correspond to the 200ms-500ms when the main peak of mVEP appears.

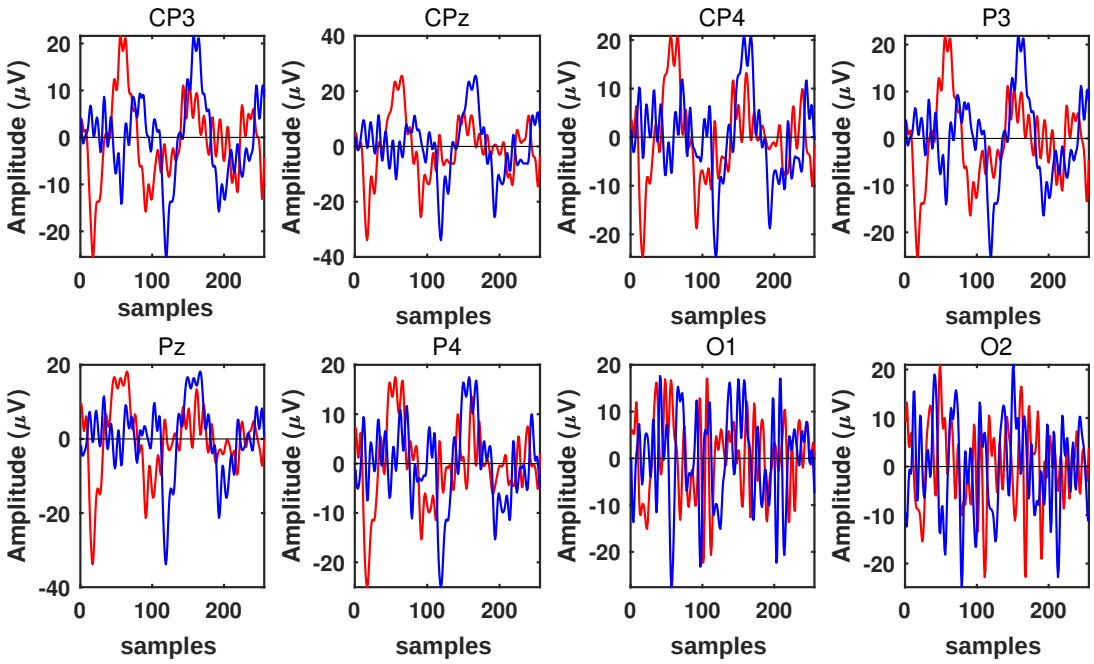


Figure 3.3: A filtered original mVEP run of subject 1 in Layout a. Red line represents a target, and the blue line represents a non-target mVEP.

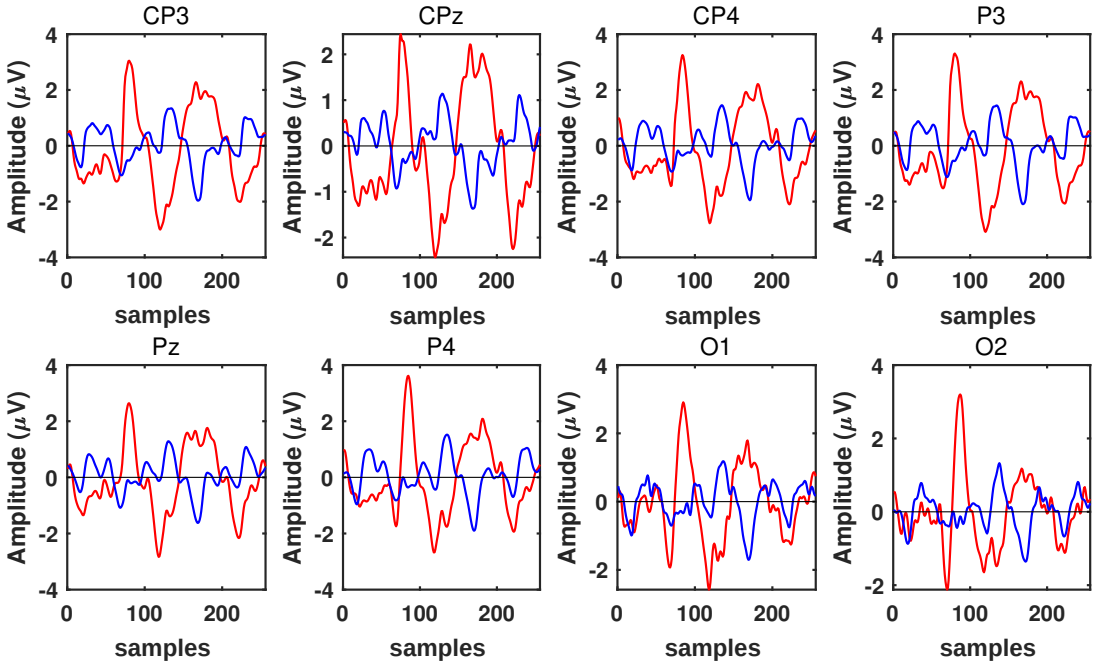


Figure 3.4: Averaged mVEP feature of subject 1 in Layout a. Red line represents a target, and the blue line represents a non-target mVEP.

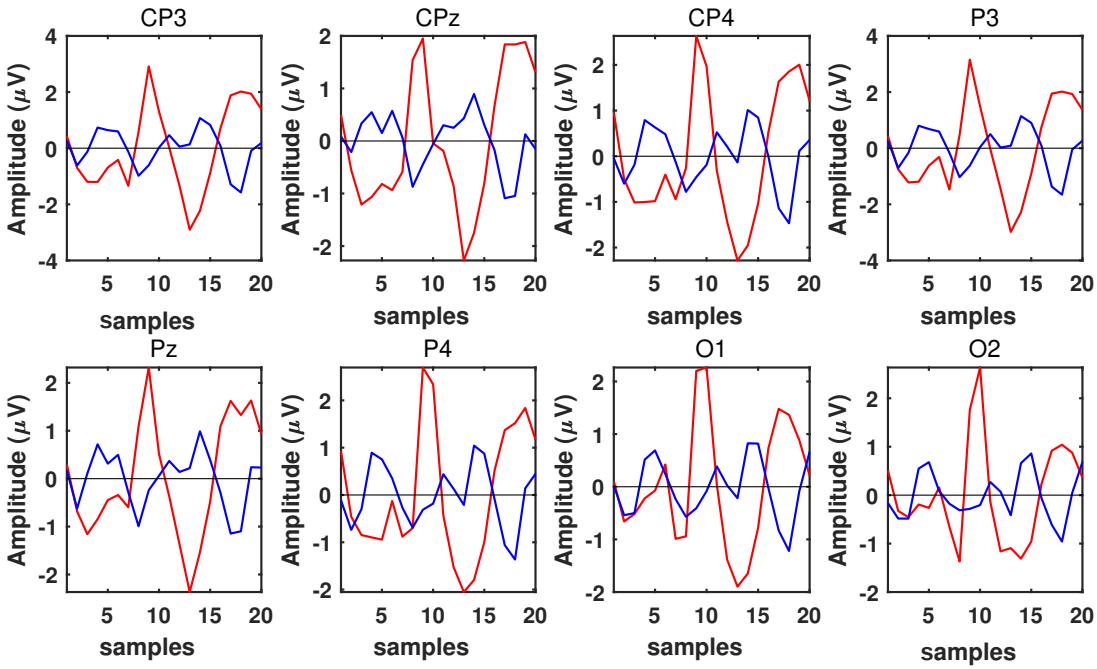


Figure 3.5: Downsampled single trial mVEP feature of subject 1 in Layout a. Red line represents a target, and the blue line represents a non-target mVEP.

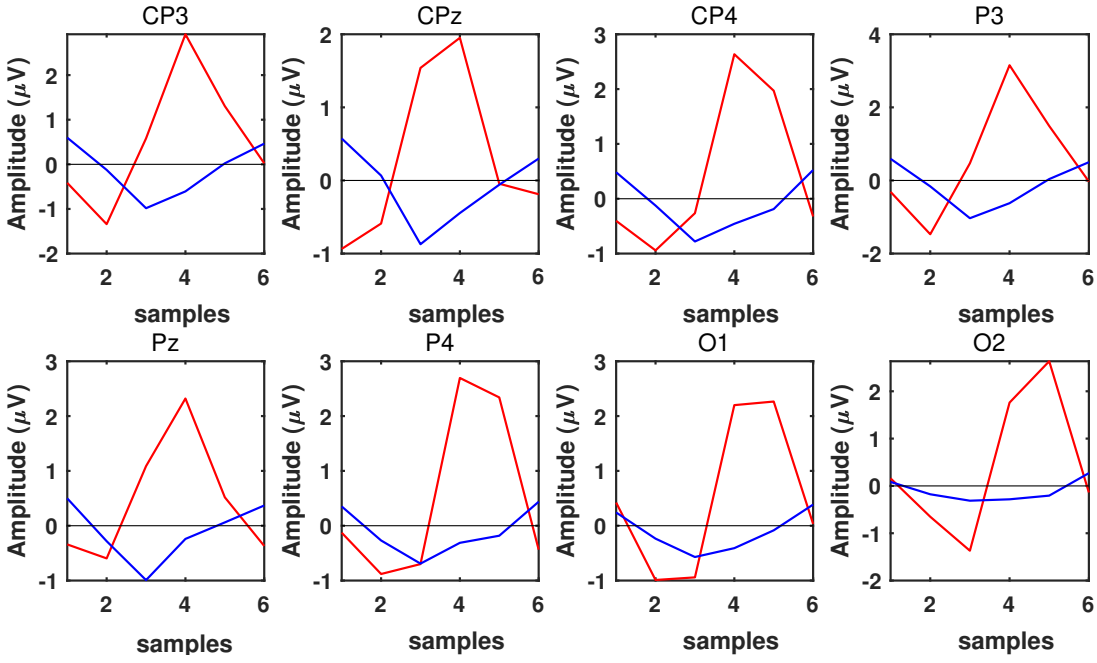


Figure 3.6: Windowed mVEP feature of subject 1 in Layout a. Red line represents a target, and the blue line represents a non-target mVEP.

3.3 Results of Offline Study

3.3.1 Morphological Analysis of Target and Non-target motion-onset Visual Evoked Potential

Fig. 3.7 shows the mVEP feature presented with 256 samples/s over a period of 1s for subject 1 (S1) on the CP3 electrode when box 4 was the target box. Each number represents a corresponding box on the GUI. It can be seen that between the window length of the 50th-128th of 256 samples, after the cue appears (200-500ms), box 4 shows an mVEP feature in red. When observing non-target mVEP in blue in the other three boxes (box 1, box 2, box 3), it is evident that the peak occurs with a delay after the 128th sample and it can be observed that there is no mVEP waveform during the corresponding time window(50th-128th samples) of these three non-targets boxes, indicating information about the leading or lagging of mVEP feature. This means that these three boxes are the targets that the subjects are not paying attention to because they are focusing on box 4 at this time.

Table 3.1: channel selection that achieved the highest classification accuracy with the varying number of channels for S1 in Layout a.

Electrode locations	O1	O1 O2	Pz O1 O2	Pz P4 O1 O2	CPz Pz P4 O1 O2	CPz P3 Pz P4 O1 O2	CP3 CPz P3 Pz P4 O1 O2	CP3 CPz CP4 P3 Pz P4 O1 O2
Accuracy (%)	64.72	72.37	75.23	80.56	81.05	81.24	79.25	78.34

3.3.2 Influence of Graphical User Interface Design on Classification Accuracy

Even though 8 electrodes were used to collect EEG signals, it is important to use fewer electrodes to reduce the step time. In order to define the suitable number of electrodes, the number of electrodes varied from 1 to 8 to find the maximum accuracy for each subject for each number of channels in layout a. Table 3.1 gives an example of S1's best classification performance achieved when the number of channels was increased from 1 to 8. The average classification accuracy of 10 subjects across 8 circumstances is shown in Fig.3.8.

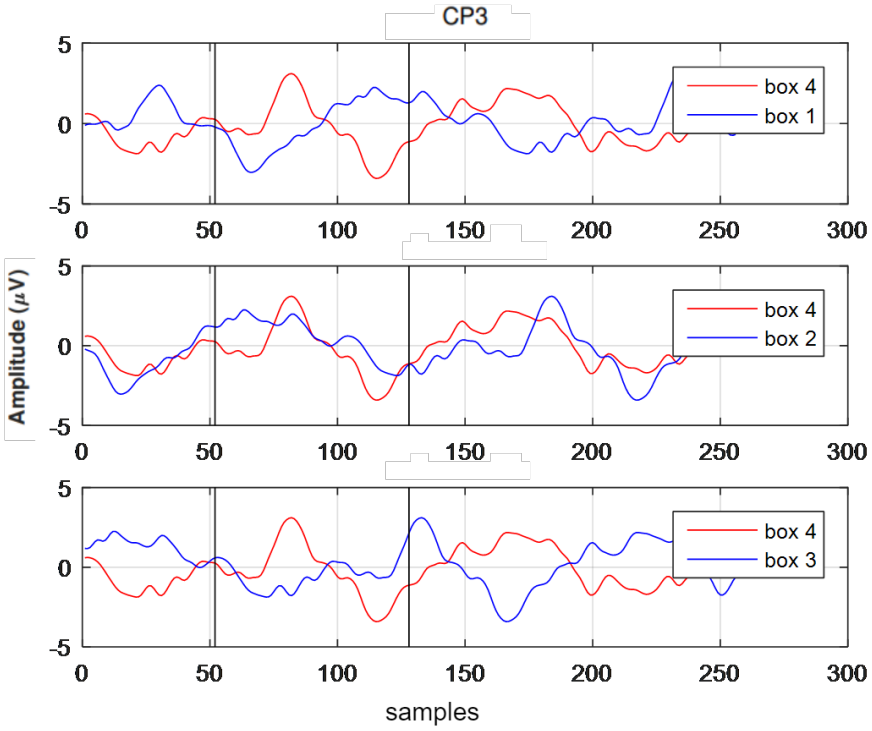


Figure 3.7: mVEP feature of box4 in 1s (256 samples). When box 4 is the target box which is the same in the three sub-figures, the red line represents that the mVEP feature appears at a time period of 200-500ms (i.e., from the 51st to 128th sample). The blue lines in the three small graphs represent non-target EEG signals from boxes 1-3. If the mVEP feature is extracted based on the time period (200-500ms), It can be observed that there is no mVEP waveform during the corresponding time window (200-500ms) of these three non-targets boxes, indicating information about the leading or lagging of mVEP feature. This means that these three boxes are the targets that the subjects are not paying attention to because they are focusing on box 4 at this time.

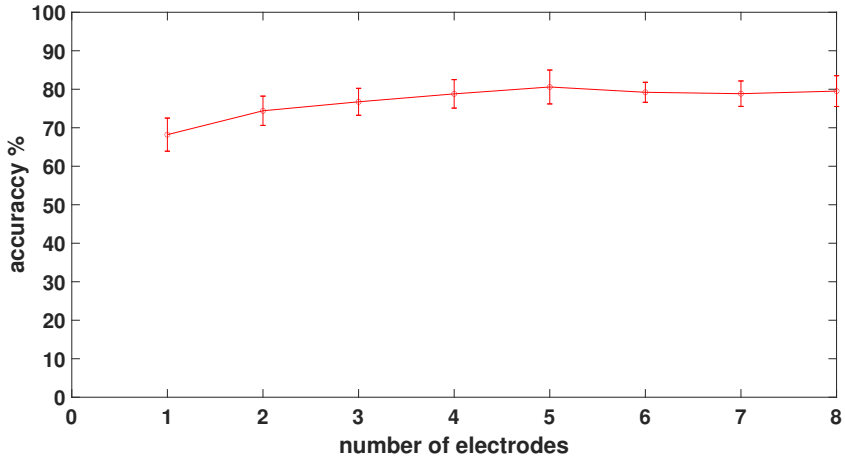


Figure 3.8: The average classification accuracy across 10 subjects for different numbers of EEG channels selected. Results are presented as mean ± std.

The average accuracy for 1-8 electrodes respectively is 68.21%±4.27%, 74.41%±3.84%, 76.74%±3.48%, 78.80%±3.73%, 80.59%±4.36%, 79.21%±2.58%, 78.85%±3.26% and 79.52%±4.02%(Fig. 3.8). Based on the results, it can be concluded that the average accuracy does not increase when more than five channels are used, and using five electrodes results in the highest classification accuracy. However, the improvement in accuracy for using five electrodes over four electrodes is only 1.79%. The average accuracy is very close for using 4 and 5 electrodes, using fewer electrodes is more user-friendly. This is why the results for different layouts are presented using four electrodes. Kolmogorov–Smirnov test suggested that classification accuracies for the different numbers of electrodes were not normally distributed. Moreover, the non-parametric Kruskal Wallis test showed no significant difference ($p > 0.05$) among eight groups of data (1-8 electrodes used). Fewer channels are more user-friendly; thus, for different GUI layouts in the second experiment, only up to 4 channels were used. Table 3.2 shows the classification results of 10 subjects under 6 different layouts and the corresponding channel selection. Figure 3.10 presents the total number of times that each channel was selected by 10 subjects from table 3.2. It can be seen that O1 and O2 are the frequently used channels.

Table 3.2: 10 subjects' offline classification results (%)using 4 channels.

Subjects	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Mean±std
Electrodes	Pz P4 O1 O2	CP3 P3 O1 O2	CPZ CP4 P4 O2	CP3 PZ P4 O1	CPZ CP4 PZ P4	CP3 CPZ CP4 O1	CPZ P3 PZ O2	CP4 P3 P4 O1	P3 PZ O1 O2	P3 P4 O1 O2	
Layout a	86.88	74.41	82.02	83.16	72.30	79.21	78.06	70.15	75.03	74.09	77.53±5.29
Layout b	84.09	77.15	88.90	82.63	73.87	67.52	80.76	81.39	81.54	80.60	79.85±5.86
Layout c	88.02	76.51	79.24	81.65	82.53	64.16	77.48	80.23	81.70	80.30	79.18±6.14
Layout d	84.31	65.38	74.18	81.01	85.96	72.94	82.36	84.45	75.76	77.63	78.40±6.46
Layout e	74.73	82.03	74.19	86.33	78.65	70.88	80.13	77.82	70.88	74.15	76.98±4.96
Layout f	81.03	79.76	76.54	93.16	84.96	74.75	78.38	82.50	85.61	82.28	81.90±5.25

In Table. 3.3-Table. 3.8, we presented the average Confusion matrix over 10 subjects for 6 layouts, it should be noted that each subject had its own classifier trained thus, the confusion matrix actually didn't indicate the performance for a specific classifier but only the overall performance of classifiers over 10 subjects.

Table 3.3: Average confusion matrix of layout a

Layout(a)	Predict positive	Predict negative
Actual positive	80.13%	19.87%
Actual negative	25.06%	74.94%

Ten subjects' performance was analyzed separately for six different GUI layouts. Kruskal Wallis test was applied to the results in table3.2 to test

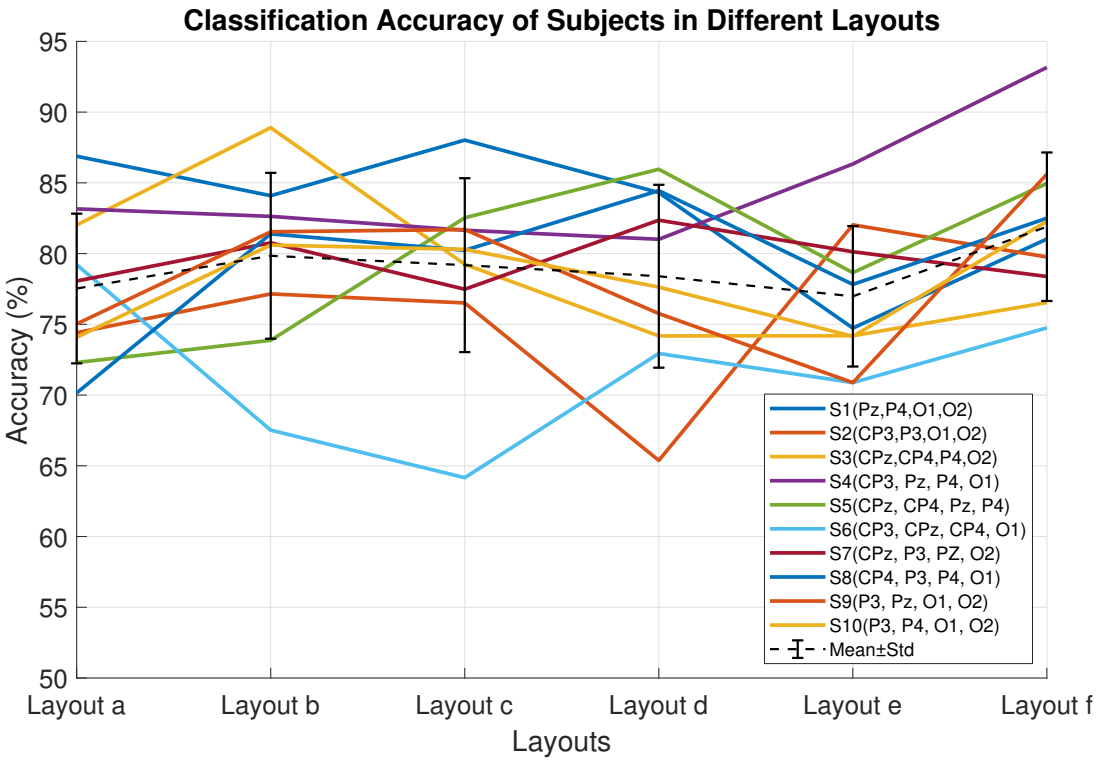


Figure 3.9: mVEP feature of subject 1 in Layout a. Red line represents a target, and the blue line represents a non-target mVEP.

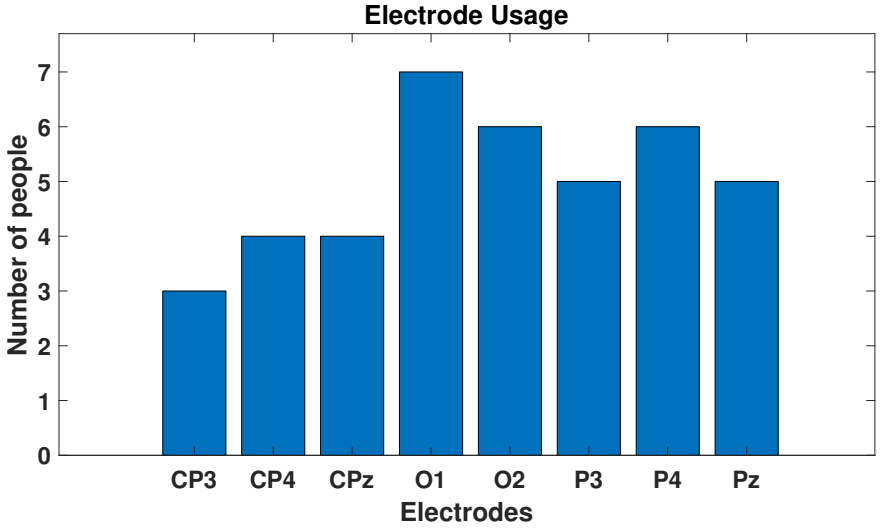


Figure 3.10: The number of times that each channel was selected for all 10 subjects.

Table 3.4: Average confusion matrix of layout b

Layout(b)	Predict positive	Predict negative
Actual positive	77.16%	22.84%
Actual negative	17.46%	82.54%

Table 3.5: Average confusion matrix of layout c

Layout(c)	Predict positive	Predict negative
Actual positive	83.66%	16.34%
Actual negative	25.29%	74.71%

Table 3.6: Average confusion matrix of layout d

Layout(d)	Predict positive	Predict negative
Actual positive	84.76%	15.24%
Actual negative	28.16%	71.84%

Table 3.7: Average confusion matrix of layout e

Layout(e)	Predict positive	Predict negative
Actual positive	75.92%	24.08%
Actual negative	21.96%	78.04%

Table 3.8: Average confusion matrix of layout f

Layout(f)	Predict positive	Predict negative
Actual positive	86.58%	13.42%
Actual negative	22.74%	77.24%

whether there was a significant difference in classification accuracy between 6 groups, showing no statistically significant difference (p=0.408). The effect size ($\eta^2 = 0.085$) suggested a moderate effect. Notably, layout f achieved the highest average accuracy rate, possibly due to its more central view angle. As a result, subjects required less eye movement when observing moving targets, thus reducing the interference of the electrooculogram (EOG).

3.3.3 Influence of Spatial Distribution of Boxes on Classification Accuracy

In the first experiment, different boxes were positioned on four sides of the screen (left, right, up, and down).To determine whether there is a specific location where mVEP can be detected exceptionally well or poorly, we selected classification results based on Layout A and 4 EEG electrodes for further analysis. This is because the placement of boxes on Layout A (on all four sides), the use of moving bars in red color, and the background color are commonly used in other mVEP

studies. By utilizing this classic layout, we can provide a valuable reference point for future mVEP research.

Features were produced when subjects are looking at box 1. Thirty-nine randomly selected target features (from box 1), and non-target features from box 2, box 3, and box 4 (39 respectively) were sent through the already trained classifier to see if the position of the box had significance. It should be noted that, since the classifier trained was a two-class classifier, the features from the target box1 and only one out of three boxes' non-target features were sent together each time, i.e., features of box1 and box2 were tested, then features of box1 and box3 are tested, finally, box1 and the box4 are tested. The number of instances when each time non-target box was correctly classified when sent together with the target box1 is shown in the Table. 3.9:

Table 3.9: The number of times non-target boxes were right classified when box 1 was a target box. (39 features for each box were tested)

	Box2	Box3	Box4
S1	35	32	30
S2	29	28	28
S3	34	30	32
S4	30	32	34
S5	30	29	27
S6	30	28	32
S7	31	31	27
S8	26	24	26
S9	28	29	30
S10	28	30	28

Table 3.10: The number of times non-target boxes were right classified when box 2 was a target box. (39 features for each box were tested)

	Box1	Box3	Box4
S1	33	33	31
S2	38	29	29
S3	35	28	37
S4	31	38	35
S5	26	39	34
S6	34	36	37
S7	35	32	28
S8	30	32	34
S9	29	26	32
S10	29	36	31

The data from the three groups (boxes 2-4) was not normally distributed, and a non-parametric Kruskall-Wallis test was applied. The results show no

Table 3.11: The number of times non-target boxes were right classified when box 3 was a target box. (39 features for each box were tested)

	Box1	Box2	Box4
S1	36	34	32
S2	31	29	37
S3	35	26	30
S4	29	37	29
S5	34	33	31
S6	37	28	38
S7	30	36	30
S8	39	30	28
S9	33	36	26
S10	28	27	32

Table 3.12: The number of times non-target boxes were right classified when box 4 was a target box. (39 features for each box were tested)

	Box1	Box2	Box3
S1	33	28	36
S2	35	27	31
S3	30	26	28
S4	36	34	26
S5	28	31	30
S6	33	29	38
S7	31	28	36
S8	29	26	34
S9	38	33	32
S10	27	34	32

difference in classification results for different non-target box positions when subjects were looking at Box1 (p=0.831). Similarly, other combinations were also investigated, with Box2(p=0.669) in Table. 3.10, Box3(p=0.516) in Table. 3.11 and Box4(p=0.206) in Table. 3.12 as the targets. No statistically significant difference was found between the different locations of the boxes. Thus it can be concluded that the location of the target box doesn't influence the performance of this offline BCI test.

In conclusion, our results show no clear relation between elements of graphical design, which suggests that the colors and background may not have an influence on mVEP detection. There has been no comprehensive discussion on the impact of graphic elements on BCI performance in the field of mVEP, as presented in this chapter. Our study is novel, and our findings provide valuable references for integrating mVEP BCI into gaming. The discussed factors are all modifiable and do not significantly affect performance, which allows for UI changes to make the experience more vivid and engaging.

3.4 Methods of Online Study for two different mVEP layouts

3.4.1 Participants

Ten participants aged from 20 to 29 years old (mean 24±5 years old, 5 female, 5 male) participated in this online study. Seven of them had experience with BCI. All the people had normal or corrected to the normal vision when they took part in this experiment. They were asked to sign an informed consent form, and the study was approved by the Ethical Committee of the College of Science and Engineering at the University of Glasgow

3.4.2 Experimental Paradigm

Based on results from study 1, the number of electrodes was reduced to 2 in this study, as shown in Fig. 3.11. This further reduces the setup time. Results from the offline study show that the best combination of two electrodes to achieve the highest accuracy is O1 and O2. In this study, we compared Layout A from study 1 with boxes in left, right, top, and bottom locations with layout B where all four boxes are lined up horizontally, as shown in Fig. 3.12. Layout B is more convenient for the design of a graphical user interface to avoid overlapping the boxes with the game in the center of the screen. The order of displayed numbers

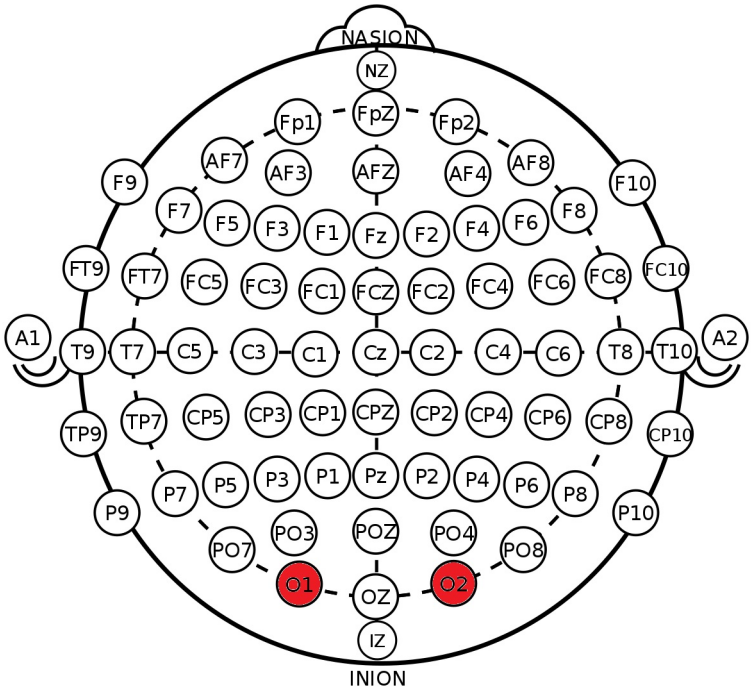
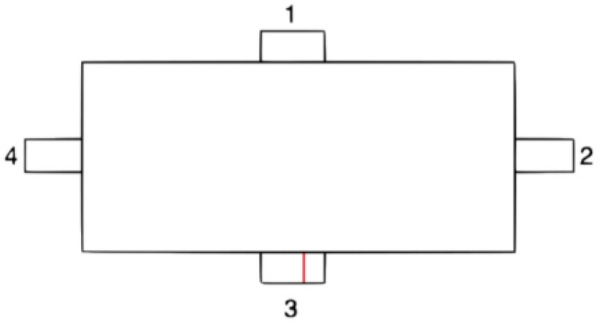


Figure 3.11: O1 and O2 are used as the electrodes for online mVEP study.

was predefined(semi-randomly, randomly). In one run, the red bar was moving from box 1 to box 4 sequentially. It took 140 ms for a red bar to cross a single box, and there was an additional 60 ms interval until the bar started to move on to the next box. The red bar passed through all four boxes for 800 ms in one run with a 200 ms delay between consecutive runs. For the offline experiment, each trial had 5 runs and there were 2s rest between consecutive trials. 100 trials in total were obtained for each subject. In the online experiment, each trial only contained 3 runs. There was also 2s' rest between trials and the target number of boxes appeared 2s in advance. 140 trials were tested(each number of box appeared 35 times).It should be noted that the predicted number of the box was calculated in real-time, namely, the code for the upcoming box would appear on the screen immediately at the end of each trial in a hidden window on the monitor. Since the participants were asked to avoid eye movement, they didn't see the predicted results until they finished all 140 online trials. On the other hand, the hidden window can reduce the distraction since the participants were also asked to be highly focused.

