Brain Activity Classification Using

Machine Learning



*By*

Muhammad Junaid

FA21-EEE-061

Umer Ali

FA21-BCE-008

Ahmed Raheel

FA21-BCE-015

BS Thesis

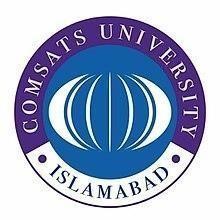
In

Electrical and Computer Engineering

COMSATS University Islamabad

Abbottabad Campus-Pakistan

Spring 2025



**COMSATS University Islamabad, Abbottabad Campus**

Brain Activity Classification Using

Machine Learning

A Thesis Presented to

COMSATS University Islamabad- Abbottabad Campus

In partial fulfillment

of the requirement for the degree of

BS Computer Engineering

By

Muhammad Junaid

CUI/FA21-EEE-061/ATD

Umer Ali

CUI/FA21-BCE-008/ATD

Ahmed Raheel

CUI/FA21-BCE-015/ATD

Spring 2025

Brain Activity Classification Using Machine Learning

An undergraduate thesis was submitted to the Department of Electrical and Computer Engineering as partial fulfillment of the requirements for the award Degree of BS in Computer Engineering.

|  |  |
| --- | --- |
| Name | Registration Number |
| Muhammad Junaid | CUI/FA21-EEE-061/ATD |
| Umer Ali | CUI/FA21-BCE-008/ATD |
| Ahmed Raheel | CUI/FA21-BCE-015/ATD |

Supervisor

Dr. Shoaib Azmat

Associate Professor, Electrical and Computer Engineering

Abbottabad Campus

COMSATS University Islamabad (CUI)

Abbottabad Campus

Co-Supervisor

Engr. Nauman Khan Tareen

Assistant Professor, Electrical and Computer Engineering

Abbottabad Campus

COMSATS University Islamabad (CUI)

Abbottabad Campus

Final Approval

This thesis titled

Brain Activity Classification Using Machine Learning*By*

*Muhammad Junaid*

*CUI/FA21-EEE-061/ATD*

*Umer Ali*

*CUI/FA21-BCE-008/ATD*

*Ahmed Raheel*

*CUI/FA21-BCE-015/ATD*

Has been approved.

For the COMSATS University Islamabad, Abbottabad Campus

Supervisor: Dr. Shoaib Azmat, Associate Professor

Department of Electrical and Computer Engineering Engineering/CUI, Abbottabad Campus

Co-Supervisor: Engr. Nauman Khan Tareen, Assistant Professor

Department of Electrical and Computer Engineering Engineering/CUI,

Abbottabad Campus

HOD: Dr. *Zahid* Mahmood *Jahangiri*, Associate Professor

Department of Electrical and Computer Engineering Engineering/CUI,

Abbottabad Campus

**Declaration**

We Muhammad Junaid **CUI/FA21-EEE-061/ATD**, Umer Ali **CUI/FA21-BCE-008/ATD**, and Ahmed Raheel **CUI/FA21-BCE-015/ATD** hereby declare that we have produced the work presented in this thesis, during the scheduled period of study. We also declare that we have not taken any material from any source except referred to wherever due that amount of plagiarism is within acceptable range. If a violation of HEC rules on research has occurred in this thesis, **we** shall be liable to punishable action under the plagiarism rules of the HEC.

|  |  |
| --- | --- |
| Date: **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | Signature of the student: |
|  | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
|  | Muhammad Junaid  CUI/FA21-EEE-061/ATD |
|  | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Umer Ali  CUI/FA21-BCE-008/ATD  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Ahmed Raheel  CUI/FA21-BCE-015/ATD |

**Certificate**

It is certified that Muhammad Junaid **CUI/FA21-EEE-061/ATD, Umer Ali CUI/FA21-BCE-008/ATD, and Ahmed Raheel CUI/FA21-BCE-015/ATD** have carried out all the work related to this thesis under my supervision at the Department of Electrical and Computer Engineering, COMSATS University Islamabad, Abbottabad campus and the work fulfills the requirement for award of BS degree.

Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Supervisor:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dr. Shoaib Azmat

Associate Professor

Department of Electrical and Computer Engineering

Comsats University Islamabad,

Abbottabad Campus

Head of Department:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dr. Zahid Mahmood Jahangiri,

Associate Professor

Department Computer Engineering

Comsats University Islamabad,

Abbottabad Campus

**DEDICATION**

We dedicate this study first and foremost to Allah, the Almighty for giving us strength, health, and guidance.

To our respected parents and family for having trust in us.

To our institute Comsats University Islamabad, Abbottabad campus.

And finally, to our respected supervisor and co-supervisor for guiding, leading and mentoring us.

**ACKNOWLEDGEMENTS**

We would like to extend our heartfelt appreciation to our esteemed supervisor, **Dr. Shoaib Azmat**, whose insightful comments, constructive remarks, and active engagement have played a pivotal role in enhancing the quality of our work. We would also like to recognize our co-supervisor **Engr. Nauman Khan Tareen** and **Dr. Muhammad Imran Shehzad** for their invaluable guidance and unwavering support throughout the completion of this bachelor's thesis.

Their support, guidance and encouragement have played a major role in our understanding of the subject matter and elevated the quality of our research. The wealth of knowledge and expertise they shared with us has not only been instrumental in our academic achievements but has also fostered a conducive learning environment.

We would like to express our sincere gratitude to Mr. Muhammad Waheed for graciously allowing us to acquire EEG data in his lab, the NCS and EEG Center. His support and provision of resources played a crucial role in the successful completion of our research.

**Muhammad Junaid CUI/FA21-EEE-061/ATD**

**Umer Ali CUI/FA21-BCE-008/ATD**

**Ahmed Raheel CUI/FA21-BCE-015/ATD**

**ABSTRACT**

Brain Activity Classification Using Machine Learning

Machine learning-based brain activity classification is a crucial part of an active and evolving research area in biomedical signal processing known as brain-computer interface (BCI). The concept is to read the brain signal, that is, the electroencephalogram (EEG) and interpret it to identify cognitive states or intentions. The EEG data used in this study was obtained on an open-source dataset and real-time recording via the Contec KT88-2400 device in different cognitive and visual tasks conditions. The signals were segmented and processed to extract statistical features after the removal of artifacts with Independent Component Analysis (ICA). A number of machine learning models were trained and tested, such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest. The Random Forest model was chosen as the one that strikes the right balance between accuracy and efficiency and was implemented on an Odroid XU4 embedded device with a custom GUI to predict the brain state in real-time. The applications of this work are varied especially in the medical field. The ultimate objective of this project was to come up with a well-implemented classification method on an embedded edge device that would be able to identify brain activity.

**TABLE OF CONTENTS**

[Chapter 1 1](#_Toc202128765)

[1.1 Introduction 2](#_Toc202128766)

[1.2 Project Background 2](#_Toc202128767)

[1.3 A BCI System 3](#_Toc202128768)

[1.4 AI, ML, and the Medical Field 3](#_Toc202128769)

[1.5 Motivation 4](#_Toc202128770)

[1.6 Problem Statement 4](#_Toc202128771)

[1.7 Objectives 5](#_Toc202128772)

[1.8 Scope of the Project 5](#_Toc202128773)

[Chapter 2 6](#_Toc202128774)

[2.1 Introduction 7](#_Toc202128775)

[2.2 Brain Function and Signals 7](#_Toc202128776)

[2.3 Brain Signals and Signal Acquisition Methods 8](#_Toc202128777)

[2.3.1 Brain Signals 8](#_Toc202128778)

[2.3.2 Signal Acquisition Methods 8](#_Toc202128779)

[2.4 Electroencephalography (EEG) 9](#_Toc202128780)

[2.4.1 Overview of EEG 9](#_Toc202128781)

[2.4.2 