3.4.3 Feature extraction and classification

For the training stage of the online experiment, in each trial(contains 3 runs), the number of boxes that the subjects should focus on would appear 2s in advance, which also corresponds to the rest time between every two trials. One hundred



(a) LayoutA of online study

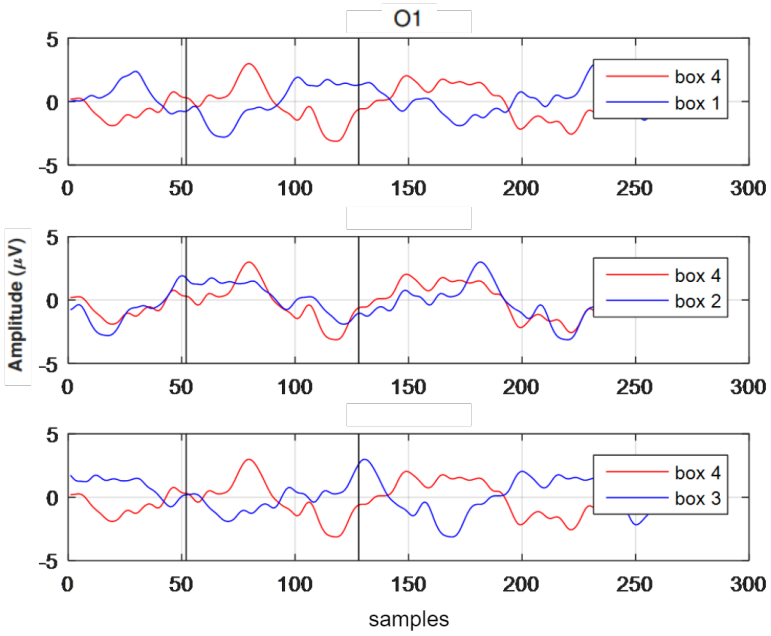
(b) LayoutB of online study

Figure 3.12: Online experiment layouts

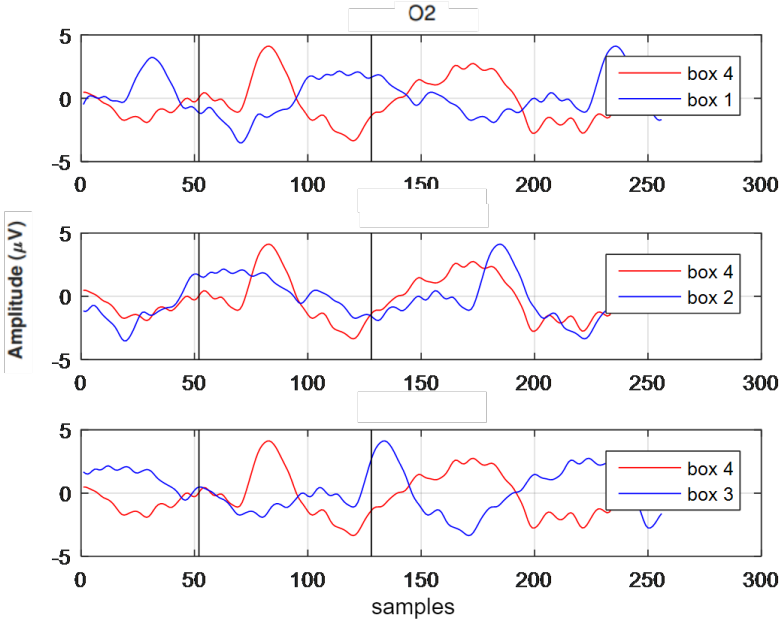
trials in total for each layout were obtained for feature extraction and training the classifier. Each trial has 3 repetitive runs which resulted in 300 runs. Every 3 runs were averaged, and 100 (300/3) features were finally obtained to train the classifier. Then the same amount of the non-target trials was added with the 100 target features to train the classifier. Finally, the pre-trained classifier in the training stage was applied in the final online experiment, where every 3 runs in each trial from 140 trials were averaged to export the predicted number of the box.

3.5 Results of Online Study

Data processing and classifications were similar to the previous study. Fig. 3.13 shows the mVEP for 256 sample points (corresponding to 1000ms) for S1's O1 and O2 electrodes (channel 1 to the left, channel 2 to the right) when box 4 is the target using Layout B. It can be seen that between the window length of 50th-128th samples (200-500ms), box 4 shows an mVEP feature. When observing mVEP from the other three boxes (box 1, box 2, box 3), the latency of the positive peak is out with the 200-500ms interval, which indicates that mVEP was evoked in the other box (i.e. box 2 in this case).Table 3.13 shows the test accuracy of 140 online trials. The two groups of data were not normally distributed and the non-parametric test results suggested no statistically significant difference between the two layouts.



(a) channel O1: mVEP features of box4 in 1s(256samples)



(b) channel O2: mVEP features of box4 in 1s(256samples)

Figure 3.13: mVEP feature for box 4 of S1

Table 3.13: Test accuracy of 140 trials for two different layouts.

Subjects	LayoutA	LayoutB
S1	67.85	74.29
S2	70.00	74.29
S3	82.14	82.86
S4	68.57	70.71
S5	71.43	67.85
S6	70.71	75.00
S7	77.8	78.57
S8	65.0	62.85
S9	73.57	78.13
S10	87.86	81.43
Average	73.49±7.09	74.60±6.17

3.6 Discussion and Conclusions

This study investigates the influence of the location and the number of EEG electrodes, and the GUI design on mVEP performance. This is an important issue for a game design, where the central content of the game may be affected by mVEP boxes.

Color may affect the performance of P300 BCI, and its performance [136] can be further improved by overlaying letters with the faces of famous people. P300 presents a higher level of cognitive processing than mVEP. Thus, improvement in mVEP with adding faces could clearly not be expected. It is surprising that the color did not make much difference. One possible explanation is that the bars had too small in size and thus, changes in color were not sufficient to affect the visual processing.

The most relevant result was from the offline research of this chapter, showing that the Layout of boxes on the screen is not relevant, which allows larger flexibility in GUI design and placement of boxes in the areas which are convenient for a particular layout of the game. Remarkably, in the offline research findings, the smaller boxes present in Layout E exhibit a diminished classification accuracy in contrast to Layout A, devoid of statistical significance. In contrast, Layout F exhibits a superior mean classification accuracy when compared to the other five layouts, albeit lacking statistical significance. The effect size ($\eta^2 = 0.085$) suggested a moderate effect of graphical factors on six different layouts. Both Layouts E and F share the attribute of smaller boxes, however, the distinguishing factor is that Layout F confers a more concentrated visual angle. As previously alluded to, this could potentially be ascribed to the fact that Layout F instigates fewer ocular movements and electrooculography signals. However,

this contradicts the general research on mVEP, which recommends the use of larger visual files [146]. It is possible that the small sample size of 10 participants led to this conflicting conclusion. Therefore, further research is needed to explore this topic in more depth. Beverage et al. [150] investigated the location of boxes on the screen (on the top or more centrally), but they were always horizontally aligned.

The systematic analysis of electrodes contributing most to classification accuracy showed a consistent selection of O1 and O2 alongside P4. The results of this study are in accordance with previously published literature showing that good classification results can be achieved with a few occipital electrodes only. In [150], 12 channels were ranked according to classification accuracy, and the top three were O1, P7, and TP7. In contrast, in our study, two out of three channels were in the right hemisphere, in which N2 dominates [146, 150]. We found the strongest responses on Oz and electrodes up to 5 cm to the left and right e.g. O1, O2, Oz, O3, O4.

In [150], mVEP performance was analyzed in the presence of various commercially available video games to assess the influence of other moving objects on the screen. They used a standard red bar with five boxes aligned at the top of the screen. Interestingly, they found a better performance when no white background was implemented when mVEP boxes overplayed the game. Classification results, however, did depend on the game.

In an analysis of mVEP outwit BCI, [146] provided recommendations for optimal screen parameters, including the type of motion, the time course of motion, duration of motion vs the interstimulus interval, monitor frame frequency, stimulus field, luminescence, and contrast. Other aspects that affect mVEP should be analyzed in further studies, such as luminescence, contrast speed of motion, or moving patterns (linear vs radial). Finally, given that some conditions that affect the visual system such as glaucoma, neuroborreliosis, amblyopia but also dyslexia, and age might affect mVEP [150] it would be of interest to check mVEP on some of these conditions, at least on some that affect a large population such as people of different age or people with dyslexia.

Chapter 4

Three Data Augmentation Methods based on Common Spatial Pattern, Band Power and Time Domain Parameter Feature Extraction Methods on Motor Imagery Data

4.1 Introduction

In the previous chapter, the impact of graphical factors on the performance of MVEP BCI was investigated. The results showed that certain graphical elements had no significant effect on the accuracy of the BCI system. However, due to the limited amount of available training data, the performance of the BCI system was still suboptimal. To address this issue, in this chapter, the potential of data augmentation techniques to improve the performance of Motor Imagery BCI was explored. By generating additional artificial training data, the training time and the classification accuracy of the BCI system can be improved. Motor imagery Brain-Computer Interface is a widely used type of BCI with the main application for rehabilitation of movement in the patient population. A typical MI BCI is based on a cue-based paradigm where participants are asked to repeat the imagined movements with different limbs several times to create a reliable classifier [75]. MI-BCI is typically based on EEG. EEG possesses excellent time resolution of tens of hundreds of milliseconds, capable of capturing cognitive

dynamics [151]. Because EEG is a multidimensional signal, represented in time, frequency, and spatial domain, it can measure complex patterns of neuronal activity. However, EEG has an inherently low signal-to-noise ratio because it is an aggregate signal, representing the activity of hundreds of thousands of neurons being dampened by several layers of tissue [152]. Another inherent problem for EEG online analysis, where models are trained on past EEG data to decode present neuronal activity, is its non-stationarity [153].

For that reason, to achieve high accuracy, BCI needs to be trained and re-trained on a large amount of data to avoid overfitting [154]. Unfortunately, collecting these data is time-consuming and often tiring for participants, in particular when they are patients [155]. It was estimated that at least 40 trials per class are necessary to obtain reasonable online BCI performance [156], and it is often necessary to recalibrate the BCI classifier for each use. There are two most widely accepted methods to improve classification accuracy without acquiring additional data from a participant. The first one is transfer learning [157, 158], where a classifier is pre-trained on data of similar participants doing the same task, and the other is DA which is a focus of this study. DA is a general approach that can be applied to EEG signals beyond BCI, such as analysis of sleep or seizure detection [159], but in the area of BCI, it has been implemented in motor imagery [160] and motor execution [161] tasks, estimation of mental workload [162], emotion recognition [163] and visual tasks [164]. A number of methods have been proposed in the literature to create DA, such as noise addition, sliding windows, sampling, Fourier Transform, recombination of segments, or General adversarial networks (GAN) [159]. While some DA methods take raw EEG directly (typically implementing deep learning classifiers on a small number of EEG electrodes), multichannel DA is typically based on EEG-extracted features such as wavelets, entropy, spatial filters, short-time Fourier transform, spatial-temporal filters (Independent Components (IC), Principal Components (PC), CSP), power spectral density [159]. Alternatively, the time-frequency decomposition of EEG signal (spectrogram) can be treated as an image to implement DA methods originally developed for DA of a static 2D image [165].

Deep learning methods are popular classifiers for data-augmented BCI, previously showing excellent results with DA in image analysis [166]. DL, however, typically requires large original data training sets, while BCI datasets are much smaller and non-stationary. For that reason, various methods such as dropout regularization, batch normalization, and transfer learning have been implemented to reduce the need for a large amount of real data [167]. Still, DL methods from the literature are based on a large amount of original data. For

instance, DA with DL for motor imagery tasks in [168] was based on 280 and 400 original trials, and in [169] on 200 trials to achieve an average improvement in classification accuracy up to 8% from 74.5%. More recently, [170] implemented a novel DA method based on brain area recombination on BCI competition IV dataset 2a and significantly improved classification accuracy with only 10 to 20 original datasets. A classifier based on EEGnet DL improved accuracy by approximately 16%. Interestingly a classifier based on the SVM also improved accuracy by approximately 7% (values read from a graph), which is comparable with improvement based on DL classifiers from other studies [170]. This is a significant improvement from previous studies. For example, [159] used the same dataset and, although they achieved excellent initial classification accuracy, improved accuracy by only 4% using DA. CSP features were used in all cases, utilizing 20 or more EEG channels.

Apart from study [170], only a few published studies on DA did not utilize DL classifiers. As already mentioned, while DL outperforms other classifiers, they typically require a large amount of data that is not achievable in most MI-based BCI applied to the patient population.

Xiong and Wei [171] implemented three different DA methods, time-shifting, time-frequency shifting, and analogy (based on independent components), that were tested on three different datasets, including MI, workload, and mental imagery, and LDA was used for classification. All three methods preserve the ERD/ERS feature of EEG signals in the frequency domain, typically within the 6-12 Hz and 18-24 Hz ranges, which is relevant for selected features i.e., CSP, and classifiers, i.e., LDA, both of which rely on calculating the covariance matrix of signals belonging to two classes. In the MI task, they used BCI competition IV dataset 2a. The classification accuracy improved for small numbers of original trials, up around 4% for ten original trials and around 2% for 20 original trials (reading from a graph) with 500 artificial trials. These three different DA methods are used in this chapter for the exploration of the influence of this DA method on different feature extraction methods of BCI classification accuracy.

Clemens et al. [172] used the Empirical Mode Decomposition method of temporal EEG frames in a task that consisted of left and right-hand motor imagery accompanied by functional electrical stimulation. They extracted CSP features and used LDA classifiers. There were only 40 trials per hand for implementing DA. They aimed to keep classification error under 33%, which they achieved by replacing 87.5% of real data with artificial trials.

Blankerz [173] implemented Empirical Mode Decomposition followed by mixing the resulting intrinsic mode functions. They used CSP features, LDA,

and a Linear Logistic Regression (LR) classifier. They tested their approach in BCI competition IV dataset 2a [174] and BCI Competition III dataset 4a [175], using 20 original EEG trials to create DA trials. They achieved up to 6.9% improvement in classification accuracy with the left-right hand classifier for the first dataset and 7.9% improvement for the second dataset.

While many different DA methods and classifiers were applied for MI tasks, there was no such variety of EEG-derived features. While CSP has shown better classification results than most other features, it requires a considerable number of channels [156]. Other methods such as Band Power or TDP [176] require a smaller number of channels, which reduces the setup time and provides an opportunity to use cheaper consumer-based BCI systems.

In this study, DA methods proposed by Lotte et al. [14] were implemented on three different features: CSP, bipolar BP, and TDP. All of these methods are related to band power: CSP optimizes spatial filters that optimally discriminate band-power features; TDP for the 0th derivate presents band power, and BP is self-explanatory. A BCI competition IV 2a dataset and a dataset collected in our laboratory [177] were used to test the robustness of the method. We asked the following research question: (i) which feature shows the best performance for a small size of the original dataset, (ii) is there an interaction between the DA method and feature selection (iii) to what extent increasing the DA dataset above the real EEG helps to achieve improvement in classification accuracy (iv) repeatability of results when tested on different datasets.

4.2 Method

4.2.1 Datasets

We used two datasets, one validated dataset from the BCI competition to compare our results with results from the literature and a dataset collected in our laboratory to test the generalizability of results.

Dataset 1 was a BCI competition IV dataset 2a [174] collected on nine healthy participants, recorded with 22 EEG channels, with a sampling frequency of 250 samples/s. The cue-based BCI paradigm consisted of four different mental tasks, namely MI of the left hand, right hand, feet, and tongue, which are indicated by arrows in the left, right, down, or up directions, respectively. Each class has 72 trials in the training set and another 72 trials in the test set (144 datasets in total). In the study, only two classes left and right-hand MI were used as the most frequently used motor rehabilitation tasks. This dataset was filtered in the 4-40

Hz range and then epoched in 4s-long trials. At the beginning of a trial at t=0s, a fixation cross and a short acoustic warning tone prompted the participant to be prepared. After two seconds at t=2s, an execution cue in the form of an arrow pointing in one of four directions (up, down, left, right) appeared and stayed on the screen for 1.25 s, prompting the participants to perform the desired MI task. Participants were instructed to sustain motor imagery for as long as the cross remained on screen, till t= 6s. A short break followed that lasted 1.5 s - 2.25 s. An epoch was extracted from t = 2s till t = 6s.

Dataset 2 also contained the MI task but was collected by our group [177]. This dataset is collected from 12 people, seven females and five males (age 25±4). Participants sit opposite the computer screen at approximately 1.5m, distance, with forearms resting on a desk. A 64-channel modular EEG amplifier (g.tec, Graz, Austria) recorded brain activity with a sampling frequency of 256 samples/s. Ground and reference electrodes were placed on the left and right earlobe. EEG impedance was kept under 5kΩ. There were three tasks: MI of the left hand, right hand, and both hands. Only MI of the left and right hands was used in this study. The experiment was split into 7 runs of 45 trials, proving 105 trials of one MI type. Each run provided an equal number of three MI types in a semi-random order.

The experimental paradigm consisted of the following: A readiness cue (a fixation cross) appeared on the screen at 0s, At t=1s an execution cue in the form of an arrow pointing left, right, or upwards (for left, right, and bimanual MI, respectively) appears and stayed on a screen and stayed for 4s, till t=5s. Participants were instructed to sustain motor imagery for as long as the arrow remained on screen. There was a semi-random period between 1.5 and 2.5s between trials.

Data were preprocessed in the following way: using the Matlab-based application Brainstorm, raw EEG data were segmented into epochs from -2 to 5 seconds relative to motor imagery onset. To reduce the effects of muscle artifacts, EEG was first bandpass filtered between 1 and 40 Hz using a 5th-order Butterworth filter, and power line noise was removed with a 50 Hz notch filter. The common average reference was applied, and noisy epochs were rejected through a process of visual inspection. Around 10% of epochs were removed per participant.

Notably, only 2s raw EEG for both datasets was finally used for the next study starting from 0.5s after the direction cue appeared. Fig. 4.1 shows the experimental paradigm of dataset 1 and dataset2, as well as the 2s long EEG signals used in this study.

For dataset 1, classifiers were built with 10, 20, 40, and 60 original trials (OT) combined with 0, 10, 20, 40, and 60 artificial trials (AT). All 10-60 AT datasets were built from the corresponding 10-60 OT training sets. For dataset 2, since it had only 105 trials, classifiers were implemented on 10, 20, 40, and 50 original trials, which were combined with 0, 10, 20, 40, and 50 artificial trials, originating from the corresponding OT trials. The testing set for dataset 1 consisted of the remaining 72 trials and for dataset 2 of the remaining 50 trials.

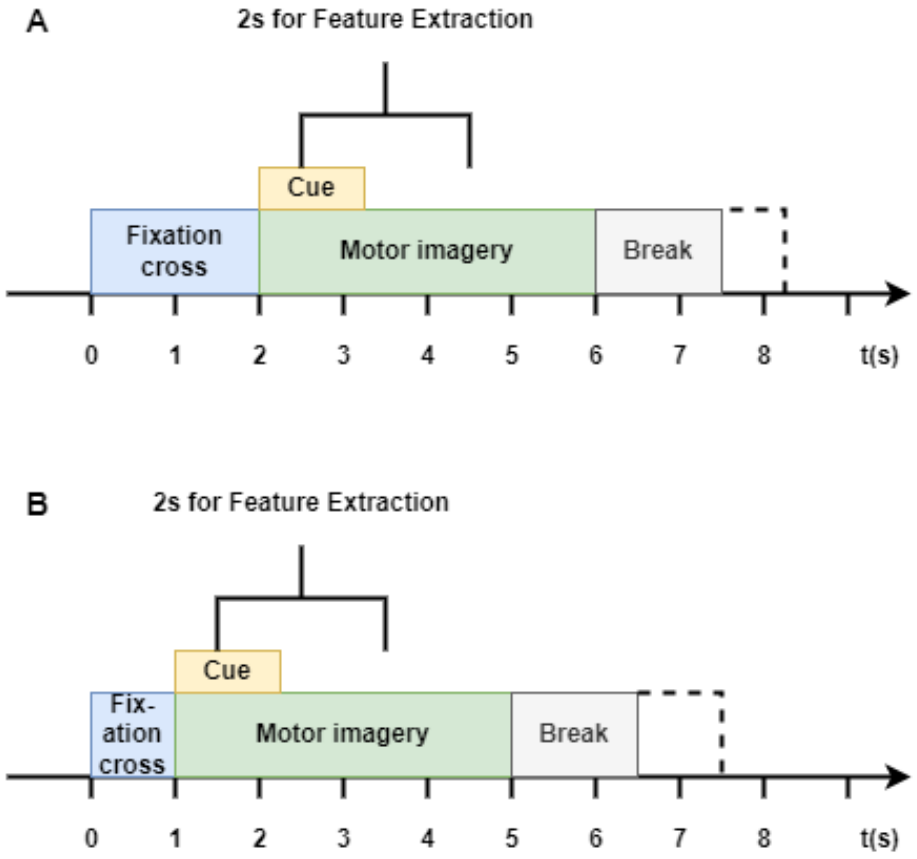


Figure 4.1: Paradigm A belongs to dataset 1 and has a preparation time of 2 seconds. Paradigm B belongs to dataset 2 and has a shorter preparation time of 1 second. Paradigms A and B have the same duration of motor imagery time. Features are extracted 0.5 seconds after the appearance of the cue in both paradigms.

4.2.2 Three Artificial Data Generation Methods

Segmentation and Recombination in Time Domain

The general idea of Segmentation and Recombination in the Time Domain (SRTD) [171] is the time series EEG data segmentation and recombination. The principle behind it is that when participants start performing MI tasks, the power of the sensory-motor rhythm (μ 8-12 Hz and β 16-24 Hz bands) drops due to a phenomenon called even related desynchronization [178] which in time domain manifest itself as a decreased amplitude of the wideband EEG signal [179].

In cue-based MI paradigms, the EEG of each MI trial is typically divided into epochs with respect to cue prior to the analysis. For SRTD, those 2s epochs can be further divided into 0.25s segments to keep consistent with Lotte's approach, and then new trials with equal time lengths can be generated from those segments (as shown in Fig. 4.2). Let's denote $\Phi = \{T_1, T_2, \cdots, T_N\}$ as the collection of preprocessed MI trials, where N is the number of trials. For $\forall i \in [1, N]$, $T_i \in R^{C \times S}$ contains data samples of the i-th trial, with C representing the channel number and S representing the number of time-domain trials. If we divide T_i into M consecutive and non-overlapping segments along with its time domain, i.e., $T_i = [T_i^1, T_i^2, \cdots, T_i^M]$ with $T_i^M \in R^{C \times \frac{S}{M}}$, a new artificial trail can be generated as $T_{arti} = [T_{s_1}^1, T_{s_2}^2, \cdots, T_{s_M}^M]$, where s_k is randomly selected from $[1, N]$ while preserving the timestamp k. Keeping the timestamp is important to preserve the ERD dynamics. The new artificial trial has the same number of samples as the real one to keep the same length. The channel order should be preserved (samples for an artificial EEG time series should be taken from one corresponding real EEG channel only); otherwise, the space information of EEG channels will be lost. Note that different classes of datasets (e.g., MI of the left and right hand) are processed separately for this artificial data generation method.

This is an extremely simple way for artificial EEG signal generation. However, the concatenation of EEG segments from different trials might cause high-frequency noise due to the inconsistency of EEG samples between boundaries.

Segmentation and Recombination in Time-Frequency Domain

Inspired by the paradigm presented in the time domain, the artificial data generation can be carried out in a similar way by Segmentation and Recombination in Time-Frequency Domain (SRTFD) [171]. Short-Time Fourier Transform (STFT) is an ideal tool to get time-frequency segments of equal size. The window size of the STFT is 50 samples(0.2s), and the step size is 25 samples (0.1s). Moreover, if such segmentation and recombination are conducted directly in a time-frequency

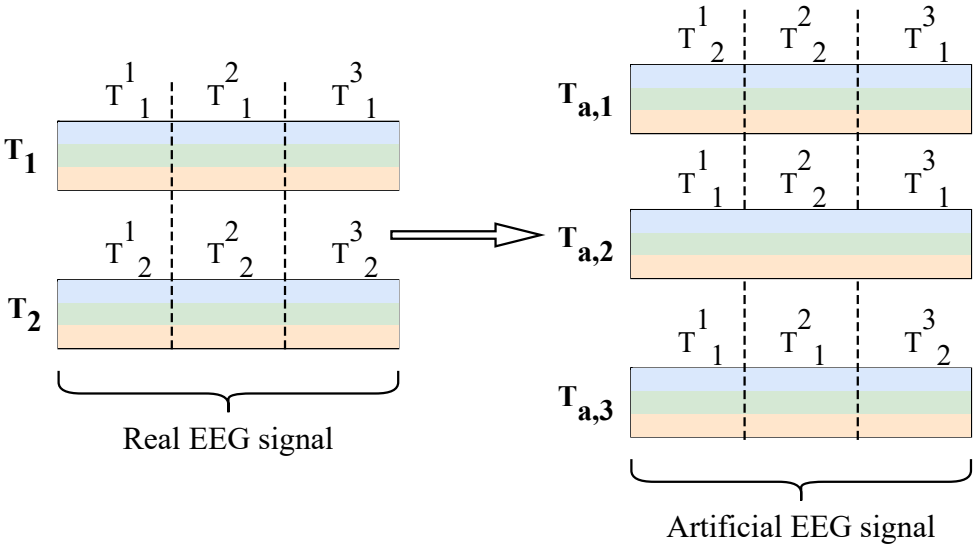


Figure 4.2: Example of generating artificial EEG signal in Time Domain with two real trails, where $T_1, T_2 \in \mathcal{R}^{3\times S}$ and M = 3 (Different colors denote different channels). Subscripts denote segment numbers, and subscripts denote trial numbers.

domain, the shortcoming of unwanted high-frequency noise that may occur in SRTD can be avoided.

We perform a channel-by-channel STFT for each segment T_i^m in T_i, where m is the index of the m-th segment and i is the index for the trail. The total number of trails is N. After the STFT, the time-frequency domain signal is generated as $FT_i^m = [FT_i^m(1); FT_i^m(2); \cdots; FT_i^m(C(channel))]$ with $FT_i^m(c) \in \mathcal{C}^{f\times w}$ being the STFT signal for the c-th channel of m-th segment in the i-th trail. The operation $[FT_i^m(1); FT_i^m(2)$ indicates the concatenation of rows in $FT_i^m(1)$ and $FT_i^m(2)$, and the value of f and w is decided by parameters of STFT, e.g., the Discrete Fourier Transform (DFT) points and the window size. An artificial trail can thus be generated as:

$$FT_{arti} = \begin{pmatrix} FT^1_{s_1}(1) & FT^2_{s_2}(1) & \cdots & FT^M_{s_M}(1) \\ FT^1_{s_{M+1}}(2) & FT^2_{s_{M+2}}(2) & \cdots & FT^M_{s_{2M}}(2) \\ \vdots & \vdots & \ddots & \vdots \\ FT^1_{s_{(C-1)M+1}}(C) & FT^2_{s_{(C-1)M+2}}(C) & \cdots & FT^M_{s_{CM}}(C) \end{pmatrix} \tag{4.1}$$

where each S_b, $(1 \le b \le C \times M)$ is a trial randomly (and independently) selected from all the trials. Then, the inverse STFT is applied channel by channel to transform FT_{arti} into the time domain and get the artificial trail. We demonstrate

the progress of this method in Fig. 4.3.

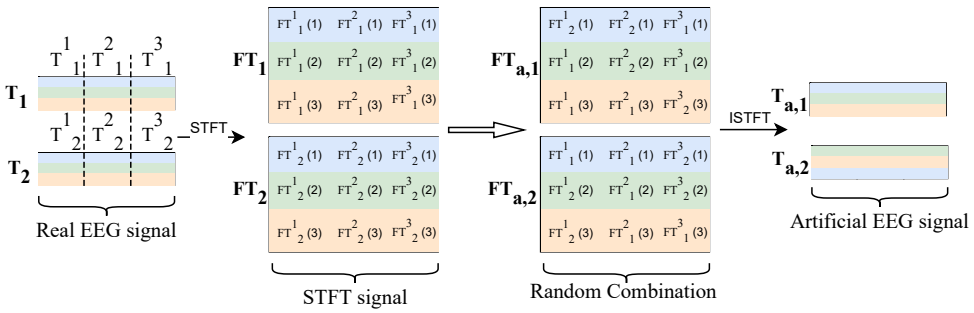


Figure 4.3: Example of generating artificial EEG signal by Segmentation and Recombination in Time-Frequency Domain with two real EEG trials, where $T_1, T_2 \in \mathcal{R}^{3\times S}$ and $M = 3$ is the Hypothetical number of segments in this figure's example within a trial, and S is the number of bins in frequency bins (Different colors denote different channels). Each "box" presents on time-frequency windows.

Spatial Analogy

The principle behind the Spatial Analogy (SA) [171] approach is to generate artificial trials according to the analogy found between the existing trials. Specifically, we first select three trials T_A, T_B, and T_C from the real data (including all channels). Then, the analogy between T_A and T_B is calculated. After that, the artificial trial T_D can be created so that the analogy between T_D and T_C is the same as that between T_A and T_B. To implement such a method, we use signal power along with the principal component as the measure of analogy between different trials. The procedure to create analogy-based artificial data generation in EEG signals is illustrated as follows:

1. Calculate the covariance matrix C for each class based on all the available data, and compute the eigenvectors V of C, which are also the principal components of the signal;

2. Randomly select three trails T_A, T_B, and T_C, and calculate the signal power of them alone each PC, i.e., $p_A(i)$, $p_B(i)$ and $p_C(i)$, where i denotes the i-th column of V;

3. Generate T_D by $T_D = T_C V P_A^{\frac{-1}{2}} P_B^{\frac{1}{2}} V^T$, where P_A, P_B, and P_C are diagonal matrices formed by $p_A(i)$, $p_B(i)$ and $p_C(i)$, respectively.

The procedure of this method is illustrated in Fig. 4.4.

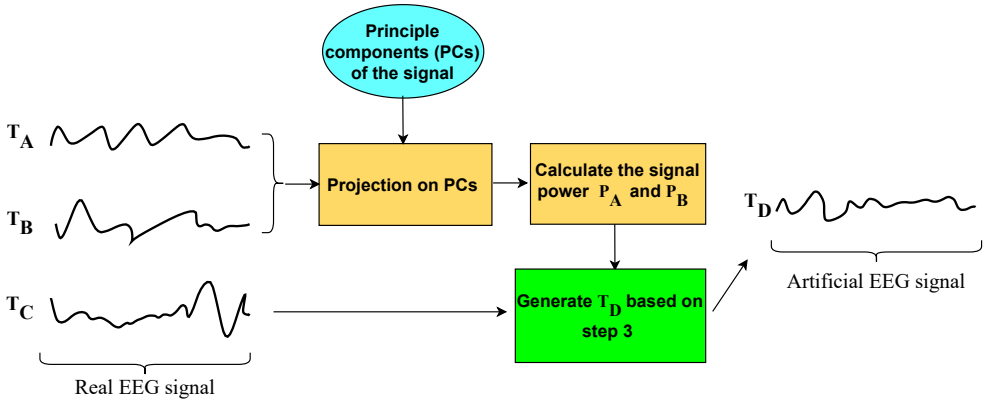


Figure 4.4: Principle of generating artificial EEG signal based on analogies.

4.3 Feature Extraction and Classification Method

4.3.1 Common Spatial Pattern

For each dataset, 22 electrodes based on a 10-20 system are used for extracting the CSP feature. The electrodes are Fz, FC3, FC1, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2, POz. The CPS filter maximizes the inter-class distance while minimizing the inner-class distance [180]. The input data matrix E is an $N \times T$ matrix where N is the number of samples per channel and T is the number of channels. Spatial covariances are calculated for each of the two classes (C1, C2) by averaging the covariances over successive trials of each class. The normalized spatial covariance of the EEG can be obtained from:

$$C = \frac{EE^T}{\text{trace}(EE^T)} \tag{4.2}$$

Then the composite spatial covariance is given by the sum of the two-class data's spatial covariance where C_l and C_r are covariances for left and right MI trials:

$$C_C = \bar{C}_l + \bar{C}_r. \tag{4.3}$$

C_C can be factored as:

$$C_C = U_C \lambda_C U_C^T, \tag{4.4}$$

where U_C is the matrix of eigenvectors and is the diagonal matrix of C_C, and λ_C contains corresponding eigenvalues. The whitening transformation:

$$P = \sqrt{\lambda_C^{-1}} U_C^T. \tag{4.5}$$

The sum of two corresponding eigenvalues is always one, i.e., $\lambda_l + \lambda_r = I$.

This property makes the eigenvectors useful for the classification of the two distributions. The projection of whitened EEG onto the first and last eigenvectors is optimal for the discrimination of two populations of EEG. The input data E can be transformed into the feature vector Z with a whitening matrix W, i.e., $Z = WE$. When synthetic data are created by any DA method, they have a different covariance matrix from the original data's covariance matrix.

Covariance matrix when DA trials are added

When synthetic data are created by any DA method, they have a covariance matrix that is different from the covariance matrix of the original data. Define C^{real} and C^{arti} as the average time-domain covariance matrix of the real and artificial of class j, respectively. The combination of them results in a new average time-domain covariance matrix C^{new}, which can be written as:

$$C_j^{new} = \frac{N_{real}}{N_{real} + N_{arti}} C_j^{real} + \frac{N_{arti}}{N_{real} + N_{arti}} C_j^{arti}, \tag{4.6}$$

where N^{real} and N^{arti} are the number of real and artificial trials, respectively.

The method we used to create the covariance matrix is the same as Lotte et al. Take dataset 1 as an example, it has 72 trials for each class. The following steps shows how the covariance matrices are calculated in two conditions.

Condition 1: Without artificial trails

Step1: calculate the covariance matrix for each trail in class 1. $E(i) = C(i)C(i)'$, C is a matrix with size $n * t$, t is the number of samples in each trail(500 sample points), n is the number of channels (22), and i is the number of trails (1-72).

Step2: calculate the average covariance matrix for 72 trails:

$C(class1) = mean(\sum_{i=1}^{72} C(i))$.

Step3: calculate the C' (class2).

Step4: [eigenvalues]=eig(C'(class1), C'(class2))

Condition 2: With artificial trails

It is obvious to see that if artificial trails are added, in the above step (2), the final C of each class is calculated for the average C of each signal trail, so it is related to the portion of artificial trails and real trails.

4.3.2 Band Power Feature Extraction Method

For dataset 1, 12 electrodes were used for 6 bipolar derivations: FC3-CP3, FCz-CPz, FC4-CP4, C3-P3, Cz-Pz, and C4-P4. Similarly, for dataset2 6 bipolar derivations were used FC3-CP3, FCz-CPz, FC4-CP4, C1-P1, Cz-Pz, and C2-P2. MI oscillatory features are more discriminable in the frequency domain than in the time domain due to event-related desynchronization/synchronization phenomena. The band power was computed in μ and β bands (8-12 Hz and 16-24 Hz) [75]. A 5th-order FIR bandpass filter is applied to the EEG signal to remove frequencies outside selected bands. The energy in selected bands was calculated by squaring the magnitude of the time series samples directly (i.e., each sample of the resulting signal is squared):

$$E[t] = y[t]^2, \tag{4.7}$$

where $E[t]$ is the energy of the signal and $y[t]$ is the signal containing the 8–12 Hz and the 18–24 Hz frequency components. With the smoothing window, w was set at 1s to reduce power fluctuation. This means that the band power for a certain sample equals the average power of w previous samples. The final feature is the logarithm of the band power.

4.3.3 Time Domain Parameter

Thirteen electrodes were used to extract these features, namely FC3, FCz, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CPz, and CP4. Vidaurre et al. provides the feature extraction method called TDP to measure the time properties of the MI EEG signals [176]. TDP gives the parameter as the variance of the first k derivatives of the EEG trace:

$$p_i(t) = \mathrm{var}\Big(\frac{d^i x(t)}{dt^i}\Big), \quad i = 0, \cdots, p, \tag{4.8}$$

where i is the i-th order derivatives. Notably, when $i = 0$, TDP can be regarded as the band power energy of the bandpass filtered EEG signal ($i = 7$ in this study).

4.3.4 Linear Discriminant Analysis

The classifier used for identifying features in this study is LDA [181]. The basic equation characterizing LDA can be presented as :

$$\mathbf{Y} = w\mathbf{X}(i) + b, \tag{4.9}$$

where w is the weights, and b is the bias. $\mathbf{X}(i)$ is the feature fed into the classifier. For training progress, a batch of labeled features is used to find the optimal w and b. These parameters can be later used for classification, where the classifier can predict the label for every new input feature.

4.4 Experimental Results

4.4.1 Testing Results

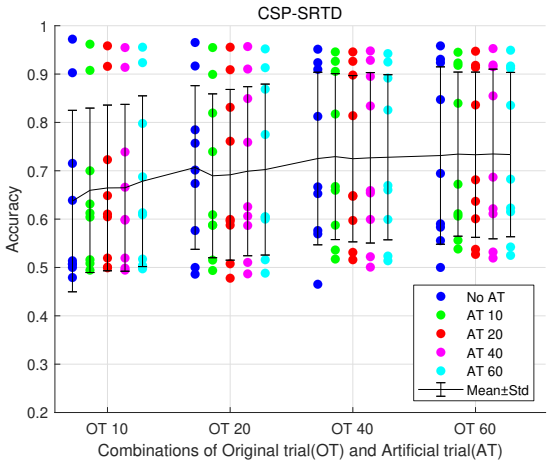
For dataset 1 without DA, average classification accuracy was in a range from 54.85±2.94% for 10 OT (TDP) to 73.15 ±17.29% with 60OT (CSP). For dataset 2, classification accuracy ranged from 52.96 ±1.64% for 10 OT (TDP) to 63.00 ±15.20% for 60 OT(CSP).

CSP features

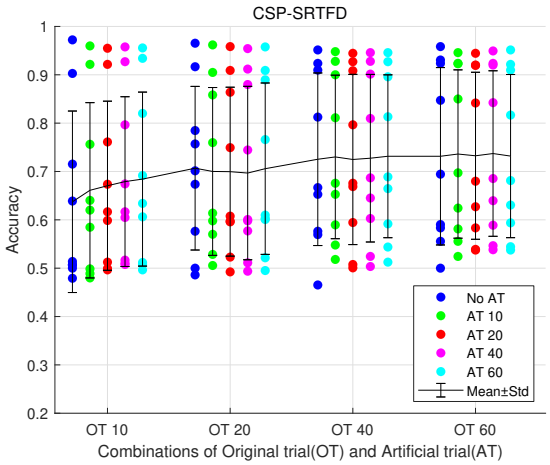
For dataset 1 (Fig. 4.5), the largest improvement was achieved for 10 OT for the average across 9 subjects, which has 5% for SRTFD (60 AT)(Fig. 4.5(b)), 4% for SRTD (60 AT)(Fig. 4.5(a)) and 2% for SA (60 AT)(Fig. 4.5(c)). For dataset 2 (Fig. 4.6) largest improvement of average accuracy was 2% for SRTD (10 OT, 50AT)(Fig. 4.6(a)), 1% for SRTFD (10 OT, 50 AT)(Fig. 4.6(b)) and 0.5% for spatial analogy (10 OT and 50 AT)(Fig. 4.6(c)). For both datasets, only 10 OT showed a consistent improvement in classification accuracy with DA. In contrast, classification accuracy decreased by about 2% but in an inconsistent manner for a larger number of OT or some combinations of OT and AT. For detailed information about classification accuracies, see Appendix. A. None of the improvements were statistically significant.

Band power features

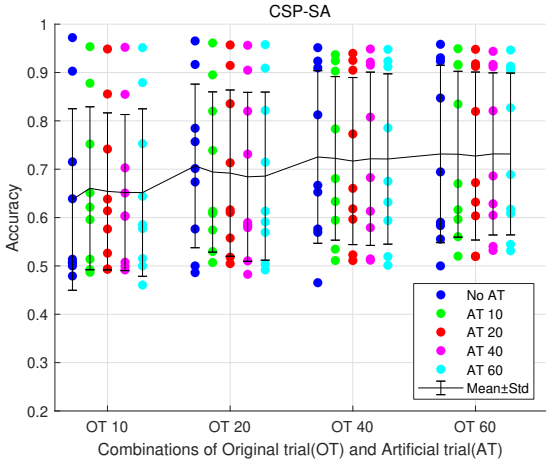
For dataset 1 (Fig. 4.7), the best improvement was achieved for 40 OT, 1% for SRTD (60 AT)(Fig. 4.7(a)), 1% for SRTFD (40 AT)(Fig. 4.7(b)) and no increase for SA(Fig. 4.7(c)). For dataset 2 (Fig. 4.8) best classification results were achieved for 10 OT: 2% for SA (10 AT)(Fig. 4.8(c)) , 1% for SRTFD (10 AT)(Fig. 4.8(b)) and 1% for SRTD (10 AT)(Fig. 4.8(a)) . For both datasets, only for 10 OT and 20 OT, there was an improvement in classification accuracy with DA, while for a larger number of OT, for some combinations of OT and AT, classification accuracy decreased by about 2% but in an inconsistent manner. None of the improvements were statistically significant. For detailed information about classification accuracies, see Appendix. A.



(a) Segmentation and Recombination in Time Domain(SRTD)

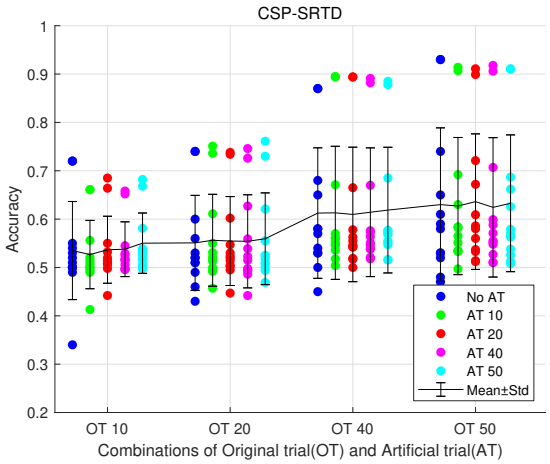


(b) Segmentation and Recombination in Time-Frequency Domain(SRTFD)

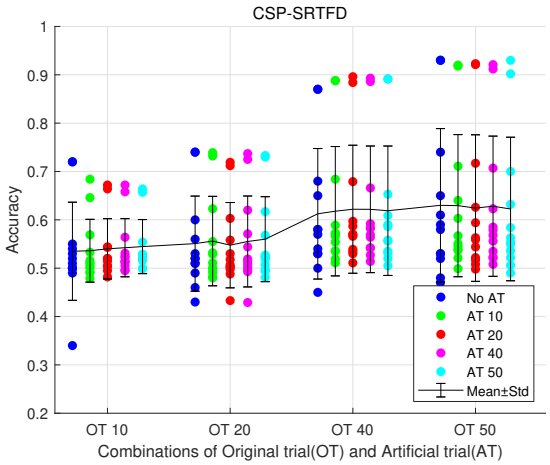


(c) Spatial Analogy(SA)

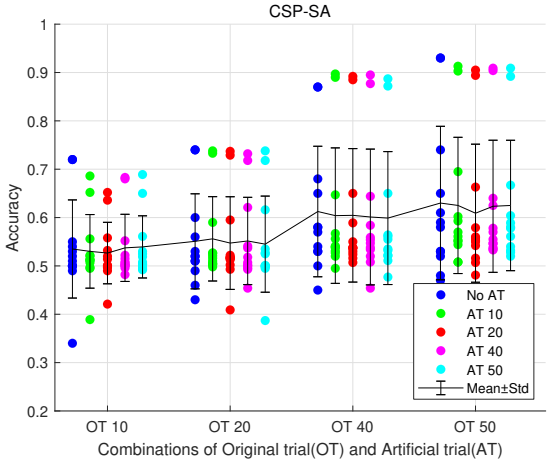
Figure 4.5: Dataset 1 Accuracy(%) of 9 participants' of different combinations of OT and AT for (a) SRTD (b) SRTFD (c) SA when using CSP feature extraction method. Mean and Standard deviation are presented with black lines.



(a) Segmentation and Recombination in Time Domain(SRTD)

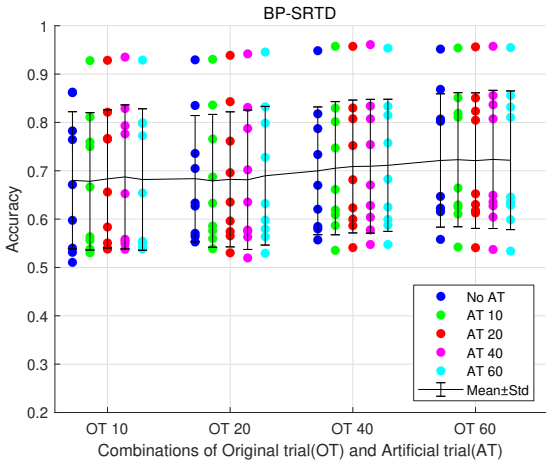


(b) Segmentation and Recombination in Time-Frequency Domain(SRTFD)

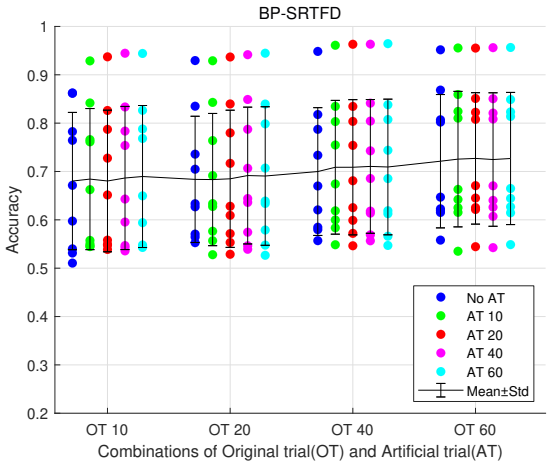


(c) Spatial Analogy(SA)

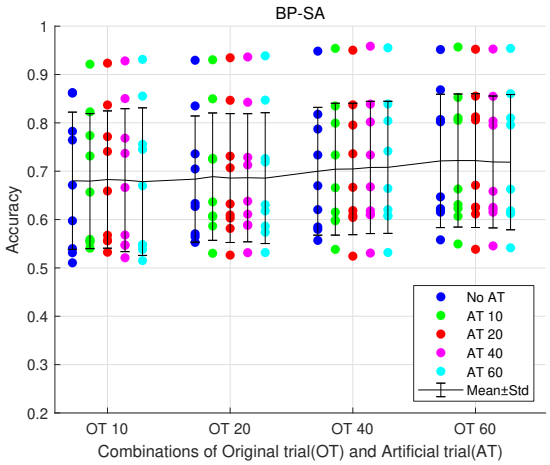
Figure 4.6: Dataset 2 Accuracy(%) of 12 participants' of different combinations of OT and AT for (a) SRTD (b) SRTFD (c) SA when using CSP feature extraction method. Mean and Standard deviation are presented with black lines.



(a) Segmentation and Recombination in Time Domain(SRTD)

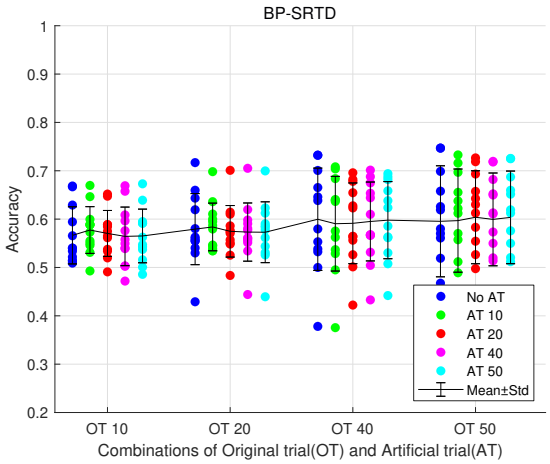


(b) Segmentation and Recombination in Time-Frequency Domain(SRTFD)

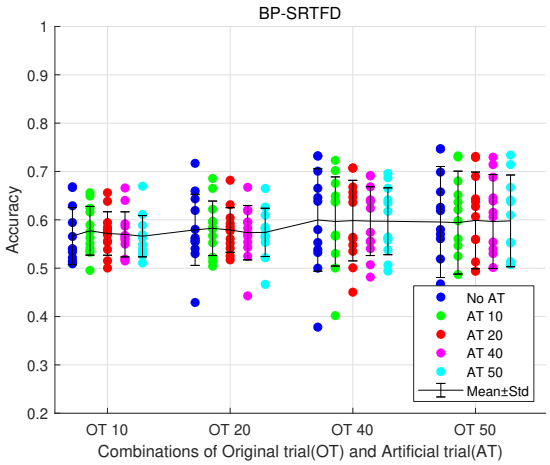


(c) Spatial Analogy(SA)

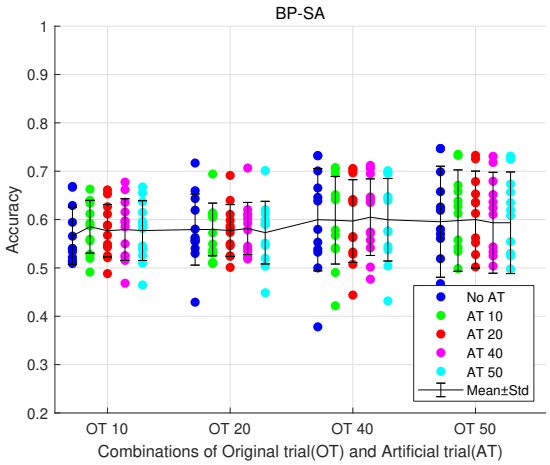
Figure 4.7: Dataset 1 Accuracy(%) of 9 participants' of different combinations of OT and AT for (a) SRTD (b) SRTFD (c) SA when using BP feature extraction method. Mean and Standard deviation are presented with black lines.



(a) Segmentation and Recombination in Time Domain(SRTD)



(b) Segmentation and Recombination in Time-Frequency Domain(SRTFD)



(c) Spatial Analogy(SA)

Figure 4.8: Dataset 2 Accuracy(%) of 12 participants' of different combinations of OT and AT for (a) SRTD (b) SRTFD (c) SA when using BP feature extraction method. Mean and Standard deviation are presented with black lines.

TDP features

For dataset 1 (Fig. 4.9), best improvement was achieved for 20 OT: 8% for SRTFD with 60 AT (55.81% to 63.81%) (Fig. 4.9(b)), 7% for SA with 60 AT (55.81% to 62.91%) (Fig. 4.9(c)) and 5% for SRTD with 60 AT (55.81% to 60.06%) (Fig. 4.9(a)). Consistent improvement was achieved for 10, 20, 40, and 60 OT. Statistically significant improvements were achieved with SRTD and SRTDF methods. A maximum statistically significant improvement of 6.2% was achieved for 10OT and 60 AT. For dataset 2 (Fig. 4.10) largest improvement was achieved for 10 OT and 50 AT, 4% for SA (52.96% to 56.66%)(Fig. 4.10(c)), 2.5% for time-frequency method (52.96% to 55.50%)(Fig. 4.10(b)) and 2% for SRTD (52.96% to 55.00%)(Fig. 4.10(a)). Consistent improvement was achieved for 10 and 20 OT. For exact numerical values, see Appendix. A.

In conclusion, TDP features offer much more consistent improvement than the other two features, with a ratio of OT to AT up to 1:6, and a very large increase in training dataset size due to DA.

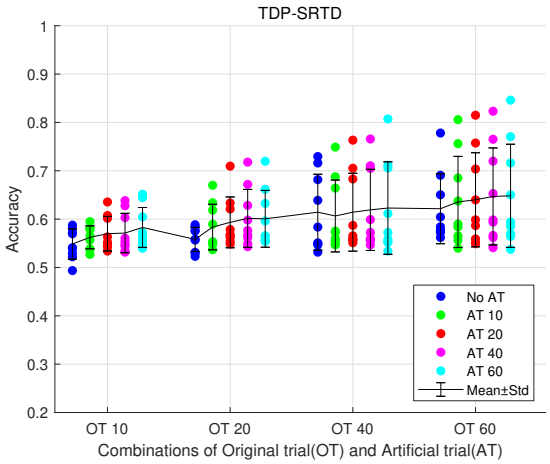
4.4.2 Statistical test results

Kruskal-Wallis tests were conducted on each feature extraction method (CSP/TDP/BP) for both datasets, see Appendix B. In dataset 1. The length oe compared to no AT adf OT was kept fixed, and for each DA method (SRTD/SRTFD/SA), different numbers of AT werded (AT=0) under the specific number of OT.

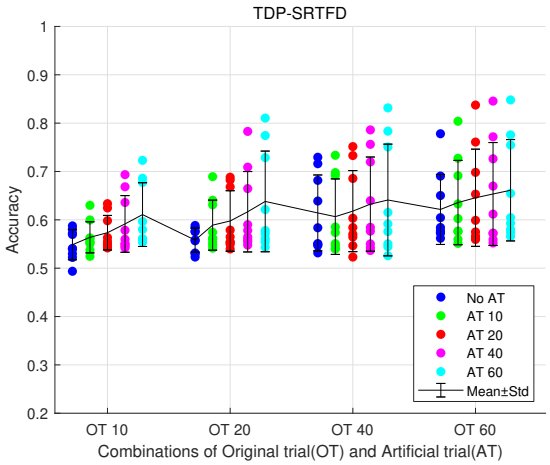
The analysis indicates that CSP and BP features were not statistically significant. However, for the TDP feature in dataset 1, there was a significant difference between OT=10/AT=60 and OT=10/AT=0 (p=0.047) for the SRTFD DA method. Another significant difference was observed for the SA DA method, where OT=20/AT=60 had a significant difference with OT=20/AT=0(p=0.014).

Regarding dataset 2, the KW test did not reveal statistical significance for CSP and BP features. For the TDP feature, statistical significance was observed for three DA methods when OT=10. For SRTD, 10AT, 20AT, 40AT, and 60AT showed significant differences with p values of 0.006, 0.002, 0.009, and 0.013, respectively, compared to no AT added. For SRTFD, p values were 0.015, 0.0567, 0.005, and 0.005, respectively, for four different ATs. For SA, the p-values were 0.149, 0.024, 0.001, and 0.002, respectively. It is noteworthy that only the SA method showed statistical significance when AT was added at OT=20. For four different numbers of AT, the p values were 0.133, 0.073, 0.038, and 0.049.

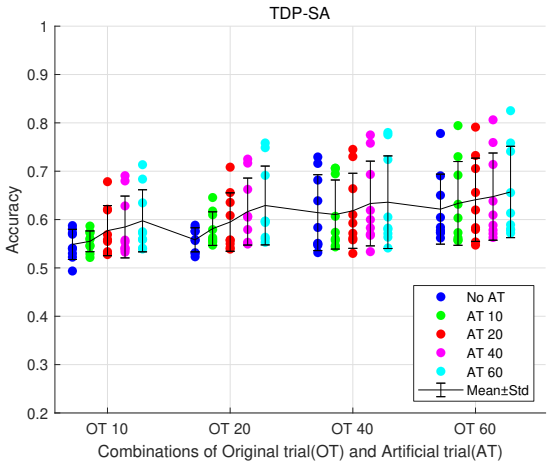
After the Holm-Bonferroni correction, only dataset 2 OT=10 remained statistically significant. For SRTD, all four numbers of AT remained significant,



(a) Segmentation and Recombination in Time Domain(SRTD)

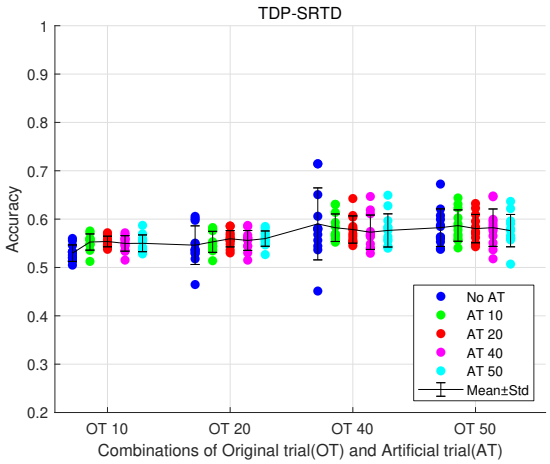


(b) Segmentation and Recombination in Time-Frequency Domain(SRTFD)

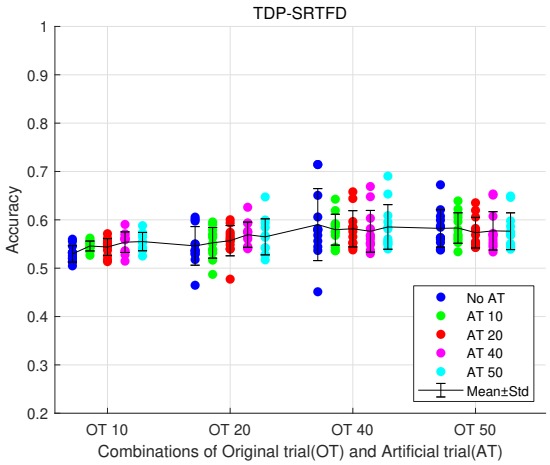


(c) Spatial Analogy(SA)

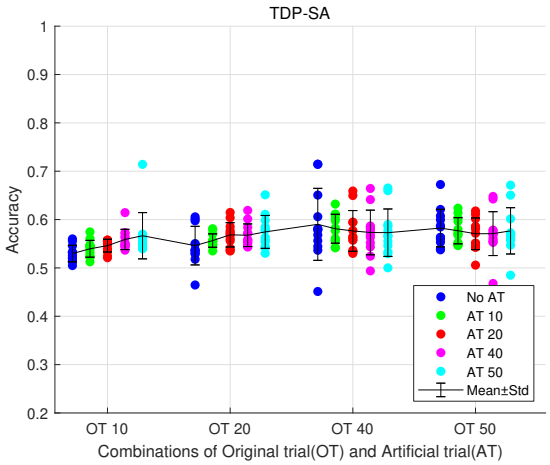
Figure 4.9: Dataset 1 Accuracy(%) of 9 participants' of different combinations of OT and AT for (a) SRTD (b) SRTFD (c) SA when using TDP feature extraction method. Mean and Standard deviation are presented with black lines.



(a) Segmentation and Recombination in Time Domain(SRTD)



(b) Segmentation and Recombination in Time-Frequency Domain(SRTFD)



(c) Spatial Analogy(SA)

Figure 4.10: Dataset 2 Accuracy(%) of 12 participants' of different combinations of OT and AT for (a) SRTD (b) SRTFD (c) SA when using TDP feature extraction method. Mean and Standard deviation are presented with black lines.

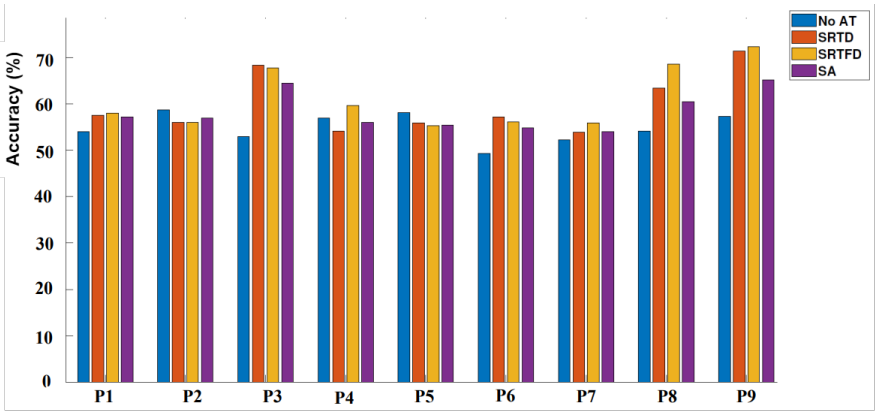


Figure 4.11: Individual Accuracy(%) of different DA methods, for TDP features for dataset 1 individual participants P1-P9. The number of original trials was OT=10 and the number of artificial trials was AT=60.

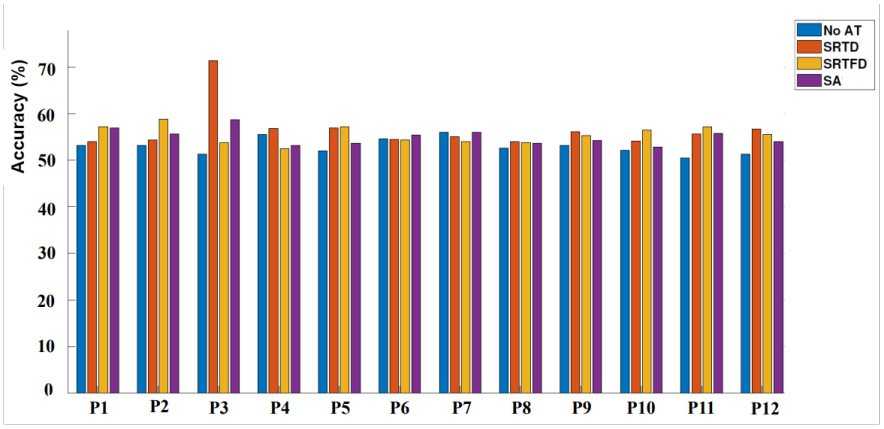


Figure 4.12: Individual Accuracy(%) of different DA methods, for TDP features for dataset 2 individual participants P1-P12. The number of original trials was OT=10, and the number of artificial trials was AT=50.

while for SRTFD, AT=10, 40, and 60 still showed statistical significance. For SA, AT=20, 40, and 60 remained statistically significant.

The statistical results showed that the addition of fake data had a significant effect when using the TDP feature extraction method compared to not adding fake data. Therefore, an example of the three DA methods was further demonstrated when using the TDP feature extraction method. Classification accuracies for 10 OT and 60 AT for dataset 1 in Fig. 4.11 for each individual participant to show between participants' variation. We also show similar results for dataset 2, for 10 OT and 50AT in Fig. 4.12.

In dataset 1, in 6 out of 9 participants, classification accuracy improved with all three DA methods, and in one, it improved with the SRTFD method only. Notably, in one participant accuracy dropped with all three DA methods. In dataset 2, classification accuracy improved in 9 out of 12 participants with all

three DA methods, while in two accuracies improved with at least one DA. In one participant, accuracy stayed the same with SA and dropped with SRTD and SRTFD.

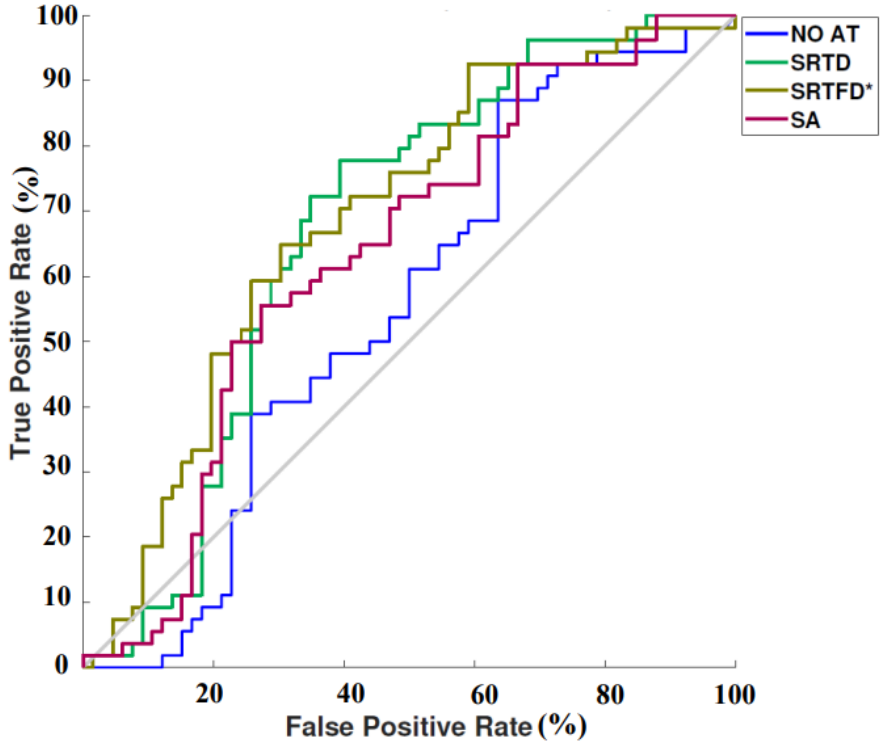
The Receiver Operating Characteristic(ROC) curve is a graphical representation of the performance of a binary classifier, plotted with TPR as the x-axis and FPR as the y-axis, where different threshold values selected from the predicted values of a classifier are used to calculate the TPR(True positive rate) and FPR(False positive rate) for each point on the curve. The ROC curve for 1 participant in each dataset was given as an example to show the classifier's performance with/without different artificial trials (Fig. 4.4.2). Participants with the highest true positive rate in both datasets were taken as an example to show the balance between the true positive and true negative classification accuracy. For dataset 1 participant P3 (Fig. 4.13(a)), all three DA have ROC located to the left from the diagonal, while for dataset 2 participants P3 (Fig. 4.13(b)), SRTD features had ROC on the let segmentation showing best performance with respect to false positives.

4.5 Discussion and Conclusions

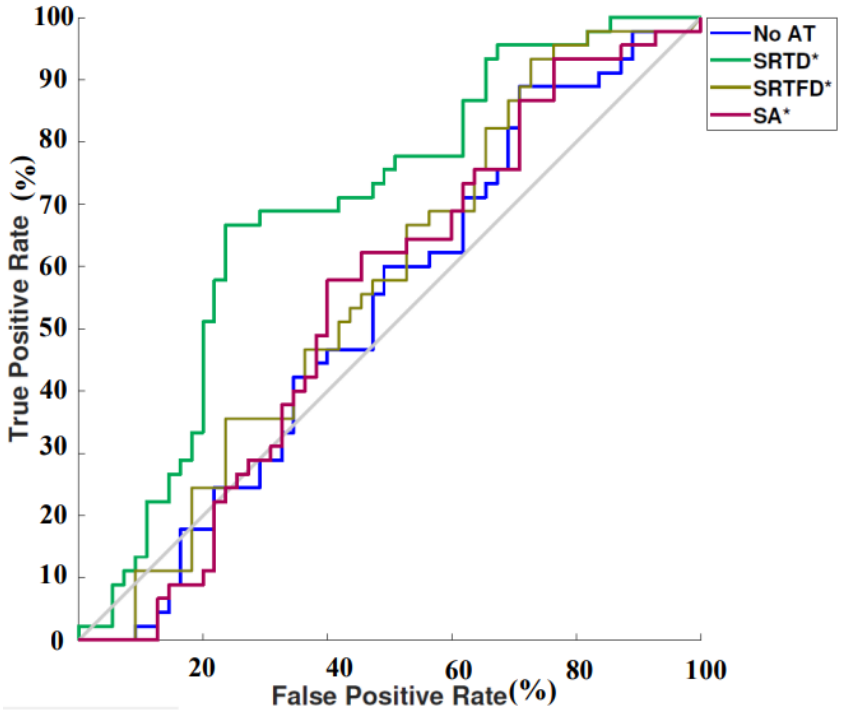
In this study, we analyzed the influence of different feature extraction methods on the improvement of classification accuracy with three different DA methods. We verified the results by testing them on two datasets, one from a BCI IV competition and our own dataset.

Through the observation of the results from two datasets, it can be noticed that when using the smallest number of OT(OT=10) and not adding AT, BP achieves the highest classification accuracy among the three feature extraction methods(68.02% for dataset 1 and 56.64% for dataset 2). However, when the number of OTs is 10 or 20, all three DA methods do not significantly improve the classification accuracy. Moreover, adding AT when increasing the number of OT not only fails to improve classification accuracy but also causes it to decrease. This phenomenon is consistent across all three DA methods. Thus, it can be concluded that these three DA methods are not suitable for the BP feature, and there is no significant difference between them for the BP feature.

When the number of OT is 10 or 20, the CSP feature has close classification accuracy compared with BP(70.08 %for dataset 1, 55.68% for dataset 2). It is worth noting that CSP uses the largest number of electrodes among the three features, so BP has the advantage to reach near classification accuracy. In dataset 1, adding AT can improve classification accuracy when the number of OTs is low,



(a) Dataset1 P3's ROC



(b) Dataset2 P3's ROC

Figure 4.13: ROC of one subject in each dataset: SRTD means Segmentation and Recombination in Time Domain, SRTFD means Segmentation and Recombination in Time-Frequency Domain, and SA means Spatial Analogy.

while in other cases, a limitation of DA methods similar to that of BP is observed. This limitation is consistent across all three DA methods in both datasets.

The original classification accuracy of the TDP feature is lower than that of BP and CSP(54.96% for dataset 1 and 52.85% for dataset 2). However, the TDP feature shows consistent sensitivity to adding AT in both dataset 1 and dataset 2, especially when the OT value is 10 and 20. All three DA methods can improve the classification accuracy of TDP feature, which is particularly significant in dataset 2. However, when the value of OT is larger, the TDP feature behaves similarly to BP and CSP, indicating the ineffectiveness of DA methods. Therefore, although the original classification accuracy of the TDP feature is lower than that of the other two features, it is more advantageous in terms of improving classification accuracy under the influence of DA methods. In dataset 1, when OT is 10, SRTFD achieved a 6.23% improvement in initial accuracy from 54.85% to 61.08%, which is a 2% improvement over SA's 59.74% and SRTD's 58.28%. Similarly, when OT is 20, SRTFD still outperforms others with a higher accuracy of 63.81%, which is an increase from the initial accuracy of 55.81%.In dataset 2, SA improved the initial accuracy from 52.96% to 56.66%, which is a 1% advantage over SRTD's 55% and SRTFD's 55.50%. Similarly, when OT is 20, SA's accuracy achieved the best improvement from 54.61% to 57.45%.

This is a meaningful discovery because previous studies have mentioned that 40 is the minimum number of trials required to obtain satisfactory MI BCI performance. To reduce calibration time, more attention should be paid to research on MI trials with less than 40 trials. Our study reveals that the TDP feature has advantages over BP and CSP, two commonly used research methods, in terms of sensitivity to DA methods.

We shall address here the research questions from the Introduction. The first research question referred to the best features. Results of the study showed that statistically significant improvements were achieved only for TDP and for small datasets with 10 to 2O OT. That also answers the second research question about the interaction between the DA methods and features. Although we could not perform a multivariate ANOVA due to non-normal data distribution, results indicate no obvious interaction. Significant improvement was achieved for all three DA methods, for TDP features only. For dataset 1's TDP feature extraction method, best improvement was achived by SRTFD(OT=20,AT=60),from 55.81% to 63.81%. For dataset 2, the largest improvement was made by the SA for TDP method, from 52.96% to 56.66%(OT=10,AT=50). In dataset 1, significant improvement was achieved with SRTD and SRTFD, while in dataset 2, significant improvement was achieved with all three DAs before the Holm-

Bonferroni correction. Only dataset 2 remains significant after the multiple comparison correction. The results of this study outperformed the original study by Lotte [14] due to novel features. In this study, for the same dataset, the biggest statistically significant improvement was 3.6% with TDF features(dataset 2, SA, OT=10, AT=50). Further improvement might be possible to achieve by applying optimization methods [182] to Lotte et al. achieved about 2% improvement with 10 OT and 500 AT for CSP features, but they did not perform the test of statistical significance. In the current study, the best improvement with CSP features was 6%, but it was not statistically significant due to the large variance between participants. The percentage of improvement of classification accuracy achieved with DA in this study is comparable with results from the literature. However, most previous studies were based on CSP features and therefore used a much larger number of channels as compared to TDP (Lotte 2015, Pei 2021, Xiong and Wei 2022) [14, 170, 173]. The results of this study are comparable with Xiong and Wei [173], who achieved an improvement of 6.9% with CSP and LDA classifiers, and with Pei et al. [170], who achieved an improvement of about 7% with CSP and SCM classifiers. We used 86% artificial data (10 OT with 60 AT), comparable with Dinares-Ferran et al. [183], which used 87.5% artificial data aiming to keep classification error under 33%. The third research question was related to the optimal ratio between OT and AT. We found that as much as 6 times the size of original data for small datasets improves the accuracy the most, which was a consistent result for both datasets. We did not test the largest ratio, but there might be a limit in artificial data where DA becomes too similar, leading to overfitting. Our results are in accordance with results from Dinares-Ferran et al. [183], who showed that 87.5% of original data could be replaced with synthetic data without significantly affecting the classification accuracy. In Lotte et al. [14], up to 500 AT were added to the original 10 OT. Since they achieved a modest improvement of 2% with no statistical significance testing, it is hard to say if getting so many AT may result in overfitting, at least in this basic approach, before applying transfer learning and other optimization methods. The final research question refers to the method's robustness when tested on two different datasets. Results were consistent for both datasets. TDP outperformed CSP and BP features for all DA methods in both cases in terms of improvement under the effect of three DA methods. Classification of the BCI competition dataset resulted in higher classification accuracy than our laboratory's dataset. This points out an important issue of validating methods for BCI competition. Although it is necessary to use standard datasets for comparison, they present an ideal scenario, with experiments performed in highly specialized labs with a low

noise level and with participants experienced in BCI. Our dataset was collected on BCI naïve participants, whose focus on the task and motivation might be lower than participants in the competition dataset. Best classification accuracy with SCP features, LDA classifier, and 10-fold classification was up to 75% [177]. The main limitation of this study is low initial classification accuracy with TDF features that did not reach 70%, which is believed to be minimal accuracy to achieve reasonable online control. In the future, additional optimization methods as proposed in [14] should be applied to boost the initial accuracy without AT. However, we believe that it is worthwhile to focus on TDF features, which require 30% fewer electrodes than CSP features. Many affordable commercially available EEG devices have a smaller number of electrodes, and in general, the price of an EEG device is proportional to the number of EEG electrodes. Focusing on DA, which requires a smaller number of channels, would facilitate the adoption of BCI by a larger number of users, including the patient population. A method proposed in the current study is suitable in cases when MI BCI should be applied to patients unable to tolerate long calibration procedures with many EEG electrodes or when data from other similar participants are not available or where due to large inter-individual variability (e.g. stroke population) user-to-user transfer may not possible.

Chapter 5

Data Augmentation Based on Generative Adversarial Networks with Transfer Learning and Cropped Training Strategies on Motor Imagery Data

5.1 Introduction

5.1.1 Exploring New Techniques for Generating EEG Signals and Overcoming Challenges in BCI Applications

In the previous chapter, traditional signal processing techniques were utilized, such as time domain or time-frequency domain decomposition and recombination, to generate artificial EEG signals. However, this approach relies on manual feature extraction and subjective combination, which may introduce limitations and biases. Moreover, the traditional methods may not be able to capture complex non-linear relationships in the EEG signals, which can limit their accuracy and reliability. In order to overcome these challenges, transfer learning techniques and GAN networks are employed to generate synthetic EEG signals. These methods not only highlight continuous exploration and innovation of new techniques but also contribute to the research objectives of the thesis.

Several factors can affect BCI performance and cause difficulties in practical use. Firstly, Conducting MI-BCI experiments requires strict experimental

conditions, such as avoiding laboratory noise or signal interference and instructing users to avoid body and even eye movements during the recording period. Secondly, the long calibration period for acquiring enough data frequently makes subjects feel tired and impatient. Thirdly, EEG signals are changing with subjects' mental and physical states. Sometimes, identifiable MI patterns obtained in one experimental session may not be applicable to another session. Last but not least, the MI features are person specific. Subjects need to do calibration individually, resulting in the low efficiency of making use of existing calibration EEG data.

5.1.2 Challenges and Solutions for Reliable EEG-Based Analysis

EEG signals are known for their instability, which is evident not only within the same individual across multiple recording sessions but also across different individuals. There can be significant variations in the extracted EEG features, which can lead to poor reproducibility of the results. This variability can be attributed to several factors, including differences in electrode placement, head geometry, individual brain anatomy and function, and the presence of noise and artifacts in the recorded signal. As a result, careful consideration of experimental design and rigorous data analysis techniques are required to address these issues and ensure the reliability of EEG-based analyses. Based on these backgrounds, researchers are striving to find the balance between better calibration performance and user-friendliness. The most common approaches are different attempts at feature extraction methods or classifiers. Previous research has proved that at least 40 trials per class are required for effective classification in MI BCI while using the traditional ML method. Data augmentation has been proposed for increasing the number of subject-specific training datasets for ML. Transfer Learning is another solution to reduce the calibration demand for a new BCI user by pre-training BCI with a dataset of other subjects doing the same MI task. Generation of new samples from existing samples by time, frequency shifting, or spatial transform or GAN is called DA. ML needs enough training trials to achieve high classification accuracy. However, acquiring large amounts of high-quality biomedical and medical data is a challenging task. On one hand, obtaining and sharing medical data is restricted by laws and ethics in order to protect patient privacy. On the other hand, biomedical signal data is often expensive and complex, requiring specialized equipment and expertise for collection and processing. Therefore, acquiring training data in biomedical and medical fields

may be more challenging than in other fields.

This chapter takes a different approach from prior papers that use deep networks to analyze vast amounts of existing data and then facilitate the classification of new subject data by learning features from the extensive data. Instead, the new user's data is directly merged with the existing data and trained in a Generative Adversarial Network (GAN). In comparison to pre-training on large-scale datasets, this method utilizes the GAN's generative modeling capacities to grasp the underlying distribution of the data and generate fresh samples that closely resemble the distribution of the target domain. Overall, We combined cropped training, TL, and GAN to generate artificial trials. Finally, we investigated the test accuracy of MI trials when only original trials without artificial trials, original trials with artificial trials, and only artificial trials.

5.2 Method

5.2.1 Experimental Paradigm

The data used for the study is BCI competition IV data 2a. It's a 4-class motor imagery dataset of 9 subjects sampling at 250hz. Each class has 72 training trials and another 72 trials for testing, i.e. validation. In the study, only two classes, left-hand MI and right-hand MI, out of four classes were chosen for DL. These datasets were filtered by 4-38hz and then epoched as 4s-long trials. Notably, these trials were further extracted into 600-sample long cropped trials. All 22 EEG channels were fed into CNN.

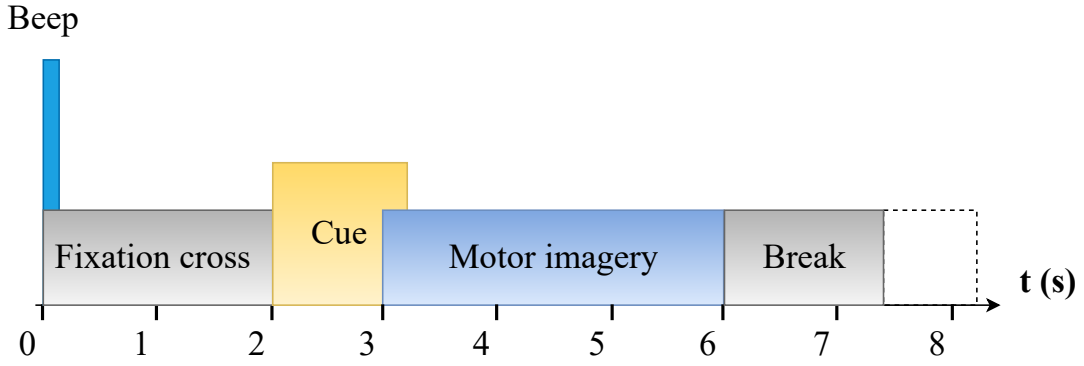


Figure 5.1: BCI Competition IV 2a data experimental paradigm.

The subjects were sitting in a comfortable armchair in front of a computer screen. At the beginning of a trial ($t = 0s$), a fixation cross appeared on the black screen. In addition, a short acoustic warning tone was presented. After two seconds ($t = 2s$), a cue in the form of an arrow pointing either to the left,

right, down, or up (corresponding to one of the four classes left hand, right hand, foot, or tongue) appeared and stayed on the screen for 1.25 s. This prompted the subjects to perform the desired motor imagery task. No feedback was provided. The subjects were asked to carry out the motor imagery task until the fixation cross disappeared from the screen at $t = 6s$. A short break followed when the screen was black again. The paradigm is illustrated in Figure 5.1.

Both ShallowFBCSPNet and Deep4Net can be used for MI EEG decomposition and classification. GAN used for DA in this study is learned from the structure of the Deep4Net. The ShallowFBCSPNet is used as the classifier for two classes of MI data to avoid repetition for each subject. Only the left and right classes of BCI competition IV dataset 2a were used in this study. There were 72 trials for training and 72 trials for testing for each class. Since GAN is extremely data-hungry, each original trial with 1000 sample points was cropped into a 600-sample cropped trial for both training and testing datasets. The step of the crop windows is 66 samples; thus, each trial yielded 7 cropped trials, and the last 4 samples were dropped. Finally, 504 cropped training trials were used to train the shallowFBCSPNet. The 504 cropped testing trials(72 original testing trials for each class) were used to test the classifier's performance. Besides, 280 cropped training trials were used to train the ShallowFBCSP, and the same 504 cropped testing trials were used for testing to investigate the performance of the ShallowFBCSP when a small number of training trials were available. The table shows the results of different conditions.

5.2.2 End-to-end Convolutional Neural Network

The Deep4Net and ShallowFBCSPNet were initially introduced by Robin Tibor et al. [83]. These networks adopt an end-to-end deep learning approach that allows the convolutional neural network (CNN) to directly learn from raw EEG data, thus avoiding the need for separate feature extraction steps.

Deep4Net

There are four main convolutional parts for Deep4Net. The first part contains two convolutional layers to handle the large data volume for raw EEG data. The input shape of the first convolutional layer is a matrix of four dimensions $(N, 1, N_t, N_c)$. The N is the number of trials for one batch of training. $N_t \times N_c$ is a typical EEG data matrix, where N_t is the time bins for one trial, and N_c is the channel number. In $(N, 1, N_t, N_c)$, the second dimension of 1 is a hypothetical spatial dimension. In reality, EEG signals are two-dimensional, namely (Nt,

Nc). However, data fed into DCGAN always contain three dimensions, height, weight, and length. The kernel size of this convolutional layer is 10×1. The second convolutional layer operates as a special filter with kernel size $(1 \times N_c)$. A Batch Normalization (BN) layer follows these two convolutional layers. The function of Batch Normalization is to pull the increasingly irregular distribution back to the standardized distribution so that the value of neurons falls into the sensitive area of the activation function. After the BN layer, the neurons pass through a nonlinear activation function, ELUs. The kernel in the MaxPooling layer preserves the largest value within its coverage after the BN layer, thus reducing the computing complexity for later layers. The kernel size of Deep4Net is (3×1), and the stride is (3×1). The second part of the Deep4Net starts with a Dropout layer with a dropout rate of 0.5, which means half of the neurons are randomly deleted from the input to avoid overfitting. Then a convolutional layer with a kernel size (of 10,1) is added. The next layers are similar to the first part: the BN layer, ELU(), and the MaxPooling layer. The next two parts have a similar structure as part two while having different input and output sizes. It should be noted that the last part is a Dence layer with softmax as an activation function for presenting classification results.

ShallowFBCSPNet

The ShallowFBCSPNet is a network with fewer layers consisting of two parts. The first part is two convolution layers with kernel sizes (25×1) and $(1 \times N_c)$. After the BN layer, the activation function is Square () instead of ELU(). A MaxPooling layer follows them with kernel size (75×1) and stride(15×1), an activation function Log(), and a dropout layer with a dropout rate of 0.5. The structure of the second part of ShallowFBCSP is the same as the last part of the Deep4Net. A softmax activation follows a Dence Layer to present the probability to a certain class.

Cropped training

The cropped training method was also proposed by Tibor et al. to compensate for DL's data-hungriness for the Deep4Net [83]. Shallow networks(ShallowFBCSPNet) typically require fewer data than deeper networks because they have fewer layers and therefore fewer parameters to learn. However, the amount of data needed ultimately depends on the complexity of the task and the specific architecture of the network. A time window with a length of about 2s is used to extract sub-trials from each trial. By doing this, the number of training trials fed into the network

Table 5.1: Deep4Net

	Deep4Net
Part1	25× Conv2D (10,1), Stride1, 1 25× Conv2D (1,n_c), Stride1, 1 BatchNorm Activation (ELU) MaxPool (3,1), Stride3, 1
Part2	Dropout (0,5) 50× Conv2D (10,1), Stride1, 1 BatchNorm Activation (ELU) MaxPool (3,1), Stride3, 1
Part3	Dropout (0,5) 100× Conv2D (10,1), Stride1, 1 BatchNorm Activation (ELU) MaxPool (3,1), Stride3, 1
Part4	Dropout (0,5) 200× Conv2D (10,1), Stride1, 1 BatchNorm Activation (ELU) MaxPool (3,1), Stride3, 1
Part5	Dense Softmax Classification

Table 5.2: ShallowFBCSPNet

	ShallowFBCSPNet
Part1	40× Conv2D (25,1), Stride1, 1 40× Conv2D (1,n_c), Stride1, 1 BatchNorm Activation (Square) MaxPool (75,1), Stride5, 1 Activation (Log) Dropout(0.5)
Part2	Dense Softmax Classification

increases. For BCI competition IV dataset 2a, each trial is extracted 4s after the direction cue(left, right up, down). The sampling frequency is 250Hz, thus yielding 1000 time points. A time window with a length of 600 sample points (around 2.5s) was used to slide along the single trial with a stride producing a certain amount of cropped trials with only one original trial. Notably, the final class of one trial in the test stage is decided by the average value of all the cropped trials in a trial. However, in this study, different window sizes and strides are used and the final classification accuracy was not given by the average label over a trial but directly presented with cropped trials, since cropped trials lead to more training trials for DCGAN when compared with using original trials.

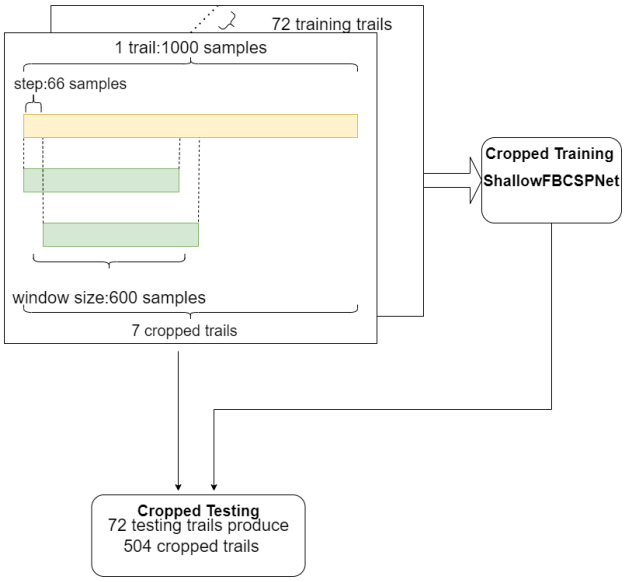


Figure 5.2: Cropped training.

5.2.3 Transfer learning

The training strategy for artificial data generation combines the TL with the cropped training. Thus, the DCGAN would have sufficient trials to learn. When it comes to TL, the purpose is to improve the classification performance of the target subject, and the source dataset contains all other existing people's experiment data (see Tab. 5.3). If subject one is selected as the target subject, only 280 cropped trials from 40 original trials (for each class) were used, then all the training trials of the rest eight subjects (572 original training trials) yielded 572*7 cropped trials. The target subjects cropped, and source cropped trials (280+572*7=4312) were prepared for the DCGAN.

Table 5.3: Transfer Learning

Target Data (280 cropped trials)	Source Dataset (572*7 cropped trials)
S1	S2, S3, S4, S5, S6, S7, S8, S9
S2	S1, S3, S4, S5, S6, S7, S8, S9
S3	S1, S2, S4, S5, S6, S7, S8, S9
S4	S1, S2, S3, S5, S6, S7, S8, S9
S5	S1, S2, S3, S4, S6, S7, S8, S9
S6	S1, S2, S3, S4, S5, S7, S8, S9
S7	S1, S2, S3, S4, S5, S6, S8, S9
S8	S1, S2, S3, S4, S5, S6, S7, S9
S9	S1, S2, S3, S4, S5, S6, S7, S8

5.2.4 Generative Adversarial Networks

The generator's(G) aim is to create artificial samples, thus maximizing the misclassification of the discriminator. The equation is

$$\mathbb{E}_{\mathbf{z}\sim p_z(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))], \tag{5.1}$$

where z is the random noise, and G(z) is the generated artificial trial. $1-D(G(\mathbf{z}))$ is the probability of artificial trials identified as original trials. The discriminator, however, is trying to improve its classification accuracy between artificial trials and real trials. $D(x)$ is the probability that x is the original data. Hence, the zero-sum game can be defined as the equation:

$$\min_G \max_D V(G, D) = \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\log(D(\mathbf{x}))] + \mathbb{E}_{\mathbf{z}\sim p_z(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))]], \tag{5.2}$$

where V represents the value function, and E represents the expected value. X is the real data, z is the random noise vector, and P() is the distribution.

Deep Convolutional Generative Adversarial Networks

DCGAN is one of the most commonly used networks for image and video areas. It consists of convolution layers without max-pooling of fully connected layers. The convolutional layers in DCGAN extract useful features of each part of the image. The feature generated by those filters ensures the quality of the generated features, hence DGGAN achieves better performance for image features or hidden space. Hartmann et al. proposed a DCGAN with opposition in layer structure between the generator and discriminator (Table. 5.4) [184]. The Deep4Net has already been applied to identify the features of MI data. A DCGAN oriented from Deep4Net is proposed for artificial data generation. The generator has the

	Generator				Discriminator		
Layer Number	Layer Type	Act./Norm.	Output shape		Layer Type	Act.	Output shape
1	Latent vector	-	200 X 1		Input signal	-	1 X 768
	Linear	LReLU	50 X 12		Conv 1	LReLU	50 X 768
2	Upsample		50 X 24		Conv 9	LReLU	50 X 768
	Conv 9	LReLU/PN	50 X 24		Conv 9	LReLU	50 X 768
	Conv 9	LReLU/PN	50 X 24		Downsample	-	50 X 384
3	Upsample	-	50 X 48		Conv 9	LReLU	50 X 384
	Conv 9	LReLU/PN	50 X 48		Conv 9	LReLU	50 X 384
	Conv 9	LReLU/PN	50 X 48		Downsample	-	50 X 192
4	Upsample	-	50 X 96		Conv 9	LReLU	50 X 192
	Conv 9	LReLU/PN	50 X 96		Conv 9	LReLU	50 X 192
	Conv 9	LReLU/PN	50 X 96		Downsample	-	50 X 96
5	Upsample	-	50 X 192		Conv 9	LReLU	50 X 96
	Conv 9	LReLU/PN	50 X 192		Conv 9	LReLU	50 X 96
	Conv 9	LReLU/PN	50 X 192		Downsample	-	50 X 48
6	Upsample	-	50 X 384		Conv 9	LReLU	50 X 48
	Conv 9	LReLU/PN	50 X 384		Conv 9	LReLU	50 X 48
	Conv 9	LReLU/PN	50 X 384		Downsample	-	50 X 24
7	Upsample	-	50 X 768		Conv 9	LReLU	50 X 24
	Conv 9	LReLU/PN	50 X 768		Conv 9	LReLU	50 X 24
	Conv 9	LReLU/PN	50 X 768		Downsample	-	50 X 12
8	Conv 1	-	1 X 768		Linear	-	1 X 1

Table 5.4: An example of EEG-DCGAN [184]

opposite structure of the Discriminator.

DCGAN Discriminator

The Discriminator of this DCGAN has the same structure as Deep4Net, which provides good criteria for the generator to learn. The input of the discriminator is $(N, 1, 600, 22)$, N is the batch size, 1 is the spacial dimension, 522 is the time length of the crop, and 22 is the channel number. The activation function for the last layer is Sigmoid().

DCGAN Generator

The generator produces artificial trials of the same size as the real cropped trials to feed into the discriminator. The hidden layers use the feedback of the loss function to adjust the parameters in layers. The generator can generate fake data that increasingly resemble the original data by iteration using only random noise. The input size of the random noise is (1,100,2,1). In the hidden layers, ConTransposed2D layers are used for expanding the tensors' size. Continuous using the Con2dTranspose layers broadcast the input dimension of (1,100,2,1) to (1,1,600,22). The activation function of Convolutional layers is ELU(), and

Table 5.5: Discriminator in DCGAN

Structure	Input dimension	Layers	Output dimension
Part1	(1*,1,600,22)	Conv2D(kernel size(10,1),stride(1,1))	(1,25,591,22)
	(1,25,591,22)	Conv2D(kernel size(1,22),stride(1,1))	(1,25,591,1)
	(1,25,591,1)	BatchNorm()	(1,25,591,1)
	(1,25,591,1)	ELU()	(1,25,591,1)
	(1,25,591,1)	MaxPool2d(kernel size(3,1),stride(3,1))	(1,25,197,1)
Part2	(1,25,197,1)	Dropout(p=0.5)	(1,25,197,1)
	(1,25,197,1)	Conv2D(kernel size(10,1),stride(1,1))	(1,50,188,1)
	(1,50,188,1)	BatchNorm()	(1,50,188,1)
	(1,50,188,1)	ELU()	(1,50,188,1)
	(1,50,188,1)	MaxPool2d(kernel size(3,1),stride(3,1))	(1,50,62,1)
Part3	(1,50,62,1)	Dropout(p=0.5)	(1,50,62,1)
	(1,50,62,1)	Conv2D(kernel size(10,1),stride(1,1))	(1,100,53,1)
	(1,100,53,1)	BatchNorm()	(1,100,53,1)
	(1,100,53,1)	ELU()	(1,100,53,1)
	(1,100,53,1)	MaxPool2d(kernel size(3,1),stride(3,1))	(1,100,17,1)
Part4	(1,100,17,1)	Dropout(p=0.5)	(1,100,17,1)
	(1,100,17,1)	Conv2D(kernel size(10,1),stride(1,1))	(1,200,8,1)
	(1,200,8,1)	BatchNorm()	(1,200,8,1)
	(1,200,8,1)	ELU()	(1,200,8,1)
	(1,200,,1)	MaxPool2d(kernel size(3,1),stride(3,1))	(1,200,2,1)
Part5	(1,200,2,1)	Flatten()	(1,400)
	(1,400)	Linear()	(1,1)
	(1,1)	Sigmoid()	(1,1)

the last activation function is tanh().Tanh() is one of the most commonly used activation functions in the generator as the final activation layer since it can map the neurons to a range of (-1,1). Notably, the real trials fed into the generator are processed with standardization to the range of [-1,1]. The Tanh() and standardization can ensure that both the real and artificial trial samples are in the same distribution.

DCGAN Training Phase

4300 cropped trials are selected from 4312, divided by the batch size of 50. The number of training epochs is 200. Adam optimizer is used for both generator and discriminator, the learning rate of the optimizer is 0.0002 and 0.0001, respectively. Finally, the generator model is saved to generate new artificial trials. The first step of the training loop is to feed the discriminator a batch of real samples. Then the loss of real samples is calculated from the label of this batch of real samples and the output of the discriminator (activation function is sigmoid). The second step is to give a batch of random noise to the generator to produce the same number of artificial samples. The discriminator also computes the loss of these artificial samples. The final loss of the discriminator should be the sum of the real samples and fake samples. The loss between the value of fake samples evaluated

Table 5.6: Generator in DCGAN

Structure	Input dimension	Layers	Output dimension
Part1	(1,100,2,1)	ConvTranspose2D(kernel size(3,1),stride(3,1))	(1,100,6,1)
	(1,100,6,1)	BatchNorm()	(1,100,6,1)
	(1,100,6,1)	ELU()	(1,100,6,1)
Part2	(1,100,6,1)	ConvTranspose2D(kernel size(12,1),stride(1,1))	(1,100,17,1)
	(1,100,17,1)	BatchNorm()	(1,100,17,1)
	(1,100,17,1)	ELU()	(1,100,17,1)
Part3	(1,100,17,1)	ConvTranspose2D(kernel size(3,1),stride(3,1))	(1,100,51,1)
	(1,100,51,1)	BatchNorm()	(1,100,51,1)
	(1,100,51,1)	ELU()	(1,100,51,1)
Part4	(1,100,51,1)	ConvTranspose2D(kernel size(10,1),stride(1,1))	(1,50,62,1)
	(1,50,62,1)	BatchNorm()	(1,50,62,1)
	(1,50,62,1)	ELU()	(1,50,62,1)
Part5	(1,50,62,1)	ConvTranspose2D(kernel size(3,1),stride(3,1))	(1,50,186,1)
	(1,50,186,1)	BatchNorm()	(1,50,186,1)
	(1,50,186,1)	ELU()	(1,50,186,1)
Part6	(1,50,186,1)	ConvTranspose2D(kernel size(10,1),stride(1,1))	(1,25,197,1)
	(1,25,197,1)	BatchNorm()	(1,25,197,1)
	(1,25,197,1)	ELU()	(1,25,197,1)
Part7	(1,25,197,1)	ConvTranspose2D(kernel size(3,1),stride(3,1))	(1,25,591,1)
	(1,25,591,1)	BatchNorm()	(1,25,591,1)
	(1,25,591,1)	ELU()	(1,25,591,1)
Part8	(1,25,591,1)	ConvTranspose2D(kernel size(1,22),stride(1,1))	(1,25,591,22)
	(1,25,591,22)	ConvTranspose2D(kernel size(10,1),stride(1,1))	(1,1,600,22)
	(1,1,600,22)	Tanh()	(1,1,600,22)

by the discriminator and real labels is a generator loss. The Deep4Net-based generator in DCGAN of this study has the ability to identify MI features from raw EEG signals. However, when training the GAN, it can be easily seen that the loss of discrimination stays in low value and the loss of discriminator stay in high value. This phenomenon proves that the Deep4Net Based discriminator brings too strict standards for the generator to learn. The strategy is to set the learning rate of the generator greater than the learning rate of the discriminator.

DCGAN Validation for artificial data

The final step of artificial data generation is to provide the pre-saved generator models with random noise to produce fake samples for each class. It should be noted that left and right MI cropped trials are trained separately since DCGAN cannot generate trials with labels, and the corresponding generator generates artificial trials. Since the trials used for DCGAN are the cropped trials, the artificial trial generated is the same size as the copped trial(600*22). Three quantities of artificial trials for each class(504/1008/1512) are appended to the initial 280 cropped trials to train the ShallowFBCSP classifier and used the same amount of cropped(504) trials to test the classifier.

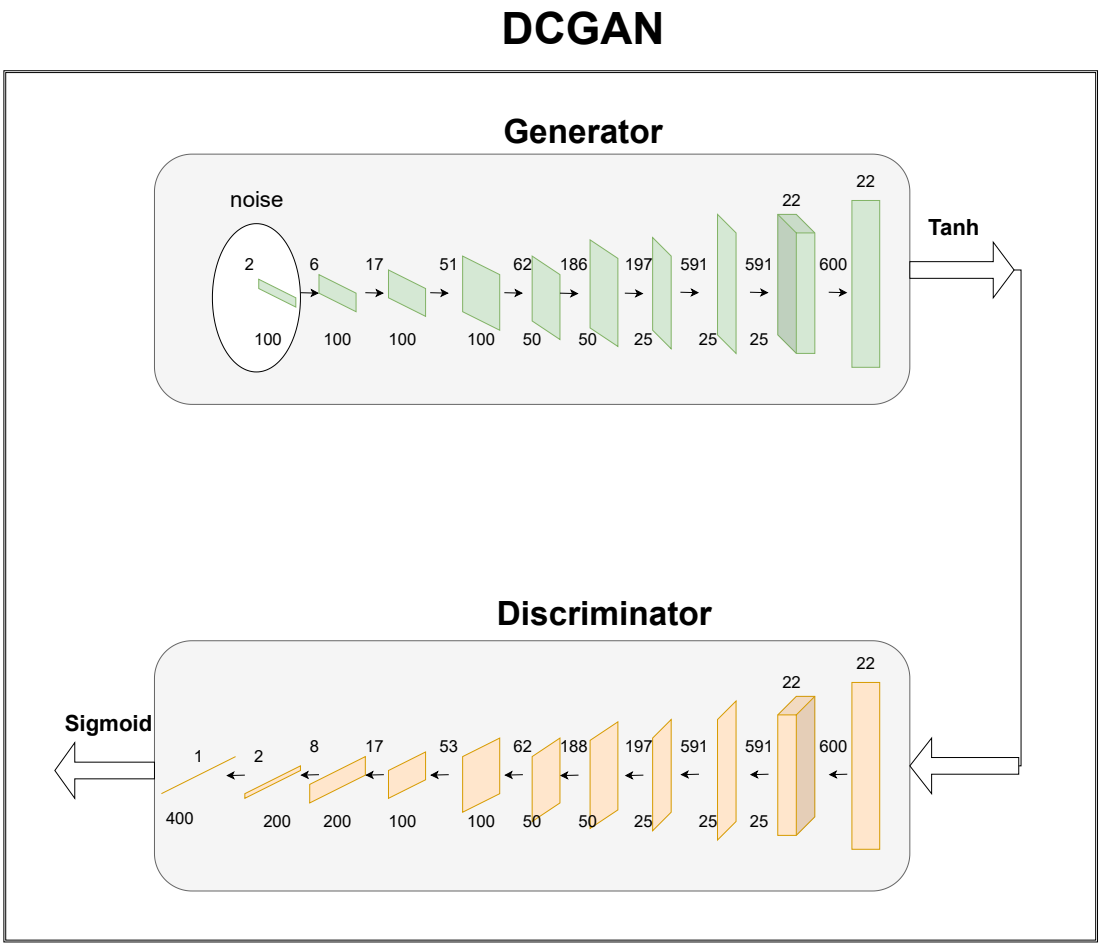


Figure 5.3: DCGAN used for artificial data generation.

5.3 Results

The ShallowFBCSP classifier was employed as a classifier to verify the classification accuracy. Three different scenarios were discussed: the first involved the use of sub-trials obtained by transforming the original trial data for classification(Table 5.7, Fig 5.4). The second involved the classification of sub-trials with the addition of synthetic data(Table 5.8, Fig 5.5). The final scenario focused on the classification results using only synthetic data(Table 5.9, Fig 5.7). Based on the classification results, the quality of generated artificial EEG trials is discussed. The characteristics of the generated MI trials are observed from three perspectives: time domain, frequency domain, and CSP filters.

5.3.1 ShallowFBCSPNet Classification Accuracy without Data Augmentation

Table 5.7: Classification testing results of ShallowFBCSP

Subject/ Classification trials	Training:280 cropped trials (40 original trials) Testing: 504 cropped trials	Training:504 cropped trials (72 original trials) Testing: 504 cropped trials
S1	77.88%	83.23%
S2	50.60%	54.56%
S3	54.86%	94.25%
S4	63.39%	69.64%
S5	56.25%	70.54%
S6	62.70%	69.54%
S7	68.95%	80.06%
S8	97.76%	92.56%
S9	80.75%	82.24%
Mean±SE	68.24±4.59%	77.40±4.20%

From the estimation results of the ShallowFBCSPNet classification, it can be concluded that the average estimation accuracy of 9 subjects is rising with the increasing amount of training trials. The table shows that for a few subjects, e.g., S2 and S4, the network cannot identify the feature, and the testing accuracy is always below the average accuracy, which might be caused by the variation of EEG among people.

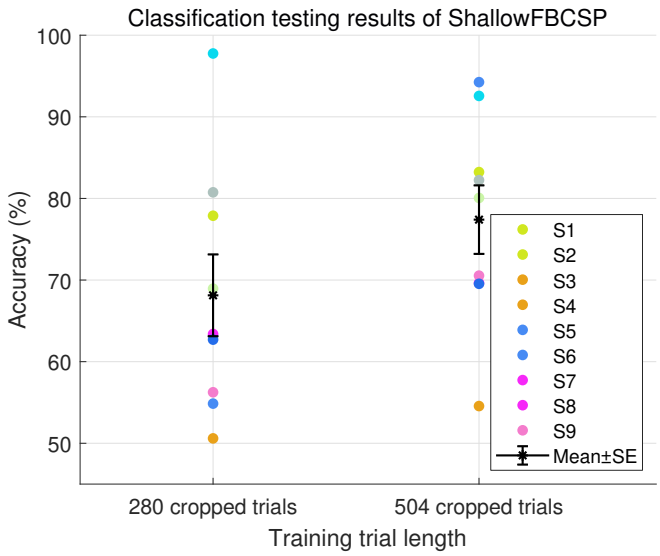


Figure 5.4: Classification testing results of ShallowFBCSP

Table 5.8: Test accuracy of the ShallowFBCSP classification with artificial trials

Subject/ Classification trials	Training: 280 cropped trials + 0 artificial trials Testing: 504 cropped trials	Training: 280 cropped trials + 504 artificial trials Testing: 504 cropped trials	Training: 280 cropped trials + 1008 artificial trials Testing: 504 cropped trials	Training: 280 cropped trials + 1512 artificial trials Testing: 504 cropped trials
S1	77.88%	74.70%	79.27%	80.26%
S2	50.60%	53.17%	50.20%	49.90%
S3	54.86%	78.37%	83.53%	87.30%
S4	63.39%	58.04%	59.23%	59.33%
S5	56.25%	58.13%	58.53%	58.83%
S6	62.70%	57.04%	55.46%	57.94%
S7	68.95%	69.94%	69.44%	71.73%
S8	97.76%	91.37%	91.87%	92.76%
S9	80.75%	72.82%	79.27%	78.27%
Mean±Standard Deviation	68.24±4.59%	68.17±4.19%	69.64±4.84%	70.70±4.97%

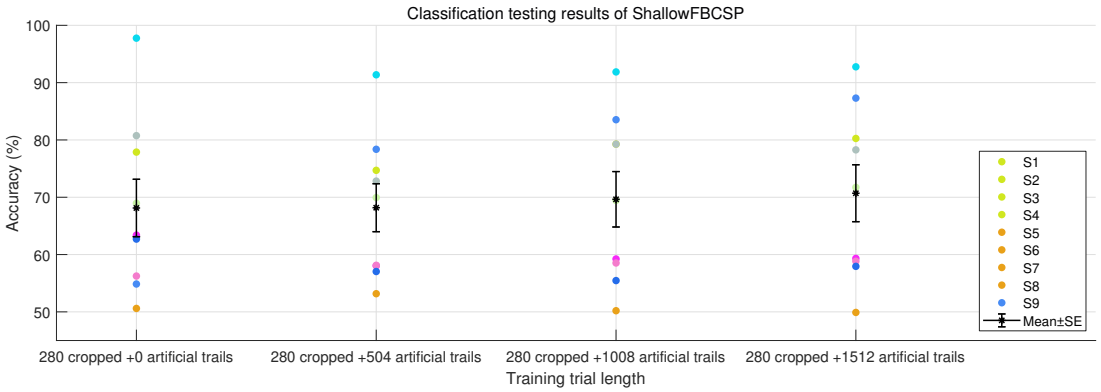


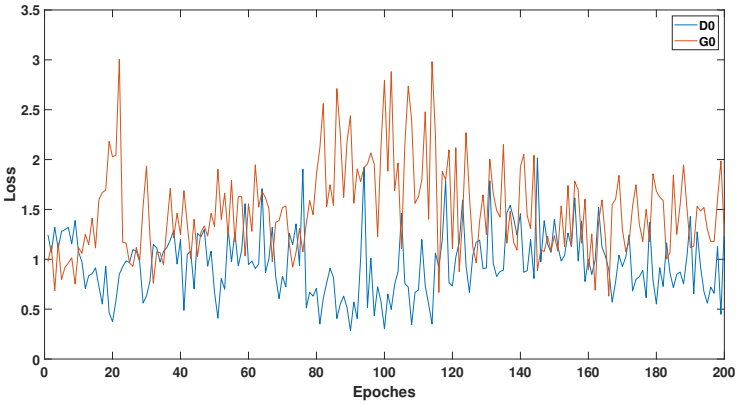
Figure 5.5: Test accuracy of the ShallowFBCSP classification with artificial trials

5.3.2 ShallowFBCSPNet Classification Accuracy with Data Augmentation

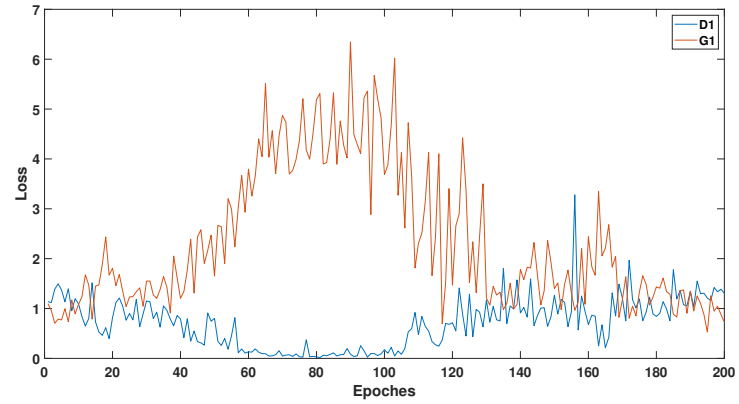
To explore the performance of the classifier when adding artificial trials to a small amount of original cropped trials, three amounts of artificial trials appended are compared with no artificial trials appended. The average classification accuracy of 9 subjects only shows a 2% improvement. Kolmogorov Smirnov test is performed on four groups of data. The results suggest that the data are not normally distributed. Then the non-parametric test Kruskal Wallis test is performed on the four data groups. The result with $p > 0.05$ indicates no significance among these four groups. For S1, S3, and S7, the artificial trials benefit the classifier training. For S3 especially, the artificial trials even bring a 35% increase to the baseline classification accuracy. However, the artificial trials even decrease the classifier's performance for the rest subjects. Fig. 5.6 shows the loss change of the Generator and Discriminator during the GAN training progress of S3, where G0/D0 represents the loss of left-hand MI trials generation/discrimination and G1/D1 represents the loss of right MI trials generation/dissemination. Convergence of the loss curve is crucial for GAN network stability. For the loss curve shown in Fig. 5.6(b), the discriminator's loss decreases while the generator's loss increases after 20 epochs, indicating that the discriminator can distinguish between real EEG data and fake data generated by the generator. As a result, the generator's performance decreases, and it cannot produce convincing fake data. However, after 100 epochs, the generator's performance improves, and the fake data it generates deceives the classifier, causing the classifier's loss to increase. Eventually, the generator's and classifier's loss converge, indicating that the GAN has reached equilibrium and cannot improve further. 4300 cropped TL trials are trained(batch= 50, training epoch=200), and the 200 points on the figure only show the batch loss at the end of every epoch.

5.3.3 ShallowFBCSPNet Classification Accuracy with only Data Augmentation

From the table 5.9, it can be seen that the classification is all around the chance level. In this part, the training process excludes the target subject data's cropped trials but only the artificial trials generated by the minimal number of targets and lots of source data. However, the test trials are from the target subject. The result shows that the artificial data generated from source data are not applicable for the direct classification of the new target person. It proves the variation of MI patterns among subjects since the S3 shows no increase this time. The better



(a) Training loss of left hand MI in 200 epoches



(b) Training Loss of right hand MI in 200 epoches

Figure 5.6: DCGAN training Loss of S3. In (a), the classifier outperforms the generator starting at around 70 epochs, while the generator's performance improves at around 100 epochs. Finally, the generator and classifier converge stably. The corresponding time points are around 40 epochs and 100 epochs in (b).

Table 5.9: Test accuracy of the ShallowFBCSP classification (artificial trials only)

Subject/ Classification trials	Training: 280 cropped trials + 0 artificial trials Testing: 504 cropped trials	Training: 504 artificial trials Testing: 504 cropped trials	Training: 1008 artificial trials Testing: 504 cropped trials	Training: 1512 artificial trials Testing: 504 cropped trials
S1	77.88%	49.50%	49.40%	48.81%
S2	50.60%	50.00%	49.40%	51.88%
S3	54.86%	43.45%	42.16%	43.15%
S4	63.39%	50.50%	49.31%	50.10%
S5	56.25%	50.00%	49.80%	49.60%
S6	62.70%	48.02%	50.00%	50.00%
S7	68.95%	48.31%	46.33%	48.12%
S8	97.76%	52.10%	50.34%	52.13%
S9	80.75%	49.40%	48.80%	49.10%

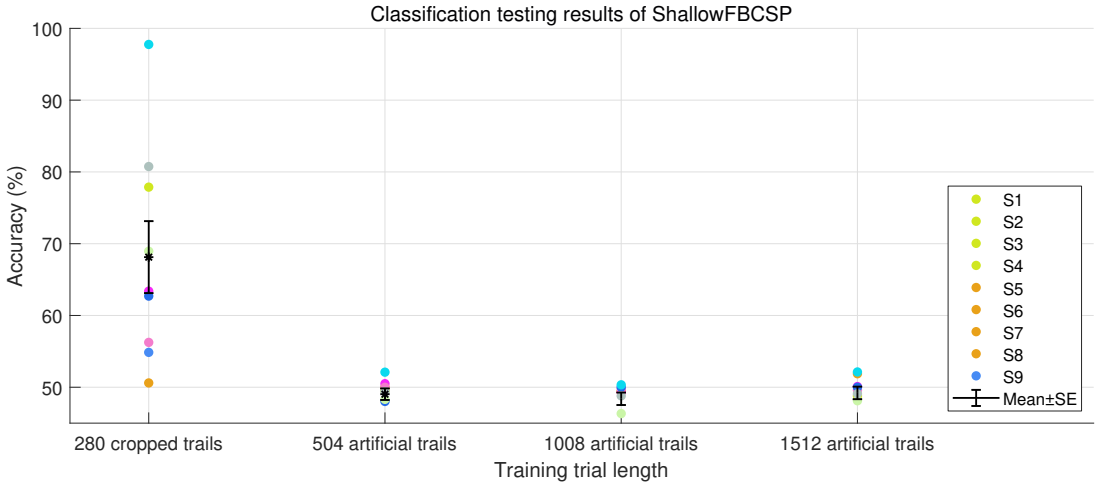


Figure 5.7: Test accuracy of the ShallowFBCSP classification (artificial trials only)

way should be as the table. 5.8, which includes the target data in the training process.

5.3.4 Evaluation for generated artificial trials

The results demonstrate the potential of Generative Adversarial Networks (GANs) for generating artificial data, however, the generated Motor Imagery (MI) trials for most participants exhibit poor quality. This section aims to investigate the quality of these artificial trials. The commonly used metrics such as Inception score(IS) and Fréchet inception distance (FID) for evaluating the quality of synthetic samples generated by GANs depend on the ImageNet pre-trained Inception network which is not suitable for temporal EEG data. Thus, generated trials are compared with the real ones in the Time domain, time-frequency domain, and their CSP filters. S1 to S3's (Fig. 5.5) data are scrutinized specifically due to their disparate initial classification performances and varying susceptibilities to artificial data. Notably, S1 exhibited a significantly higher initial classification accuracy of 77.88%(280 cropped trials) in comparison to S2's 50.60%. Nevertheless, the incremental addition of artificial data (1008 artificial trials added) led to a mere 2% uptick in S1's classification accuracy, culminating in an 83.23% classification accuracy. This implies that the strategic employment of synthetic data yields minimal gains in scenarios where the initial classification accuracy is relatively robust. Conversely, S2's initial classification accuracy approximated the chance level, and despite efforts to escalate the number of cropped trials (from 280 to 504), the accuracy rating was only marginally enhanced to 54.56%. This result is reflective of the daunting classification

challenge posed by this subject, rendering the addition of copious artificial data an ineffectual measure, with the classification accuracy hovering around the chance level. In contrast, S3's initial classification accuracy resembled that of S2 but was inferior to S1. However, the infusion of synthetic data precipitated a noteworthy uptick from 54.86% to 87.30%.

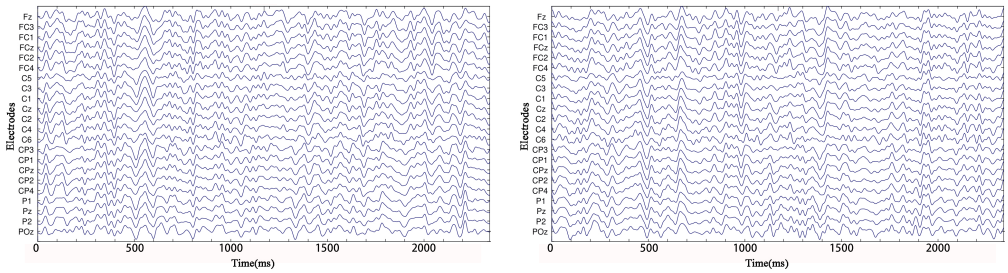
Time domain observation

One artificial trial from each class of S1 to S3 was randomly selected and compared with the real trials in the time domain. It should be noted that the real sub-trials were standardized to [-1,1] to ensure consistency with the artificial trial exported from the activation function tanh() in the GAN generator. The comparison between the real and generated trials in MI (motor imagery) in S1 to S3 (Fig. 5.8 - Fig. 5.10) reveals that in the time domain, the synthetic EEG signals appear to have waveform similarity with the generated EEG signals. However, as depicted in Fig. 5.8(c), the artificial data exhibits high-frequency noise around 0ms and low-frequency noise around 2200ms. Thus, compared to the real EEG trials, the generated signals are more susceptible to noise interference. Furthermore, the blank space on the top of S1's artificial trial(Fig. 5.8(d)) indicates that the generated signal produces a negative bias, whereas the real EEG signal should fluctuate around 0 μV. Similarly, the high-frequency noise near 0ms and 2200ms can also be observed in S2's fake trial (Fig. 5.9). The blank space on the bottom of S2's generated trial (Fig. 5.9(c)) suggests that the generated signal produces a positive bias, while the blank on the top of S2(d) (Fig. 5.9(d)) indicates a negative bias based on 0 μV. In S3(Fig. 5.10), the impact of high-frequency noise at the beginning and end of the generated trial is lower than that in S1 and S2. However, the generated signal in S3 (Fig. 5.10(d)) also exhibits a positive bias. These biases stem from the inadequacies in the DCGAN's structure design. Specifically, the initial layer of the proposed generator of DCGAN serves to project the two-dimensional EEG data onto a different space, while the final layer of the discriminator in the DCGAN functions to project it back. The intervening layers perform non-linear transformations that can induce biases at the final projection back process.

Time-frequency domain observation

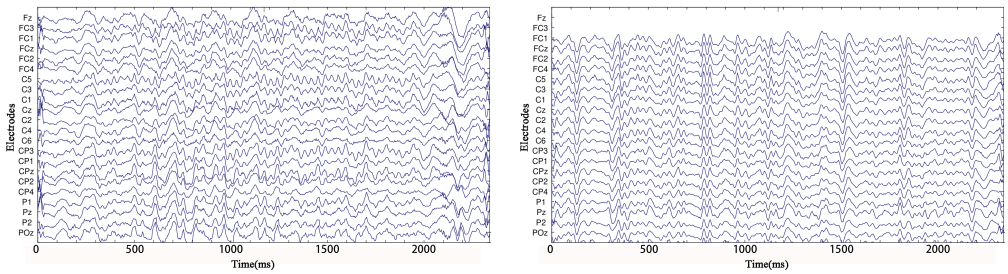
In addition to viewing the signal in time series, the signal was also analyzed in time frequency for more comparison.

As mentioned earlier, the cropped trial strategy involves taking a 600-sample



(a) a real left hand MI trial

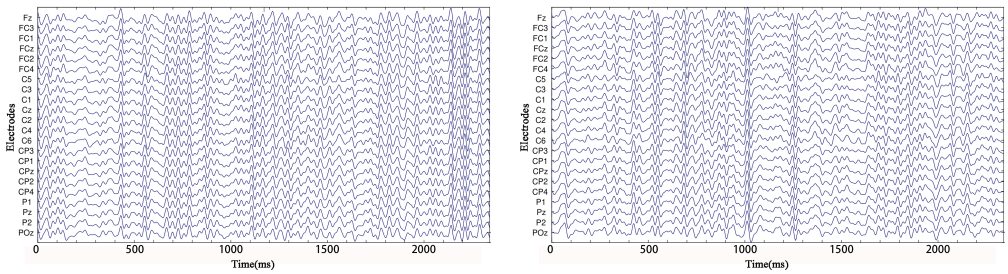
(b) a real right hand MI trial



(c) an artificial left hand MI trial

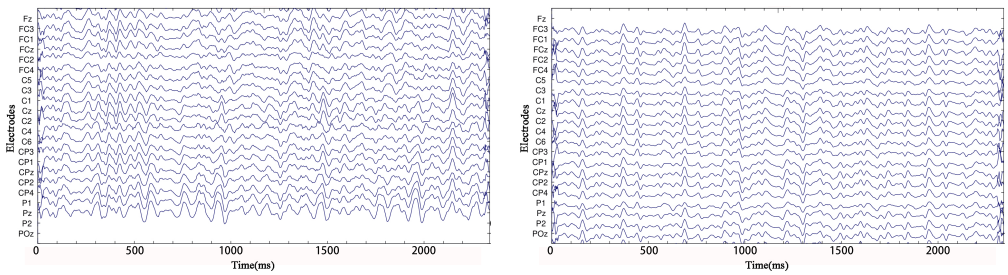
(d) an artificial right hand MI trial

Figure 5.8: Comparison of artificial trials and the real trials of S1



(a) a real left hand MI trial

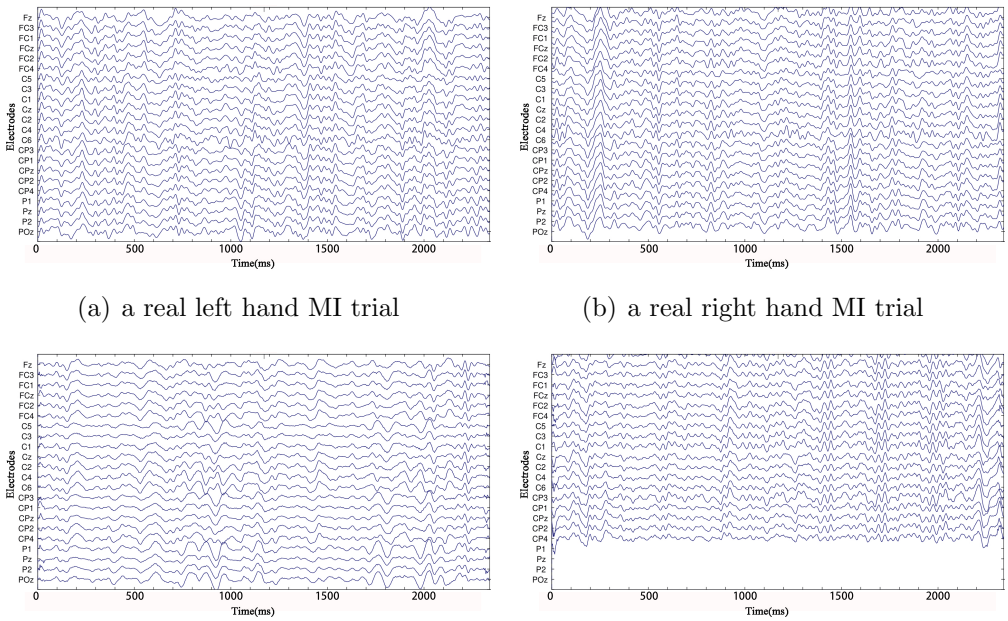
(b) a real right hand MI trial



(c) an artificial left hand MI trial

(d) an artificial right hand MI trial

Figure 5.9: Comparison of artificial trials and the real trials of S2



(a) a real left hand MI trial

(b) a real right hand MI trial

(c) an artificial left hand MI trial

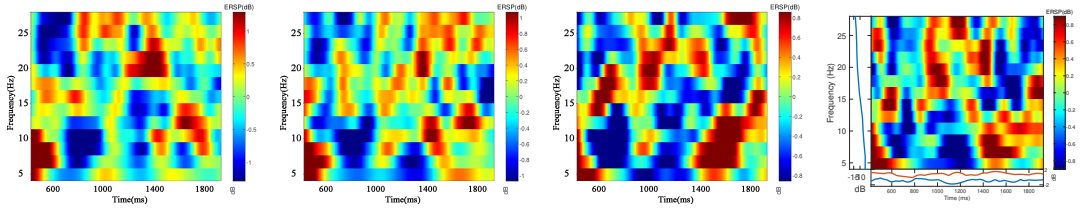
(d) an artificial left hand MI trial

Figure 5.10: Comparison of artificial trials and the real trials of S3

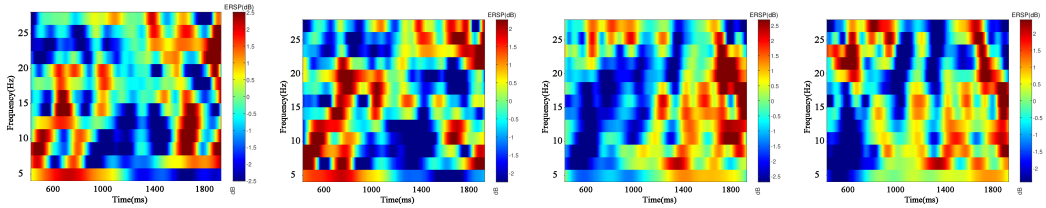
window from the original trial of 1000 sample points as a cropped trial. In the time-frequency analysis, the first cropped trial in every 72 original trials is collected for time-frequency analysis to avoid time shifting causing energy cancellation in the spectrogram, which would impede the observation of ERD and ERS phenomena. For example, on the time-frequency map of the collection of the first cropped trial, a distinct ERD phenomenon emerges around 8-10 Hz at 500 ms on electrodes C3 and C4 for left (Fig. 5.11(a), 5.11(b)) and right-hand (Fig. 5.11(c), 5.11(d)) MI in S1. Since the window length is 66 sample points, the second cropped trial extracted from every original trial will show an earlier appearance of the ERD by 264 ms (66/250). This pattern continues for the subsequent cropped trials. As expected, due to the cropping strategy, the spectrogram of the generated fake data (Fig. 5.11(e) - 5.11(h)) lacks the clear ERD phenomenon around 500 ms and 6-10 Hz that can be observed in the real data. Instead, it displays irregular spectral energy changes. Fortunately, even without precise timing, the ERD/ERS phenomenon can still be observed around 10 Hz in these fake trials. On the other hand, the more pronounced ERD/ERS contrast for real left-hand MI should appear in the contralateral electrode, namely C4, whereas for real right-hand MI, it should occur in the C3 electrode. In S1's real left-hand MI, there is indeed a more noticeable ERD/ERS phenomenon in C4 (Fig. 5.11(b)) than in C3 (Fig. 5.11(a)) around 6-12 Hz. C3 (Fig. 5.11(c)) exhibits more significant changes than C4 (Fig. 5.11(d)) in real right-hand MI.

This difference is a crucial factor affecting classification accuracy. However, it becomes difficult to distinguish between the ERD/ERS intensities in fake trial C3 (Fig. 5.11(e), 5.11(g)) and C4 (Fig. 5.11(f), 5.11(h)) of S1. For S2 (Fig. 5.9), real cropped trials (Fig. 5.12(a), 5.12(d)) of S2 display minimal differences in patterns between C3 and C4 for left and right-hand MI. This explains why the classification accuracy of S2 is only around the chance level. The generated fake trials (Fig. 5.12(e), 5.12(h)) do not resemble the real trials' spectrograms and do not enhance the distinction between the two MI classes across the left and right electrodes. Lastly, When performing left and right-hand motor imagery, S3's real trials (Fig. 5.13(a), 5.13(d)) demonstrate ERD/ERS in C3 and C4 near 10 Hz, with a stronger phenomenon observed in the contralateral electrode, adhering to the principle. Moreover, an intriguing phenomenon is observed, indicating significant differences between the spectrograms of left and right-hand motor imagery represented by Fig. 5.13(a), 5.13(b) and Fig. 5.13(c), 5.13(d), respectively. For example, at around 1400 ms, ERD appears in the left-hand trial for both C3 and C4, while the opposite ERS is observed in the right-hand trial at the same time point. However, for S1 and S2, no such inter-class differences are evident. For S3, not only does the ERD/ERS intensity difference on two electrodes during two tasks contribute the classification accuracy, but the inter-class differences are also highly beneficial for classification. However, in the generated trials for right-hand MI (Fig. 5.13(g), 5.13(h)), the ERD/ERS intensities are stronger in C3 than in C4, which is obviously incorrect and inconsistent with the real trial. Fortunately, the fake trials capture the inter-class difference between left and right-hand motor imagery at around 1400ms, which is consistent with the real trials.

It can be concluded that, in terms of the time-frequency domain, the generated trials resemble the real trials for ERD/ERS phenomenon. However, the false data of the right-hand MI appeared contradictory ERD/ERS comparisons on the C3 and C4 electrodes of S3. This indicates that the GAN can learn the frequency information of the Electroencephalogram (EEG), but ERD/ERS on electrodes are not consistent with the real situation. This could be attributed to the strategy used to generate artificial trials, whereby the fake trials are produced without labels. Our approach is to train each class separately. Although the DCGAN can capture the ERD/ERS information of both classes, it lacks the ability to differentiate between the two classes (i.e., generate trials with labels). Notably, the inter-class difference of S3 benefits the calculation of the CSP feature. That is the reason why S3 achieves benefit from the DCGAN based on shallowFBCSPNet

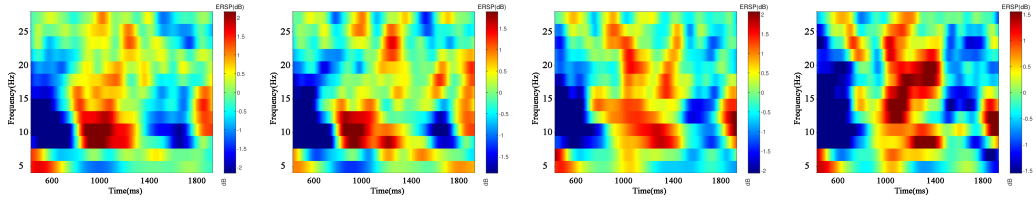


(a) C3:a real left hand MI trial (b) C4:a real left hand MI trial (c) C3:a real right hand MI trial (d) C4:a real right hand MI trial

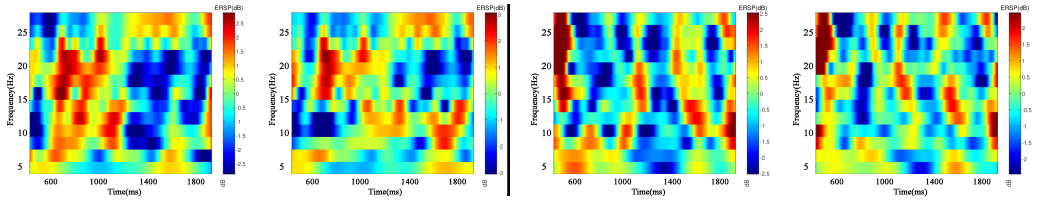


(e) C3:an artificial left hand MI trial (f) C4:an artificial left hand MI trial (g) C3:an artificial right hand MI trial (h) C4:an artificial right hand MI trial

Figure 5.11: Comparison of artificial trials and the real trials with channel time-frequency map for S1

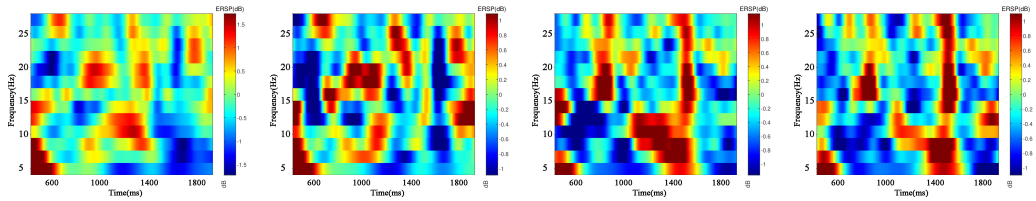


(a) C3:a real left hand MI trial (b) C4:a real left hand MI trial (c) C3:a real right hand MI trial (d) C4:a real right hand MI trial

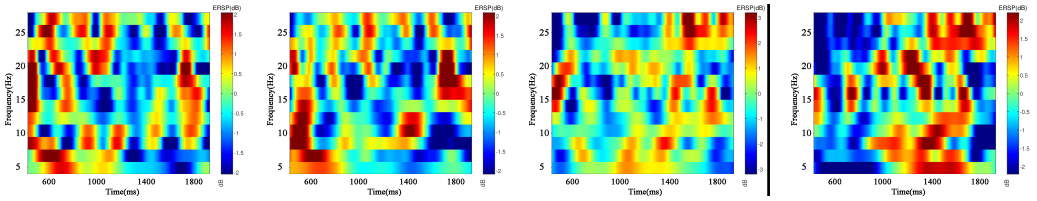


(e) C3:an artificial left hand MI trial (f) C4:an artificial left hand MI trial (g) C3:an artificial right hand MI trial (h) C4:an artificial right hand MI trial

Figure 5.12: Comparison of artificial trials and the real trials with channel time-frequency map for S2



(a) C3:a real left hand MI trial (b) C4:a real left hand MI trial (c) C3:a real right hand MI trial (d) C4:a real right hand MI trial



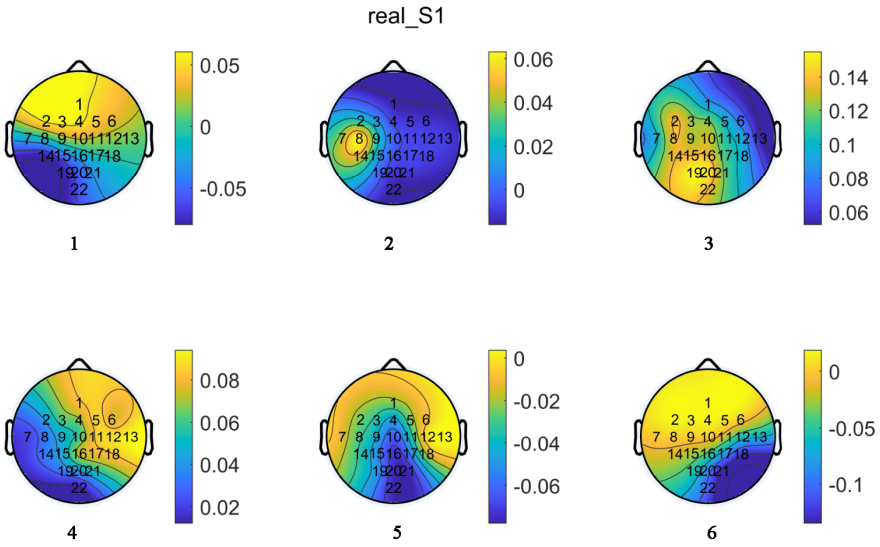
(e) C3:an artificial left hand MI trial (f) C4:an artificial left hand MI trial (g) C3:an artificial right hand MI trial (h) C4:an artificial right hand MI trial

Figure 5.13: Comparison of artificial trials and the real trials with channel time-frequency map for S3

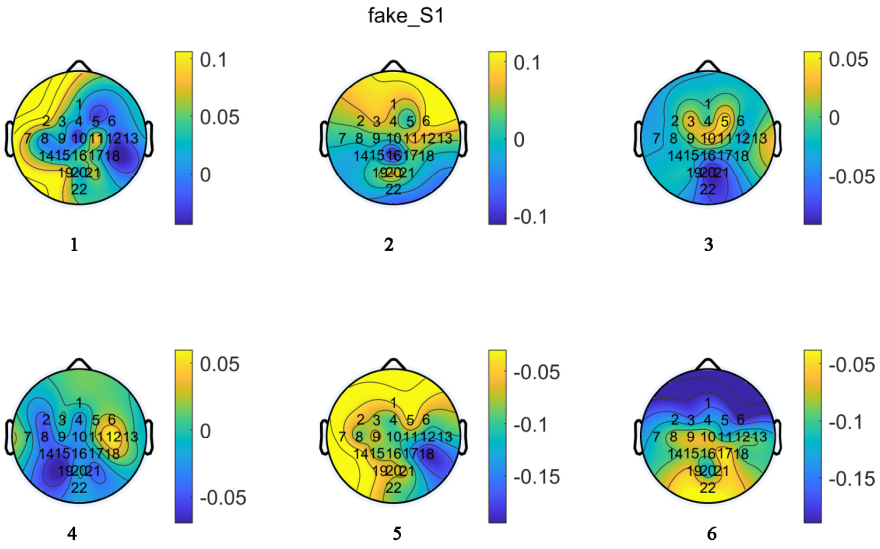
Spatial filters observation

The proposed DCGAN is based on the Common Spatial Pattern (CSP) spatial pattern and incorporates a CSP-related CNN classifier. This section compares the difference in 6 CSP filters between the real and generated trials. No similar distribution is apparent between the CSP filters calculated from the generated and real trials on the topography for S1 (Fig. 5.14) and S2 (Fig. 5.15). However, similarities can be observed in the distribution of CSP filters between the generated and real trials on S3 (Fig. 5.16).

The first filter displays a small value in the eigenvector obtained by decomposing the sum of the covariance matrices of two classes of MI data with the blue color on the bottom left section. The second filter indicates a larger value in the eigenvector in yellow on the bottom section. These two distributions are consistent in the real and fake CSP filters. In addition, the large (yellow) value on the right side of the third filter and on the right bottom of the sixth filter also indicates a similar distribution between the real and fake CSP filters. This observation supports the notion that S3 benefits from the GAN, as the generated trials display a distribution of CSP filters that is similar to that of the real trials. This is consistent with the conclusion from the time-frequency observation that the DCGAN has learned the inter-class variation of two classes of MI.

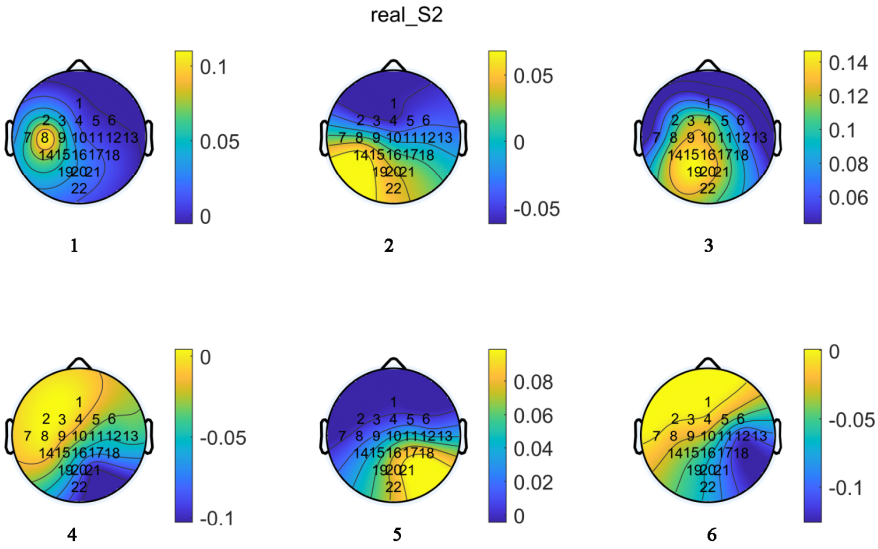


(a) CSP filters calculated from real trials for S1

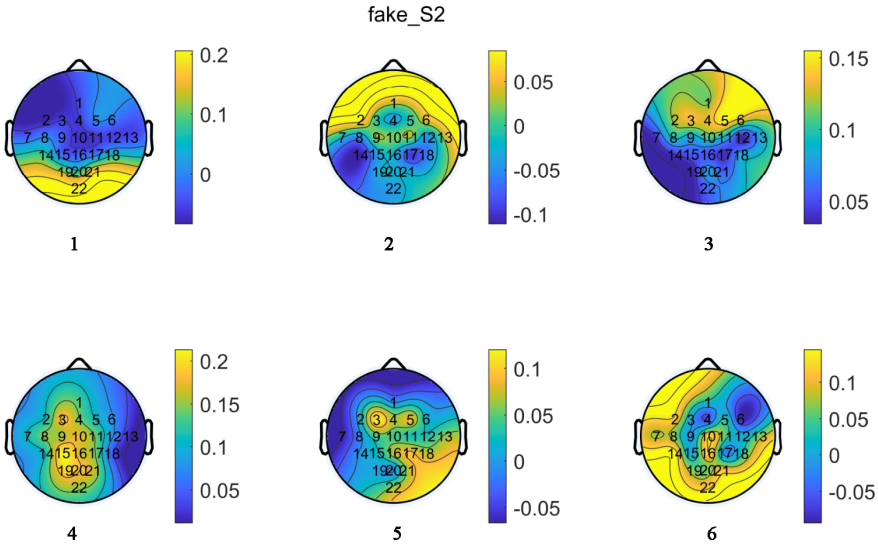


(b) CSP filters calculated from artificial trials

Figure 5.14: Comparison of artificial trials and the real trials with CSP filters

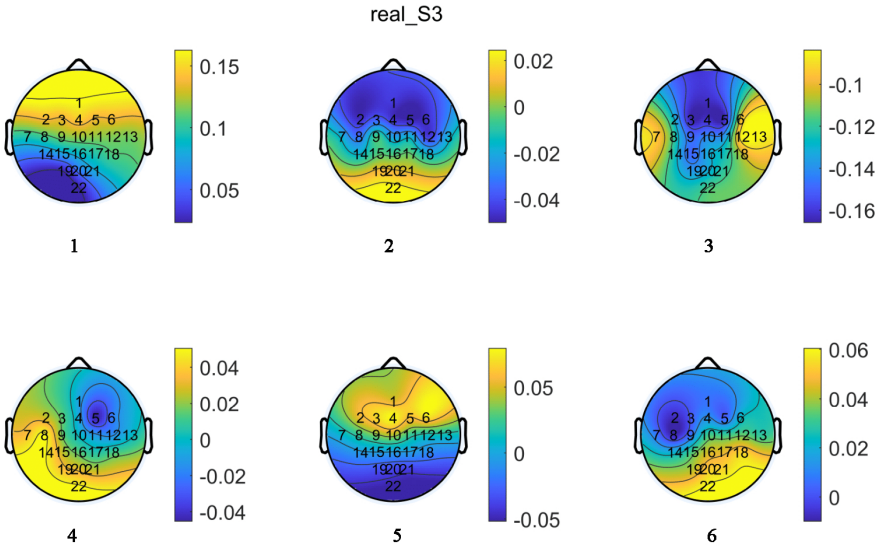


(a) CSP filters calculated from real trials

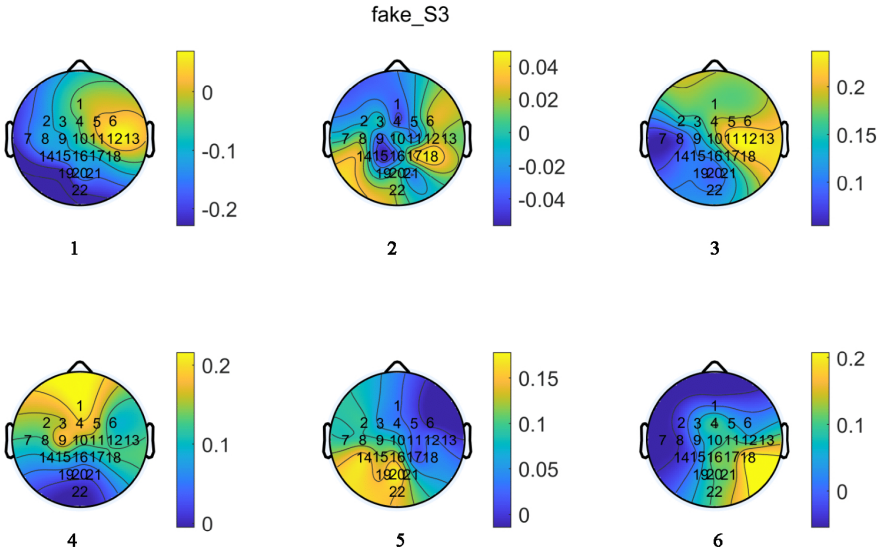


(b) CSP filters calculated from artificial trials

Figure 5.15: Comparison of artificial trials and the real trials with CSP filters for S2



(a) CSP filters calculated from real trials



(b) CSP filters calculated from artificial trials

Figure 5.16: Comparison of artificial trials and the real trials with CSP filters for S3

5.4 Discussion and Conclusions

In this part, we combined the cropped training and TL to generate artificial trials in the cropped trial scale. This method compensates for the lack of training trials and makes good use of existing data from subjects from the same session. Another similar study also showed the importance of the number of training trials. Sujit et al. performed a MI EEG GAN to generate artificial trials for MI data. In this study, they took the strategy which was the same as the cropped training. They extracted 2s-long EEG data from each 4s-long trial of only channel C3 in BCI competition IV dataset 2b as a training sample for GAN and used a time window of 50 samples to slide along the 4s-long trial. Thus, yielding more trials to feed into the GAN. They didn't apply transfer learning thus up to 420 original trials were used for their study. They evaluated the quality of artificial trials using STFT image and Power Spectrum Density (PSD) but no classification results related to artificial trials [185].

However, state-of-the-art MI classification accuracy for BCI competition IV dataset 2a is 91.57% [186], conducted by Dai et al. where they combined a hybrid scale CNN architecture with data augmentation. Others can reach 68.32% [187], 78.01% [188],And 79.93% [189] respectively for the same dataset.Our classification accuracy with smaller training trials (cropped trials from 40 original trials for each class) and artificial trials appended is only 70.70%, This is still very low when compared the state of the art method. Even though our study shows no significant improvement, no similar study which combined TL, cropped training, and GAN in MI classification was found in the published literature. However, the evaluation of artificial data implies the feasibility of the proposed GAN in subjects with different patterns of ERD/ERS in left and right hand MI. When compared to a similar study where they also combined the TL with DCGAN but did not use cropped training, they used up to hundreds of trials to achieve 93.2% classification accuracy for BCI competition IV dataset 2b [169]. Another advantage of this method is that it doesn't need the representation of the raw EEG signals, e.g., transferring to a 2D image or pre-extracting the CSP feature [169]. Still, it enables DNN to directly learn from the time series EEG. EEG patterns are not easily identified like the image or audio features in DL, so EEG-based CNNs are specially designed to classify different EEG features. This study also provides insight view into how to apply the existing CNN to a GAN. Although the average testing classification accuracy after artificial trials appended shows no significant increase. The dramatic increase for the specific subject 3 in our study proves that the artificial trial generation based on the combination of TL and DCGAN is a

promising method for reducing the calibration time of MI EEG. Future research can focus on reducing the ineffective trials in the source data when applying TL learning. Another direction could be to apply different weights to different states to different datasets in the source dataset, increasing the proportion of trials that are more similar to the target dataset, although this method might need to establish the criteria to judge the similarity. On the other hand, improvements can be made to DCGAN. The new structure of the proposed can be considered to avoid bias thus improving the generation ability of the DCGAN. Other CNNs designed for MI classification are also worthy of reconstruction into a GAN, consisting of sequential placement and inverting placement of the layers of the CNN. The best solution should combine the improvement of both the TL and the GAN. In the study of Tang et al., they also realized the problem of a shortage of training MI trials. Their solution was to design a Conditional Domain Adaption Neural Network (CDAN) to decode the MI features and then perform classification. After this CDAN learn sufficient intra-subject features of MI, it was used to classify the target person's MI data [190]. The condition means labeling the EEG data with left/right throughout the training procedure. This allows for producing data specific to the desired task.

Chapter 6

Discussion and Future work

6.1 Discussion and limitation

Hybrid BCI typically combines different types of BCI, providing robustness and increased BCI degrees of freedom. A BCI based on VEP is often combined with MI BCI. That was the main initial motivation for focusing on mVEP and MI BCI in this thesis.

The mVEP BCI, inducing little visual fatigue compared to SSVEP, has a good information transfer rate requiring a small number of repetitions to achieve good accuracy and does not require training. Therefore it is a VEP of choice for increasing the degrees of freedom of MI BCI.

Previous research from the literature indicated that VEPs, in general, are highly related to graphical factors in the Graphical User Interface, for example, colors. For P300, using stimuli of different colors can affect its response to tasks. For example, associating stimuli of a relevant color with specific options during decision-making using ERP can improve response accuracy. Stimuli of different colors can also produce practical effects in BCI applications. In this thesis, the influence of graphic factors of mVEP is explored.

Results of the study presented in Chapter 3 indicate that the color of the moving bar and the background color does not significantly influence the classification accuracy of mVEP BCI. Furthermore, geometrical factors, such as the size of the box and its layout on the screen, do not significantly influence the classification accuracy either. Results are in agreement with previous results from the literature, which showed that dynamics of the background on which boxes are superimposed do not affect mVEP BCI performance. This is in contrast to VEP BCI based on P300, which is highly dependent on the GUI design, probably because P300 is related to higher cognitive processing than mVEP. The results are important because they indicate good flexibility of mVEP that can be used

in different BCI scenarios and could therefore be combined with the GUI layout required for other types of BCIs such as MI VEP. Thus mVEP is a good candidate for complex BCI games that include a hybrid design.

A limitation of the study is the relatively small number of graphical factors tested; however, increasing the number of parameters would increase the duration of the experiment beyond the level comfortable to participants. On the other hand, in order to make a fair comparison, it was not possible to test different geometrical factors over different experimental sessions. The other limitation is a relatively small number of participants, which limited statistical methods and might have partially contributed to the lack of statistical significance between conditions. Future studies should test the influence of GUI on mVEP on a larger number of participants and should also include background dynamics as one of the GUI factors.

In contrast to mVEP BCI, MI BCI requires user training, also demanding lengthy offline data collection that can often be tiring for the user. Therefore the second and third studies of this thesis focused on data augmentation methods in order to reduce offline training by increasing the size of the available dataset.

Chapter 4 was designed for smaller datasets when only a limited number of training data for an individual participant were available. It is therefore based on features and classifiers traditionally used in BCI. Chapter 5 explores advanced data augmentation and classification methods which typically require a larger number of dataset. For that reason, transfer learning was introduced in order to increase the size of the original EEG dataset.

The novelty of chapter 3 is that it explores several different EEG-based features as candidates for DA, going beyond CSP, which was widely used in previous DA studies. It further combines different EEG features with different DA methods to explore whether there is an optimal combination of features (time, frequency, spatial) and DA methods (time, time-frequency, and domain transformation, to some extent similar to CSP).In summary, while simpler DA methods achieve lower classification accuracies, their results seem to be more robust and replicable on different datasets.

Amongst all three types of features, and for a larger number of original datasets (more than 40 trials), CSP achieved the best classification accuracy, which would be expected based on results from the literature. However, adding DA did not improve BCI classification performance, irrespective of the DA method. On the other hand, time-domain features achieved a lower initial classification accuracy, but a small number of original trials (10-20) resulted in a significant improvement in classification accuracy when combined with any of

the three DA methods. This data augmentation seems to be more dependent on the type of EEG features than on the DA method, at least for three selected DA methods. Importantly, the consistency of this finding was confirmed on two different datasets. An important advantage of time-domain parameter features over CSP is the smaller number of required EEG electrodes. Consumer-grade EEG devices have a typically smaller number of electrodes, and BCI gaming is an area appealing to able-bodied users that would typically choose affordable devices. A limitation of this study is that it does not include the novelty of DA method.

The DCGAN-based DA method is one of the most promising DA methods since Deep Neural Network(DNN) provides an opportunity to unearth hidden EEG features. This study brings a few novel elements. In chatpter 5, we use CNN to design a DCGAN for MI data. Although the average classification shows no statistically significant improvement on the group level, classification performance has improved for some individual participants. The number of trials used in conventional BCI classifiers is up to a few hundred, while GAN requires thousands of trials. Based on this background, the novelty of this study is a method to increase the number of trials for GAN by combining TL and cropped trails. This, however, did not significantly improve classification accuracy. Again, classification accuracy improved in several participants. A general limitation of CNN classifiers is their complexity with lots of potentially adjustable parameters. In this study, a fixed CNN architecture was used based on published results from the literature. While most of the DA studies based on ANN use standard competition datasets in order to compare the performance between methods, it is questionable whether similar accuracies could be achieved with another dataset due to fine-tuning of ANN parameters to the competition dataset.CNN-based DA methods have greater versatility and could potentially provide better classification accuracy but require adjustment of a large number of parameters and might be prone to overfitting.

6.2 Future Work

One general limitation of this thesis is the relatively small number of participants in each dataset due to the COVID-19 pandemic. For chapter 3, we utilized a validated competition dataset and were able to compare our results with those from the existing literature. Notably, the results from the smaller visual angle in Study 3 displayed relatively higher average accuracy than those from the larger visual angle, which contradicts previous literature. Given that our experiment

only involved ten subjects, future research should include a larger sample size to verify this result. On the other hand, more laboratory dataset, rather than a competition dataset can be used. Furthermore, while our study showed that static visual factors do not have a significant impact, in practical applications, games are often dynamic, with moving objects, flashing indicators, and sudden scene changes, which may have an effect. Therefore, future work could incorporate these dynamic factors to investigate their impact on mVEP BCI. Regarding our study on the location of boxes, we found no significant differences, but in a game setting, some regions may contain more game-related content, requiring more frequent attention from the user, while others may not. Thus, future work could place these boxes in a real game environment to investigate whether the location is related to the richness of game content.

In chapter 4, the TDP feature extraction method did not show any advantages in terms of the original accuracy compared to the other two methods using fewer electrodes.Thus, there are relatively fewer studies on TDP compared to CSP. However, the shallowFBCSP Net constructed based on CSP in Study 5 could perform the classification task of MI. Therefore, future research can perform on constructing a deep learning classifier based on the TDP method to maximize the potential of TDP features. Furthermore, the data augmentation method based on SA DA, which is related to spatial transformation, reminded us of the similarity to ICA analysis commonly used in EEG analysis. Therefore, future work could initially conduct ICA analysis on a certain number of single-class MI EEGs, mark the most important principal components related to EEG, swap them with the same number of principal components from other EEGs, and then transform them back to generate new artificial MI EEGs.

Regarding chapter 5 , where the DCGAN proposed in this study has limited applicability for nine subjects,. However, by evaluating the fake data generated by DCGAN of S1-S3, we discovered the uniqueness of S3 in MI tasks, where the left and right hand tasks exhibit completely different ERS/ERD patterns. For future MI BCI research, when we observe such subjects, can we ask about their experience and consider whether this experience can be learned by other subjects through training? Returning to the DCGAN itself, there are still many shortcomings that need to be improved. Although the DCGAN is based on learning from temporal EEG data, we found that it captured the ERD/ERS phenomenon in frequency domain analysis. However, DCGAN cannot generate labeled data, and to separate the two types of MI data, we trained them separately. This resulted in DCGAN learning ERD/ERS but having no chance to learn the distinction between left and right hand tasks. This issue can be

resolved by using a conditional GAN (CGAN), as it can train labeled data. To address the bias observed in artificial trails, the structure of the DCGAN could be improved by placing the projection layer in the middle layer. Wasserstein GAN can be tried since it is more stable than DCGAN.

The results of all three studies together should inform the design of a future hybrid BCI system based on mVEP and MI. This system would require only a few trials for calibration, allowing users to quickly and easily use it at home. The system would provide up to six degrees of freedom, making it suitable for a range of applications beyond gaming. For example, it could be integrated into assistive devices such as wheelchairs, and the use of mVEP could reduce visual fatigue. Overall, the statement suggests that the combination of mVEP and MI has the potential to create a highly efficient and user-friendly BCI system with a range of potential applications. By reducing the calibration time and increasing the degrees of freedom, this system could be used in a variety of scenarios beyond gaming, making it an exciting prospect for the future of BCI technology.

Appendix A

Dataset1: Average testing accuracy across 9 subjects. The testing trails are always 72 trails for all circumstance and different form the original trails.

Table A.1: CSP with Dataset 1

Segmentation and recombination in time domain				
	10 OT	**20 OT**	**40 OT**	**60 OT**
0 AT	63.73 ±17.70	70.68 ±15.95	72.53 ±16.83	73.15 ±17.29
10 AT	65.96 ±16.03	68.96 ±15.98	72.92 ±16.18	73.44 ±16.00
20 AT	66.45 ±16.16	69.17 ±16.65	72.49 ±16.21	73.32 ±16.11
40 AT	66.48 ±16.27	69.91 ±16.49	72.68 ±16.62	73.47 ±16.54
60 AT	67.84 ±16.65	70.25 ±16.67	72.79 ±16.10	73.35 ±16.02
Segmentation and recombination in time-frequency domain				
0 AT	63.73 ±17.70	70.68 ±15.95	72.53 ±16.83	73.15 ±17.29
10 AT	66.11 ±17.09	70.01 ±16.37	73.01 ±15.94	73.61 ±16.41
20 AT	67.05 ±16.49	69.96 ±16.47	72.48 ±16.60	73.25 ±16.28
40 AT	67.91 ±16.55	69.68 ±16.91	72.75 ±16.34	73.71 ±16.14
60 AT	68.43 ±16.97	70.58 ±16.71	73.15 ±15.88	73.19 ±15.91
Spatial analogy				
0 AT	63.73 ±17.70	70.68 ±15.95	72.53 ±16.83	73.15 ±17.29
10 AT	66.06 ±15.89	69.42 ±15.63	72.24 ±15.95	73.10 ±16.17
20 AT	65.42 ±15.31	69.19 ±16.21	71.68 ±16.29	72.72 ±16.38
40 AT	65.17 ±15.22	68.43 ±16.47	72.16 ±16.89	73.16 ±15.81
60 AT	65.17 ±16.33	68.57 ±16.39	72.11 ±16.61	73.15 ±15.78

Dataset2: Average testing accuracy across 9 subjects. The testing trails are always 50 for all circumstance and different form the original trails.

Table A.2: BP with Dataset 1

Segmentation and recombination in time domain				
	10 OT	**20 OT**	**40 OT**	**60 OT**
0 AT	68.02 ±13.38	68.37 ±12.31	69.99± 12.46	72.13 ±13.00
10 AT	67.82 ±13.39	67.92 ±12.95	70.53 ±12.98	72.29 ±13.07
20 AT	68.33 ±13.46	68.22 ±13.17	70.88 ±12.95	72.10 ±13.22
40 AT	68.73 ±14.05	68.14 ±13.59	70.94 ±13.04	72.35 ±13.47
60 AT	68.19 ±13.79	68.97± 13.52	71.13 ±12.89	72.17 ±13.52
Segmentation and recombination in time-frequency domain				
0 AT	68.02 ±13.38	68.37± 12.31	69.99 ±12.46	72.13 ±13.00
10 AT	68.43 ±13.76	68.33 ±12.89	70.87 ±13.04	72.56 ±13.21
20 AT	68.04 ±13.80	68.49 ±13.37	70.87± 13.19	72.71 ±12.81
40 AT	68.65 ±13.96	69.16 ±13.34	71.05 ±13.04	72.47 ±13.01
60 AT	68.96 ±13.85	69.06 ±13.51	70.93 ±13.23	72.69± 12.89
Spatial analogy				
0 AT	68.02 ±13.38	68.37 ±12.31	69.99 ±12.46	72.13 ±13.00
10 AT	67.95 ±13.17	68.88 ±12.42	70.42 ±12.86	72.21 ±12.98
20 AT	68.28 ±13.39	68.58 ±12.56	70.46 ±12.82	72.19 ±13.04
40 AT	68.15 ±13.94	68.65 ±12.49	70.78 ±12.89	71.91 ±12.86
60 AT	67.85 ±14.39	68.57± 12.76	70.80 ±12.88	71.87± 13.19

Table A.3: TDP with Dataset 1

Segmentation and recombination in time domain				
	10 OT	**20 OT**	**40 OT**	**60 OT**
0 AT	54.85±2.94	55.81±2.35	61.43±7.40	62.15±6.82
10 AT	56.23±2.24	58.35±4.44	60.65±7.00	63.55±8.89
20 AT	56.99±3.37	59.33±4.96	61.43±7.60	64.00±9.17
40 AT	57.12±3.85	60.15±5.66	61.92±7.91	64.69±9.46
60 AT	58.28±3.89	60.06±5.54	62.29±9.02	64.82±10.10
Segmentation and recombination in time-frequency domain				
0 AT	54.85±2.94	55.81±2.35	61.43±7.40	62.15±6.82
10 AT	56.38±3.03	58.90±4.88	60.66±7.36	63.57±8.22
20 AT	57.34±3.37	59.76±5.90	61.80±7.90	64.57±9.48
40 AT	59.16±5.52	61.67±7.84	63.25±9.19	65.34±10.03
60 AT	61.08±6.21	63.81±9.82	64.10±10.90	66.12±9.88
Spatial analogy				
0 AT	54.85±2.94	55.81±2.35	61.43±7.40	62.15±6.82
10 AT	55.50±2.02	58.13±3.27	61.04±6.73	63.34±8.17
20 AT	57.71±4.88	59.48±5.70	61.81±7.32	64.09±8.12
40 AT	58.46±6.03	61.65±6.53	63.32±8.26	64.75±8.50
60 AT	59.74±6.05	62.91±7.69	63.59±9.03	65.71±8.91

Table A.4: CSP with Dataset 2

Segmentation and recombination in time domain				
	10 OT	**20 OT**	**40 OT**	**50 OT**
0 AT	53.50 ±9.72	55.08 ±9.41	61.25 ±12.92	63.00 ±15.20
10 AT	52.67 ±6.77	55.62 ±9.10	61.30 ±13.18	62.70 ±13.58
20 AT	53.67 ±6.63	55.47 ±8.81	60.97 ±13.32	63.62 ±13.42
40 AT	53.77 ±5.42	55.41± 9.22	61.43 ±12.75	62.43 ±13.80
50 AT	55.03 ±5.97	55.94 ±9.08	61.87 ±12.45	63.28 ±13.54
Segmentation and recombination in time-frequency domain				
0 AT	53.5 ±9.72	55.08 ±9.41	61.25 ±12.92	63.00 ±15.20
10 AT	53.61 ±6.22	55.61 ±8.86	61.78 ±12.81	62.93 ±14.07
20 AT	54.01 ±5.96	54.76 ±8.45	62.19 ±12.67	62.43 ±14.51
40 AT	54.23 ±5.76	55.52 ±9.01	62.18 ±12.54	62.82 ±13.88
50 AT	54.47 ±5.36	56.0 ±8.41	61.88 ±12.81	62.24 ±14.21
Spatial analogy				
0 AT	53.50 ±9.72	55.08 ±9.41	61.25 ±12.92	63.00 ±15.20
10 AT	53.01 ±7.29	55.60 ±8.35	60.40 ±13.41	62.51 ±13.49
20 AT	52.66 ±6.09	54.72 ±9.17	60.46 ±13.21	60.90 ±13.67
40 AT	53.73 ±6.65	55.16 ±8.64	60.12 ±13.44	62.33 ±13.08
50 AT	53.93 ±6.16	54.50 ±9.52	59.89 ±13.16	62.50 ±12.91

Table A.5: BP with Dataset 2

Segmentation and recombination in time domain				
	10 OT	**20 OT**	**40 OT**	**50 OT**
0 AT	56.64 ±5.66	57.93 ±7.04	59.97 ±10.13	59.55 ±11.00
10 AT	57.74 ±4.63	58.38 ±4.72	59.05 ±9.37	59.69 ±10.22
20 AT	57.04 ±4.55	57.48 ±5.10	59.13 ±8.00	60.39 ±9.20
40 AT	56.43 ±5.80	57.32 ±5.76	59.51 ±7.81	59.94 ±9.19
50 AT	56.52 ±5.31	57.29 ±6.02	59.78 ±7.64	60.36 ±9.17
Segmentation and recombination in time-frequency domain				
0 AT	56.64 ±5.66	57.93 ±7.04	59.97 ±10.13	59.55 ±11.00
10 AT	57.74 ±4.82	58.25 ±5.39	59.66 ±8.83	59.42 ±10.19
20 AT	57.18 ±4.31	57.90 ±4.41	59.84 ±7.97	59.88 ±9.58
40 AT	56.99 ±4.47	57.34 ±5.39	59.74 ±6.82	59.66 ±9.30
50 AT	56.61 ±4.09	57.40 ±4.78	59.70 ±6.62	59.79 ±9.08
Spatial analogy				
0 AT	56.64 ±5.66	57.93 ±7.04	59.97 ±10.13	59.55 ±11.00
10 AT	58.51 ±5.23	55.65 ±5.22	58.10 ±8.67	57.69 ±10.04
20 AT	57.69 ±5.21	56.84 ±5.14	57.65 ±8.20	57.06 ±9.60
40 AT	57.92 ±6.11	56.76 ±5.16	57.33 ±7.59	57.09 ±10.00
50 AT	57.72 ±5.92	57.45 ±6.20	57.29 ±8.17	57.67 ±10.06

Table A.6: TDP with Dataset 2

Segmentation and recombination in time domain				
	10 OT	**20 OT**	**40 OT**	**50 OT**
0 AT	52.96 ±1.64	54.61 ±3.82	59.01 ±7.13	58.23 ±3.72
10 AT	55.26 ±1.60	55.30 ±2.08	58.22 ±2.72	58.66 ±3.10
20 AT	55.37 ±1.05	55.94 ±1.62	57.84 ±2.70	58.04 ±2.80
40 AT	54.97 ±1.53	55.59 ±1.96	57.29 ±3.40	58.23 ±3.70
50 AT	55.00 ±1.66	55.97 ±1.51	57.66 ±3.28	57.62 ±3.20
Segmentation and recombination in time-frequency domain				
0 AT	52.96 ±1.64	54.61 ±3.82	59.01 ±7.13	58.23 ±3.72
10 AT	54.59 ±0.98	55.24 ±3.03	57.96 ±3.06	58.31 ±3.02
20 AT	54.38 ±1.67	55.69 ±3.00	58.15 ±3.56	57.34 ±3.03
40 AT	55.38 ±2.04	56.93 ±2.50	57.63 ±4.13	57.72 ±3.79
50 AT	55.50 ±1.81	56.48 ±3.57	58.53 ±4.42	57.65 ±3.64
Spatial analogy				
0 AT	52.96 ±1.64	54.61 ±3.82	59.01 ±7.13	58.23 ±3.72
10 AT	53.96 ±1.66	55.65 ±1.32	58.10 ±2.86	57.69 ±2.59
20 AT	54.61 ±1.26	56.84 ±2.40	57.65 ±4.03	57.06 ±3.14
40 AT	55.89 ±2.01	56.76 ±2.24	57.33 ±4.42	57.09 ±4.34
50 AT	56.66 ±4.58	57.45 ±3.26	57.29 ±4.70	57.67 ±4.58

Appendix B

Table B.1: KW Test for TDP Feature in dataset 1

SRTD				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.4529	0.2004	0.2697	0.1711
20 OT	0.3538	0.2004	0.1023	0.1711
40 OT	0.9648	0.5078	0.6911	0.6911
60 OT	0.9648	0.9648	0.7573	0.7573
SRTFD				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.5078	0.1451	0.1023	0.0469
20 OT	0.3538	0.3538	0.1023	0.1023
40 OT	0.8253	0.7573	0.6272	0.6911
60 OT	0.8946	0.9648	0.9648	0.3538
SA				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.7573	0.4015	0.3099	0.1223
20 OT	0.2697	0.2697	0.0703	0.0193
40 OT	0.9648	0.6272	0.5078	0.5078
60 OT	0.8946	0.6272	0.5078	0.3099

Table B.2: KW Test for CSP Feature in dataset 1

SRTD				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.5658	0.5658	0.6270	0.5658
20 OT	0.8946	0.8946	0.9648	0.9648
40 OT	0.8946	0.8946	0.9648	0.8946
60 OT	0.9648	0.8946	0.9648	0.8946
SRTFD				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.8946	0.6910	0.3096	0.5076
20 OT	0.9648	0.8946	0.8946	0.9648
40 OT	0.9648	0.8946	0.9648	0.8946
60 OT	0.9648	0.7910	0.8946	0.8946
SA				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.7572	0.7572	0.8252	0.7237
20 OT	0.8946	0.8946	0.8253	0.8253
40 OT	0.9648	0.8946	0.9648	1.0000
60 OT	0.9648	0.8946	0.8946	0.8946

Table B.3: KW Test for BP Feature in dataset 1

SRTD				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.8946	0.8946	0.7573	0.7573
20 OT	0.9648	0.8946	0.9648	0.9648
40 OT	0.8253	0.6911	0.8253	0.6911
60 OT	0.7573	0.9648	0.7573	0.8253
SRTFD				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.8253	0.8253	0.8253	0.6911
20 OT	0.9648	0.9648	0.8253	0.8253
40 OT	0.8253	0.8253	0.8946	0.8946
60 OT	0.6911	0.6272	0.7573	0.7573
SA				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.8253	0.8946	0.9648	0.8946
20 OT	0.8253	0.8946	0.8253	0.8946
40 OT	0.8946	0.8253	0.8253	0.8253
60 OT	0.8253	0.8253	0.9648	0.7573

Table B.4: KW Test for TDP Feature in dataset 2

SRTD				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.0056	0.0022	0.0094	0.013
20 OT	0.4528	0.1332	0.2726	0.1488
40 OT	0.9539	0.8173	0.3864	0.6441
60 OT	0.5636	0.9081	0.9539	0.9539
SRTFD				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.0153	0.0567	0.0047	0.0047
20 OT	0.5253	0.1841	0.0734	0.4188
40 OT	0.8173	0.9081	0.4528	0.9539
60 OT	0.8173	0.5253	0.686	0.686
SA				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.1488	0.0243	0.0015	0.0018
20 OT	0.1332	0.0734	0.0376	0.0496
40 OT	1	0.4528	0.6033	0.4188
60 OT	0.9081	0.6032	0.6441	0.9081

Table B.5: KW Test for CSP Feature in dataset 2

SRTD				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.5825	0.8623	1.0000	0.5438
20 OT	0.8624	0.9539	1.0000	0.9539
40 OT	0.9770	0.9080	1.0000	0.8623
60 OT	0.8173	0.8398	0.8624	0.9539
SRTFD				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.8395	0.7725	0.6857	0.5434
20 OT	0.9539	0.6859	0.8624	0.8172
40 OT	0.8172	0.7725	0.6031	0.6859
60 OT	0.9080	0.8173	0.9539	0.9539
SA				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.9769	0.7068	0.9539	0.9539
20 OT	1.0000	0.7726	0.7727	0.7726
40 OT	0.6440	0.6858	0.8172	0.5634
60 OT	1.0000	0.6860	0.8173	0.9539

Table B.6: KW Test for BP Feature in dataset 2

SRTD				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.3555	0.6860	0.9081	0.8173
20 OT	0.9539	0.8625	0.9081	1.0000
40 OT	0.6860	0.7728	0.8625	0.8173
60 OT	0.9539	0.9081	0.9539	0.9539
SRTFD				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.4883	0.5636	0.6033	0.6033
20 OT	0.8173	1.0000	0.9081	0.7290
40 OT	0.8625	0.9539	0.8173	0.7290
60 OT	0.9081	0.9081	0.8625	0.9081
SA				
	10 AT	**20 AT**	**40 AT**	**60 AT**
10 OT	0.4188	0.5253	0.4703	0.5636
20 OT	0.7728	0.8173	0.8173	0.8625
40 OT	0.9539	0.7728	1.0000	0.9081
60 OT	1.0000	1.0000	0.9081	0.9081

Bibliography

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

[2] J. J. Vidal, "Toward direct brain-computer communication," *Annual review of Biophysics and Bioengineering*, vol. 2, no. 1, pp. 157–180, 1973.

[3] B. H. 2020, "Basics about bci," http://bnci-horizon-2020.eu/about/basics, accessed Jun 15, 2022.

[4] J. N. Mak, D. J. McFarland, T. M. Vaughan, L. M. McCane, P. Z. Tsui, D. J. Zeitlin, E. W. Sellers, and J. R. Wolpaw, "Eeg correlates of p300-based brain–computer interface (bci) performance in people with amyotrophic lateral sclerosis," *Journal of neural engineering*, vol. 9, no. 2, p. 026014, 2012.

[5] N. Riva, A. Falini, A. Inuggi, J. Gonzalez-Rosa, S. Amadio, F. Cerri, R. Fazio, U. Del Carro, M. Comola, G. Comi *et al.*, "Cortical activation to voluntary movement in amyotrophic lateral sclerosis is related to corticospinal damage: electrophysiological evidence," *Clinical Neurophysiology*, vol. 123, no. 8, pp. 1586–1592, 2012.

[6] B. Steenbergen, C. Craje, D. M. Nilsen, and A. M. Gordon, "Motor imagery training in hemiplegic cerebral palsy: a potentially useful therapeutic tool for rehabilitation," *Developmental Medicine & Child Neurology*, vol. 51, no. 9, pp. 690–696, 2009.

[7] I. Daly, M. Billinger, J. Laparra-Hernández, F. Aloise, M. L. García, J. Faller, R. Scherer, and G. Müller-Putz, "On the control of brain-computer interfaces by users with cerebral palsy," *Clinical Neurophysiology*, vol. 124, no. 9, pp. 1787–1797, 2013.

[8] S. Ikegami, K. Takano, N. Saeki, and K. Kansaku, "Operation of a p300-based brain–computer interface by individuals with cervical spinal cord injury," *Clinical Neurophysiology*, vol. 122, no. 5, pp. 991–996, 2011.

[9] X. Huo and M. Ghovanloo, "Evaluation of a wireless wearable tongue–computer interface by individuals with high-level spinal cord injuries," *Journal of neural engineering*, vol. 7, no. 2, p. 026008, 2010.

[10] G. Pfurtscheller, B. Z. Allison, G. Bauernfeind, C. Brunner, T. Solis Escalante, R. Scherer, T. O. Zander, G. Mueller-Putz, C. Neuper, and N. Birbaumer, "The hybrid bci," *Frontiers in neuroscience*, vol. 4, p. 3, 2010.

[11] C. Brunner, B. Blankertz, F. Cincotti, A. Kübler, D. Mattia, F. Miralles, A. Nijholt, B. Otal, P. Salomon, and G. R. Müller-Putz, "Bnci horizon 2020–towards a roadmap for brain/neural computer interaction," in *International Conference on Universal Access in Human-Computer Interaction*. Springer, 2014, pp. 475–486.

[12] J. Gomez-Gil *et al.*, "Brain computer interfaces, a review," *Sensors*, pp. 1211–1279, 2012.

[13] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Muller, "Boosting bit rates in noninvasive eeg single-trial classifications by feature combination and multiclass paradigms," *IEEE transactions on biomedical engineering*, vol. 51, no. 6, pp. 993–1002, 2004.

[14] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 871–890, 2015.

[15] M. Bear, B. Connors, and M. A. Paradiso, *Neuroscience: Exploring the Brain, Enhanced Edition: Exploring the Brain*. Jones & Bartlett Learning, 2020.

[16] K. W. Horch and D. R. Kipke, *Neuroprosthetics: theory and practice*. World Scientific, 2017, vol. 8.

[17] M. Jeannerod, "Neural simulation of action: a unifying mechanism for motor cognition," *Neuroimage*, vol. 14, no. 1, pp. S103–S109, 2001.

[18] A. Sosnowska, H. Gollee, and A. Vučković, "Mrcp as a biomarker of motor action with varying degree of central and peripheral contribution as defined

by ultrasound imaging," *Journal of Neurophysiology*, vol. 126, no. 1, pp. 249–263, 2021.

[19] C. Neuper, R. Scherer, M. Reiner, and G. Pfurtscheller, "Imagery of motor actions: Differential effects of kinesthetic and visual–motor mode of imagery in single-trial eeg," *Cognitive brain research*, vol. 25, no. 3, pp. 668–677, 2005.

[20] Wikipedia, "Human visual pathways," https://en.wikipedia.org/wiki/Visual_system#/media/File:Human_visual_pathway.svg, accessed Jun 11, 2022.

[21] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neurociências: desvendando o sistema nervoso*. Artmed editora, 2002.

[22] A. M. Dale and E. Halgren, "Spatiotemporal mapping of brain activity by integration of multiple imaging modalities," *Current opinion in neurobiology*, vol. 11, no. 2, pp. 202–208, 2001.

[23] H.-J. Hwang, H. Choi, J.-Y. Kim, W.-D. Chang, D.-W. Kim, K. Kim, S. Jo, and C.-H. Im, "Toward more intuitive brain–computer interfacing: classification of binary covert intentions using functional near-infrared spectroscopy," *Journal of biomedical optics*, vol. 21, no. 9, pp. 091 303–091 303, 2016.

[24] N. Naseer and K.-S. Hong, "Classification of functional near-infrared spectroscopy signals corresponding to the right-and left-wrist motor imagery for development of a brain–computer interface," *Neuroscience letters*, vol. 553, pp. 84–89, 2013.

[25] L. C. Schudlo and T. Chau, "Development of a ternary near-infrared spectroscopy brain-computer interface: Online classification of verbal fluency task, stroop task and rest," *International journal of neural systems*, vol. 28, no. 04, p. 1750052, 2018.

[26] E. Niedermeyer *et al.*, "The normal eeg of the waking adult," *Electroencephalography: Basic principles, clinical applications, and related fields*, vol. 167, pp. 155–164, 2005.

[27] G. Pfurtscheller and F. L. Da Silva, "Event-related eeg/meg synchronization and desynchronization: basic principles," *Clinical neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.

[28] C. Babiloni, R. J. Barry, E. Başar, K. J. Blinowska, A. Cichocki, W. H. Drinkenburg, W. Klimesch, R. T. Knight, F. L. da Silva, P. Nunez *et al.*, "International federation of clinical neurophysiology (ifcn)–eeg research workgroup: Recommendations on frequency and topographic analysis of resting state eeg rhythms. part 1: Applications in clinical research studies," *Clinical Neurophysiology*, vol. 131, no. 1, pp. 285–307, 2020.

[29] K. J. Blinowska, "Review of the methods of determination of directed connectivity from multichannel data," *Medical & biological engineering & computing*, vol. 49, no. 5, pp. 521–529, 2011.

[30] D.-W. Kim and C.-H. Im, "Eeg spectral analysis," *Computational EEG analysis: Methods and applications*, pp. 35–53, 2018.

[31] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems," *Neuroimage*, vol. 34, no. 4, pp. 1600–1611, 2007.

[32] J. N. Acharya and V. J. Acharya, "Overview of eeg montages and principles of localization," *Journal of Clinical Neurophysiology*, vol. 36, no. 5, pp. 325–329, 2019.

[33] R. Gordon and E. J. Rzempoluck, "Introduction to laplacian montages," *American journal of electroneurodiagnostic technology*, vol. 44, no. 2, pp. 98–102, 2004.

[34] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Handwritten digit recognition with a back-propagation network," *Advances in neural information processing systems*, vol. 2, 1989.

[35] P. Kant, S. H. Laskar, J. Hazarika, and R. Mahamune, "Cwt based transfer learning for motor imagery classification for brain computer interfaces," *Journal of Neuroscience Methods*, vol. 345, p. 108886, 2020.

[36] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[37] J. V. Odom, M. Bach, C. Barber, M. Brigell, M. F. Marmor, A. P. Tormene, G. E. Holder *et al.*, "Visual evoked potentials standard (2004)," *Documenta ophthalmologica*, vol. 108, no. 2, pp. 115–123, 2004.

[38] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[39] Z. Wu, Y. Lai, Y. Xia, D. Wu, and D. Yao, "Stimulator selection in ssvep-based bci," *Medical engineering & physics*, vol. 30, no. 8, pp. 1079–1088, 2008.

[40] Y. Zhang, P. Xu, T. Liu, J. Hu, R. Zhang, and D. Yao, "Multiple frequencies sequential coding for ssvep-based brain-computer interface," *PloS one*, vol. 7, no. 3, p. e29519, 2012.

[41] M. Cheng, X. Gao, S. Gao, and D. Xu, "Design and implementation of a brain-computer interface with high transfer rates," *IEEE transactions on biomedical engineering*, vol. 49, no. 10, pp. 1181–1186, 2002.

[42] D. Regan, "Evoked potentials and evoked magnetic fields in science and medicine," *Human brain electrophysiology*, pp. 59–61, 1989.

[43] M. Middendorf, G. McMillan, G. Calhoun, and K. S. Jones, "Brain-computer interfaces based on the steady-state visual-evoked response," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 2, pp. 211–214, 2000.

[44] F. Guo, B. Hong, X. Gao, and S. Gao, "A brain–computer interface using motion-onset visual evoked potential," *Journal of neural engineering*, vol. 5, no. 4, p. 477, 2008.

[45] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[46] E. W. Sellers and E. Donchin, "A p300-based brain–computer interface: initial tests by als patients," *Clinical neurophysiology*, vol. 117, no. 3, pp. 538–548, 2006.

[47] F. Nijboer, E. Sellers, J. Mellinger, M. A. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D. Krusienski, T. Vaughan *et al.*, "A p300-based brain–computer interface for people with amyotrophic lateral sclerosis," *Clinical neurophysiology*, vol. 119, no. 8, pp. 1909–1916, 2008.

[48] F. Piccione, F. Giorgi, P. Tonin, K. Priftis, S. Giove, S. Silvoni, G. Palmas, and F. Beverina, "P300-based brain computer interface:

reliability and performance in healthy and paralysed participants," *Clinical neurophysiology*, vol. 117, no. 3, pp. 531–537, 2006.

[49] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient p300-based brain–computer interface for disabled subjects," *Journal of Neuroscience methods*, vol. 167, no. 1, pp. 115–125, 2008.

[50] A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey, "A brain–computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients," *Annals of the New York Academy of Sciences*, vol. 1157, no. 1, pp. 90–100, 2009.

[51] M. E. Cano, Q. A. Class, and J. Polich, "Affective valence, stimulus attributes, and p300: color vs. black/white and normal vs. scrambled images," *International Journal of Psychophysiology*, vol. 71, no. 1, pp. 17–24, 2009.

[52] M. Bekdash, V. S. Asirvadam, N. Kamel, and D. K. Hutapea, "Identifying the human attention to different colors and intensities using p300," in *2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, 2015, pp. 538–541.

[53] X. Zhang, J. Jin, S. Li, X. Wang, and A. Cichocki, "Evaluation of color modulation in visual p300-speller using new stimulus patterns," *Cognitive Neurodynamics*, vol. 15, no. 5, pp. 873–886, 2021.

[54] M. Li, G. Yang, and G. Xu, "The effect of the graphic structures of humanoid robot on n200 and p300 potentials," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 9, pp. 1944–1954, 2020.

[55] Y. Sakai and T. Yagi, "Alphabet matrix layout in p300 speller may alter its performance," in *The 4th 2011 Biomedical Engineering International Conference*. IEEE, 2012, pp. 89–92.

[56] C. Guger, G. Krausz, B. Z. Allison, and G. Edlinger, "Comparison of dry and gel based electrodes for p300 brain–computer interfaces," *Frontiers in neuroscience*, vol. 6, p. 60, 2012.

[57] Y. Zhao, J. Tang, Y. Cao, X. Jiao, M. Xu, P. Zhou, D. Ming, and H. Qi, "Effects of distracting task with different mental workload on steady-state visual evoked potential based brain computer interfaces—an offline study," *Frontiers in neuroscience*, vol. 12, p. 79, 2018.

[58] J. Zhao, W. Li, and M. Li, "Comparative study of ssvep-and p300-based models for the telepresence control of humanoid robots," *PLoS One*, vol. 10, no. 11, p. e0142168, 2015.

[59] J. Xie, G. Xu, A. Luo, M. Li, S. Zhang, C. Han, and W. Yan, "The role of visual noise in influencing mental load and fatigue in a steady-state motion visual evoked potential-based brain-computer interface," *Sensors*, vol. 17, no. 8, p. 1873, 2017.

[60] J. Xie, G. Xu, J. Wang, M. Li, C. Han, and Y. Jia, "Effects of mental load and fatigue on steady-state evoked potential based brain computer interface tasks: a comparison of periodic flickering and motion-reversal based visual attention," *PloS one*, vol. 11, no. 9, p. e0163426, 2016.

[61] M. E. Cano, Q. A. Class, and J. Polich, "Affective valence, stimulus attributes, and p300: color vs. black/white and normal vs. scrambled images," *International Journal of Psychophysiology*, vol. 71, no. 1, pp. 17–24, 2009.

[62] S. Ikegami, K. Takano, M. Wada, N. Saeki, and K. Kansaku, "Effect of the green/blue flicker matrix for p300-based brain–computer interface: An eeg–fmri study," *Frontiers in neurology*, vol. 3, p. 113, 2012.

[63] M. Bekdash, V. S. Asirvadam, N. Kamel, and D. K. Hutapea, "Identifying the human attention to different colors and intensities using p300," in *2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, 2015, pp. 538–541.

[64] D. Regan, "Some characteristics of average steady-state and transient responses evoked by modulated light," *Electroencephalography and clinical neurophysiology*, vol. 20, no. 3, pp. 238–248, 1966.

[65] K. Arakawa, S. Tobimatsu, H. Tomoda, J.-i. Kira, and M. Kato, "The effect of spatial frequency on chromatic and achromatic steady-state visual evoked potentials," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1959–1964, 1999.

[66] J. Bieger, G. G. Molina, and D. Zhu, "Effects of stimulation properties in steady-state visual evoked potential based brain-computer interfaces," in *Proceedings of 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Engineering in Medicine and Biology Society, 2010, pp. 3345–8.

[67] T. Cao, F. Wan, P. U. Mak, P.-I. Mak, M. I. Vai, and Y. Hu, "Flashing color on the performance of ssvep-based brain-computer interfaces," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 1819–1822.

[68] R. Singla, A. Khosla, and R. Jha, "Influence of stimuli colour in ssvep-based bci wheelchair control using support vector machines," *Journal of medical engineering & technology*, vol. 38, no. 3, pp. 125–134, 2014.

[69] R. Beveridge, D. Marshall, S. Wilson, and D. Coyle, "3d game graphic complexity effects on motion-onset visual evoked potentials," in *8th Annual International Conference on Computer Games, Multimedia and Allied Technology (CGAT 2015)*. Global Science and Technology Forum, 2015, pp. 139–147.

[70] ——, "Classification effects on motion-onset visual evoked potentials using commercially available video games," in *2015 Computer Games: AI, Animation, Mobile, Multimedia, Educational and Serious Games (CGAMES)*. IEEE, 2015, pp. 28–37.

[71] D. Marshall, R. Beveridge, S. Wilson, and D. Coyle, "Interacting with multiple game genres using motion onset visual evoked potentials," in *2015 Computer Games: AI, Animation, Mobile, Multimedia, Educational and Serious Games (CGAMES)*. IEEE, 2015, pp. 18–27.

[72] M. A. Khan, R. Das, H. K. Iversen, and S. Puthusserypady, "Review on motor imagery based bci systems for upper limb post-stroke neurorehabilitation: From designing to application," *Computers in Biology and Medicine*, vol. 123, p. 103843, 2020.

[73] M. Tariq, P. M. Trivailo, and M. Simic, "Eeg-based bci control schemes for lower-limb assistive-robots," *Frontiers in human neuroscience*, vol. 12, p. 312, 2018.

[74] J. A. Gaxiola-Tirado, E. Iáñez, M. Ortíz, D. Gutiérrez, and J. M. Azorín, "Effects of an exoskeleton-assisted gait motor imagery training in functional brain connectivity," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 429–432.

[75] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[76] E. B. Sadeghian and M. H. Moradi, "Continuous detection of motor imagery in a four-class asynchronous bci," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 3241–3244.

[77] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[78] A. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[79] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "Eeg-based brain-computer interfaces using motor-imagery: Techniques and challenges," *Sensors*, vol. 19, no. 6, p. 1423, 2019.

[80] A. Singh, A. A. Hussain, S. Lal, and H. W. Guesgen, "A comprehensive review on critical issues and possible solutions of motor imagery based electroencephalography brain-computer interface," *Sensors*, vol. 21, no. 6, p. 2173, 2021.

[81] X. Xiao and Y. Fang, "Motor imagery eeg signal recognition using deep convolution neural network," *Frontiers in Neuroscience*, vol. 15, p. 655599, 2021.

[82] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, and Y.-Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery eeg decoding," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 6, pp. 1170–1180, 2019.

[83] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[84] R. Alazrai, M. Abuhijleh, H. Alwanni, and M. I. Daoud, "A deep learning framework for decoding motor imagery tasks of the same hand using eeg signals," *IEEE Access*, vol. 7, pp. 109 612–109 627, 2019.

[85] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of eeg motor imagery signals," *Journal of neural engineering*, vol. 14, no. 1, p. 016003, 2016.

[86] G. Xu, X. Shen, S. Chen, Y. Zong, C. Zhang, H. Yue, M. Liu, F. Chen, and W. Che, "A deep transfer convolutional neural network framework for eeg signal classification," *IEEE Access*, vol. 7, pp. 112 767–112 776, 2019.

[87] R. Zhang, Q. Zong, L. Dou, and X. Zhao, "A novel hybrid deep learning scheme for four-class motor imagery classification," *Journal of neural engineering*, vol. 16, no. 6, p. 066004, 2019.

[88] S. U. Amin, M. Alsulaiman, G. Muhammad, M. S. Hossain, and M. Guizani, "Deep learning for eeg motor imagery-based cognitive healthcare," in *Connected Health in Smart Cities*. Springer, 2020, pp. 233–254.

[89] R. Zhang, Q. Zong, L. Dou, and X. Zhao, "A novel hybrid deep learning scheme for four-class motor imagery classification," *Journal of neural engineering*, vol. 16, no. 6, p. 066004, 2019.

[90] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, and M. S. Hossain, "Deep learning for eeg motor imagery classification based on multi-layer cnns feature fusion," *Future Generation computer systems*, vol. 101, pp. 542–554, 2019.

[91] H. Wu, Y. Niu, F. Li, Y. Li, B. Fu, G. Shi, and M. Dong, "A parallel multiscale filter bank convolutional neural networks for motor imagery eeg classification," *Frontiers in neuroscience*, vol. 13, p. 1275, 2019.

[92] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain, "Multilevel weighted feature fusion using convolutional neural networks for eeg motor imagery classification," *Ieee Access*, vol. 7, pp. 18 940–18 950, 2019.

[93] H. Wu, Y. Niu, F. Li, Y. Li, B. Fu, G. Shi, and M. Dong, "A parallel multiscale filter bank convolutional neural networks for motor imagery eeg classification," *Frontiers in neuroscience*, vol. 13, p. 1275, 2019.

[94] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, and Y.-Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery eeg decoding," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 6, pp. 1170–1180, 2019.

[95] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain, "Multilevel weighted feature fusion using convolutional neural networks for eeg motor imagery classification," *Ieee Access*, vol. 7, pp. 18 940–18 950, 2019.

[96] B. E. Olivas-Padilla and M. I. Chacon-Murguia, "Classification of multiple motor imagery using deep convolutional neural networks and spatial filters," *Applied Soft Computing*, vol. 75, pp. 461–472, 2019.

[97] R. Zhang, Q. Zong, L. Dou, and X. Zhao, "A novel hybrid deep learning scheme for four-class motor imagery classification," *Journal of neural engineering*, vol. 16, no. 6, p. 066004, 2019.

[98] I. Majidov and T. Whangbo, "Efficient classification of motor imagery electroencephalography signals using deep learning methods," *Sensors*, vol. 19, no. 7, p. 1736, 2019.

[99] T.-j. Luo, C.-l. Zhou, and F. Chao, "Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network," *BMC bioinformatics*, vol. 19, no. 1, pp. 1–18, 2018.

[100] D. Li, J. Wang, J. Xu, and X. Fang, "Densely feature fusion based on convolutional neural networks for motor imagery eeg classification," *IEEE Access*, vol. 7, pp. 132 720–132 730, 2019.

[101] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, "A multi-branch 3d convolutional neural network for eeg-based motor imagery classification," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 27, no. 10, pp. 2164–2177, 2019.

[102] Z. Zhang, F. Duan, J. Sole-Casals, J. Dinares-Ferran, A. Cichocki, Z. Yang, and Z. Sun, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15 945–15 954, 2019.

[103] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from eeg with deep recurrent-convolutional neural networks," *arXiv preprint arXiv:1511.06448*, 2015.

[104] F. Wang, S.-h. Zhong, J. Peng, J. Jiang, and Y. Liu, "Data augmentation for eeg-based emotion recognition with deep convolutional neural networks," in *International conference on multimedia modeling*. Springer, 2018, pp. 82–93.

[105] M. Parvan, A. R. Ghiasi, T. Y. Rezaii, and A. Farzamnia, "Transfer learning based motor imagery classification using convolutional neural networks," in *2019 27th Iranian Conference on Electrical Engineering (ICEE)*. IEEE, 2019, pp. 1825–1828.

[106] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[107] H. Huang, P. S. Yu, and C. Wang, "An introduction to image synthesis with generative adversarial nets," *arXiv preprint arXiv:1803.04469*, 2018.

[108] H. Wang, Z. Qin, and T. Wan, "Text generation based on generative adversarial nets with latent variables," in *Advances in Knowledge Discovery and Data Mining: 22nd Pacific-Asia Conference, PAKDD 2018, Melbourne, VIC, Australia, June 3-6, 2018, Proceedings, Part II 22*. Springer, 2018, pp. 92–103.

[109] N. Aldausari, A. Sowmya, N. Marcus, and G. Mohammadi, "Video generative adversarial networks: a review," *ACM Computing Surveys (CSUR)*, vol. 55, no. 2, pp. 1–25, 2022.

[110] H.-W. Dong and Y.-H. Yang, "Convolutional generative adversarial networks with binary neurons for polyphonic music generation," *arXiv preprint arXiv:1804.09399*, 2018.

[111] Y. Khaldi and A. Benzaoui, "A new framework for grayscale ear images recognition using generative adversarial networks under unconstrained conditions," *Evolving Systems*, vol. 12, no. 4, pp. 923–934, 2021.

[112] L. Sun, K. Wang, K. Yang, and K. Xiang, "See clearer at night: towards robust nighttime semantic segmentation through day-night image conversion," in *Artificial Intelligence and Machine Learning in Defense Applications*, vol. 11169. SPIE, 2019, pp. 77–89.

[113] A. Kammoun, R. Slama, H. Tabia, T. Ouni, and M. Abid, "Generative adversarial networks for face generation: A survey," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1–37, 2022.

[114] H. Berard, G. Gidel, A. Almahairi, P. Vincent, and S. Lacoste-Julien, "A closer look at the optimization landscapes of generative adversarial networks," *arXiv preprint arXiv:1906.04848*, 2019.

[115] J. Xu, X. Ren, J. Lin, and X. Sun, "Diversity-promoting gan: A cross-entropy based generative adversarial network for diversified text generation," in *Proceedings of the 2018 conference on empirical methods in natural language processing*, 2018, pp. 3940–3949.

[116] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *International conference on machine learning*. PMLR, 2016, pp. 1060–1069.

[117] Z. Yang, W. Chen, F. Wang, and B. Xu, "Improving neural machine translation with conditional sequence generative adversarial nets," *arXiv preprint arXiv:1703.04887*, 2017.

[118] F. Fahimi, S. Dosen, K. K. Ang, N. Mrachacz-Kersting, and C. Guan, "Generative adversarial networks-based data augmentation for brain–computer interface," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 9, pp. 4039–4051, 2020.

[119] K. Zhang, G. Xu, Z. Han, K. Ma, X. Zheng, L. Chen, N. Duan, and S. Zhang, "Data augmentation for motor imagery signal classification based on a hybrid neural network," *Sensors*, vol. 20, no. 16, p. 4485, 2020.

[120] Z. Li and Y. Yu, "Improving eeg-based motor imagery classification with conditional wasserstein gan," in *2020 International Conference on Image, Video Processing and Artificial Intelligence*, vol. 11584. SPIE, 2020, pp. 437–443.

[121] M. Parvan, A. R. Ghiasi, T. Y. Rezaii, and A. Farzamnia, "Transfer learning based motor imagery classification using convolutional neural networks," in *2019 27th Iranian Conference on Electrical Engineering (ICEE)*. IEEE, 2019, pp. 1825–1828.

[122] G. Dai, J. Zhou, J. Huang, and N. Wang, "Hs-cnn: a cnn with hybrid convolution scale for eeg motor imagery classification," *Journal of neural engineering*, vol. 17, no. 1, p. 016025, 2020.

[123] O. Özdenizci, Y. Wang, T. Koike-Akino, and D. Erdoğmuş, "Adversarial deep learning in eeg biometrics," *IEEE signal processing letters*, vol. 26, no. 5, pp. 710–714, 2019.

[124] Y. Ming, W. Ding, D. Pelusi, D. Wu, Y.-K. Wang, M. Prasad, and C.-T. Lin, "Subject adaptation network for eeg data analysis," *Applied Soft Computing*, vol. 84, p. 105689, 2019.

[125] V. Kushwaha, G. Nandi *et al.*, "Study of prevention of mode collapse in generative adversarial network (gan)," in *2020 IEEE 4th Conference on Information & Communication Technology (CICT)*. IEEE, 2020, pp. 1–6.

[126] H. Lu, H.-L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for eeg classification in small-sample setting," *IEEE transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2936–2946, 2010.

[127] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Processing Letters*, vol. 16, no. 8, pp. 683–686, 2009.

[128] F. Lotte and C. Guan, "Learning from other subjects helps reducing brain-computer interface calibration time," in *2010 IEEE International conference on acoustics, speech and signal processing*. IEEE, 2010, pp. 614–617.

[129] D. Heger, F. Putze, C. Herff, and T. Schultz, "Subject-to-subject transfer for csp based bcis: Feature space transformation and decision-level fusion," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 5614–5617.

[130] A. M. Azab, J. Toth, L. S. Mihaylova, and M. Arvaneh, "A review on transfer learning approaches in brain–computer interface," *Signal Processing and Machine Learning for Brain-Machine Interfaces*, pp. 81–98, 2018.

[131] A. M. Norcia, L. G. Appelbaum, J. M. Ales, B. R. Cottereau, and B. Rossion, "The steady-state visual evoked potential in vision research: A review," *Journal of vision*, vol. 15, no. 6, pp. 4–4, 2015.

[132] P. L. Nunez and B. A. Cutillo, *Neocortical dynamics and human EEG rhythms*. Oxford University Press, USA, 1995.

[133] R. Singla, "Ssvep-based bcis," *Evolving BCI Therapy: Engaging Brain State Dynamics*, p. 91, 2018.

[134] I. Schechter, P. D. Butler, V. M. Zemon, N. Revheim, A. M. Saperstein, M. Jalbrzikowski, R. Pasternak, G. Silipo, and D. C. Javitt, "Impairments in generation of early-stage transient visual evoked potentials to magno-and parvocellular-selective stimuli in schizophrenia," *Clinical Neurophysiology*, vol. 116, no. 9, pp. 2204–2215, 2005.

[135] P. M. Siper, V. Zemon, J. Gordon, J. George-Jones, S. Lurie, J. Zweifach, T. Tavassoli, A. T. Wang, J. Jamison, J. D. Buxbaum *et al.*, "Rapid and

objective assessment of neural function in autism spectrum disorder using transient visual evoked potentials," *PLoS One*, vol. 11, no. 10, p. e0164422, 2016.

[136] D. Ryan, G. Townsend, N. Gates, K. Colwell, and E. Sellers, "Evaluating brain-computer interface performance using color in the p300 checkerboard speller," *Clinical Neurophysiology*, vol. 128, no. 10, pp. 2050–2057, 2017.

[137] T. Ma, F. Li, P. Li, D. Yao, Y. Zhang, and P. Xu, "An adaptive calibration framework for mvep-based brain-computer interface," *Computational and Mathematical Methods in Medicine*, vol. 2018, 2018.

[138] D. Marshall, D. Coyle, and S. Wilson, "Motion-onset visual evoked potentials for gaming: A pilot study," *IET*, 2013.

[139] F. Guo, B. Hong, X. Gao, and S. Gao, "A brain–computer interface using motion-onset visual evoked potential," *Journal of neural engineering*, vol. 5, no. 4, p. 477, 2008.

[140] E. Göpfert, R. Müller, F. Markwardt, and L. Schlykowa, "Visuell evozierte potentiale bei musterbewegung," *EEG-EMG Zeitschrift für Elektroenzephalographie, Elektromyographie und verwandte Gebiete*, vol. 14, no. 01, pp. 47–51, 1983.

[141] M. Kuba and Z. Kubová, "Visual evoked potentials specific for motion onset," *Documenta Ophthalmologica*, vol. 80, no. 1, pp. 83–89, 1992.

[142] M. Kuba, N. Toyonaga, and Z. Kubová, "Motion-reversal visual evoked response," *Physiological Research*, vol. 41, pp. 369–369, 1992.

[143] M. Bach and D. Ullrich, "Motion adaptation governs the shape of motion-evoked cortical potentials," *Vision Research*, vol. 34, no. 12, pp. 1541–1547, 1994.

[144] W. Skrandies, A. Jedynak, and R. Kleiser, "Scalp distribution components of brain activity evoked by visual motion stimuli," *Experimental brain research*, vol. 122, no. 1, pp. 62–70, 1998.

[145] S. P. Heinrich and M. Bach, "Adaptation characteristics of steady-state motion visual evoked potentials," *Clinical Neurophysiology*, vol. 114, no. 7, pp. 1359–1366, 2003.

[146] M. Kuba, Z. Kubová, J. Kremláček, and J. Langrová, "Motion-onset veps: characteristics, methods, and diagnostic use," *Vision Research*, vol. 47, no. 2, pp. 189–202, 2007.

[147] J. Kremláček, M. Kuba, Z. Kubova, and J. Chlubnova, "Motion-onset veps to translating, radial, rotating and spiral stimuli," *Documenta Ophthalmologica*, vol. 109, no. 2, pp. 169–175, 2004.

[148] T. Lyu, A. Vuckovic, and T. Liu, "The influence of graphical presentation on motion onset visual evoked potentialfor brain computer interface," in *8th International BCI Meeting*. BCI Society, 2021, p. 125.

[149] R. Zhang, P. Xu, R. Chen, T. Ma, X. Lv, F. Li, P. Li, T. Liu, and D. Yao, "An adaptive motion-onset vep-based brain-computer interface," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 4, pp. 349–356, 2015.

[150] R. Beveridge, S. Wilson, and D. Coyle, "3d graphics, virtual reality, and motion-onset visual evoked potentials in neurogaming," *Progress in brain research*, vol. 228, pp. 329–353, 2016.

[151] E. Niedermeyer and F. L. da Silva, *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[152] G. R. Müller-Putz, "Electroencephalography," *Handbook of Clinical Neurology*, vol. 168, pp. 249–262, 2020.

[153] W. Klonowski, "Everything you wanted to ask about eeg but were afraid to get the right answer," *Nonlinear biomedical physics*, vol. 3, no. 1, pp. 1–5, 2009.

[154] J. León, J. J. Escobar, A. Ortiz, J. Ortega, J. González, P. Martín-Smith, J. Q. Gan, and M. Damas, "Deep learning for eeg-based motor imagery classification: Accuracy-cost trade-off," *Plos one*, vol. 15, no. 6, p. e0234178, 2020.

[155] R. Rupp, "Challenges in clinical applications of brain computer interfaces in individuals with spinal cord injury," *Frontiers in neuroengineering*, vol. 7, p. 38, 2014.

[156] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, and K.-R. Müller, "The berlin brain-computer interface: Accurate performance

from first-session in bci-naive subjects," *IEEE transactions on biomedical engineering*, vol. 55, no. 10, pp. 2452–2462, 2008.

[157] M. Zheng, B. Yang, and Y. Xie, "Eeg classification across sessions and across subjects through transfer learning in motor imagery-based brain-machine interface system," *Medical & biological engineering & computing*, vol. 58, no. 7, pp. 1515–1528, 2020.

[158] L. Bougrain, S. Rimbert, P. L. Rodrigues, G. Canron, and F. Lotte, "Guidelines to use transfer learning for motor imagery detection: an experimental study," in *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2021, pp. 5–8.

[159] E. Lashgari, D. Liang, and U. Maoz, "Data augmentation for deep-learning-based electroencephalography," *Journal of Neuroscience Methods*, vol. 346, p. 108885, 2020.

[160] Z. Zhang, F. Duan, J. Sole-Casals, J. Dinares-Ferran, A. Cichocki, Z. Yang, and Z. Sun, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15 945–15 954, 2019.

[161] A. Sakai, Y. Minoda, and K. Morikawa, "Data augmentation methods for machine-learning-based classification of bio-signals," in *2017 10th Biomedical Engineering International Conference (BMEiCON)*. IEEE, 2017, pp. 1–4.

[162] S. Kuanar, V. Athitsos, N. Pradhan, A. Mishra, and K. R. Rao, "Cognitive analysis of working memory load from eeg, by a deep recurrent neural network," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2576–2580.

[163] F. Wang, S.-h. Zhong, J. Peng, J. Jiang, and Y. Liu, "Data augmentation for eeg-based emotion recognition with deep convolutional neural networks," in *International conference on multimedia modeling*. Springer, 2018, pp. 82–93.

[164] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PloS one*, vol. 12, no. 2, p. e0172578, 2017.

[165] Z. Chen, Y. Wang, and Z. Song, "Classification of motor imagery electroencephalography signals based on image processing method," *Sensors*, vol. 21, no. 14, p. 4646, 2021.

[166] Y. Aslam and N. Santhi, "A review of deep learning approaches for image analysis," in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE, 2019, pp. 709–714.

[167] J. Kukačka, V. Golkov, and D. Cremers, "Regularization for deep learning: A taxonomy," *arXiv preprint arXiv:1710.10686*, 2017.

[168] Q. Zhang and Y. Liu, "Improving brain computer interface performance by data augmentation with conditional deep convolutional generative adversarial networks," *arXiv preprint arXiv:1806.07108*, 2018.

[169] K. Zhang, G. Xu, Z. Han, K. Ma, X. Zheng, L. Chen, N. Duan, and S. Zhang, "Data augmentation for motor imagery signal classification based on a hybrid neural network," *Sensors*, vol. 20, no. 16, p. 4485, 2020.

[170] Y. Pei, Z. Luo, Y. Yan, H. Yan, J. Jiang, W. Li, L. Xie, and E. Yin, "Data augmentation: Using channel-level recombination to improve classification performance for motor imagery eeg," *Frontiers in Human Neuroscience*, vol. 15, p. 645952, 2021.

[171] F. Lotte, "Generating artificial eeg signals to reduce bci calibration time," in *5th International Brain-Computer Interface Workshop*, 2011, pp. 176–179.

[172] D.-F. J. O. R. Guger, "C solé-casals j a new method to generate artificial frames using the empirical mode decomposition for an eeg-based motor imagery bci front," *Neurosci*, vol. 12, no. 308, pp. 10–3389, 2018.

[173] W. Xiong and Q. Wei, "Reducing calibration time in motor imagery-based bcis by data alignment and empirical mode decomposition," *Plos one*, vol. 17, no. 2, p. e0263641, 2022.

[174] B. Clemens, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "Bci competition iv," https://bbci.de/competition/iv/, accessed Jun 19, 2022.

[175] B. Blankertz, "Bci competition iii," https://www.bbci.de/competition/iii/, accessed Jun 15, 2022.

[176] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl, "Time domain parameters as a feature for eeg-based brain–computer interfaces," *Neural Networks*, vol. 22, no. 9, pp. 1313–1319, 2009.

[177] C. McGeady and A. Vuckovic, "Is classifying uni- and bimanual motor imagery feasible as a three class bci problem?" in *8th International BCI Meeting*. BCI Society, 2021, p. 27.

[178] G. Pfurtscheller and F. L. Da Silva, "Event-related eeg/meg synchronization and desynchronization: basic principles," *Clinical neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.

[179] Y. Jeon, C. S. Nam, Y.-J. Kim, and M. C. Whang, "Event-related (de) synchronization (erd/ers) during motor imagery tasks: Implications for brain–computer interfaces," *International Journal of Industrial Ergonomics*, vol. 41, no. 5, pp. 428–436, 2011.

[180] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *2005 IEEE engineering in medicine and biology 27th annual conference*. IEEE, 2006, pp. 5392–5395.

[181] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Recipes for the linear analysis of eeg," *Neuroimage*, vol. 28, no. 2, pp. 326–341, 2005.

[182] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update," *Journal of neural engineering*, vol. 15, no. 3, p. 031005, 2018.

[183] J. Dinarès-Ferran, R. Ortner, C. Guger, and J. Solé-Casals, "A new method to generate artificial frames using the empirical mode decomposition for an eeg-based motor imagery bci," *Frontiers in neuroscience*, vol. 12, p. 308, 2018.

[184] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, "Eeg-gan: Generative adversarial networks for electroencephalograhic (eeg) brain signals," *arXiv preprint arXiv:1806.01875*, 2018.

[185] S. Roy, S. Dora, K. McCreadie, and G. Prasad, "Mieeg-gan: generating artificial motor imagery electroencephalography signals," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[186] G. Dai, J. Zhou, J. Huang, and N. Wang, "Hs-cnn: a cnn with hybrid convolution scale for eeg motor imagery classification," *Journal of neural engineering*, vol. 17, no. 1, p. 016025, 2020.

[187] J. Luo, Z. Feng, J. Zhang, and N. Lu, "Dynamic frequency feature selection based approach for classification of motor imageries," *Computers in biology and medicine*, vol. 75, pp. 45–53, 2016.

[188] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms," *IEEE Transactions on biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2010.

[189] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "A multi-class eeg-based bci classification using multivariate empirical mode decomposition based filtering and riemannian geometry," *Expert Systems with Applications*, vol. 95, pp. 201–211, 2018.

[190] X. Tang and X. Zhang, "Conditional adversarial domain adaptation neural network for motor imagery eeg decoding," *Entropy*, vol. 22, no. 1, p. 96, 2020.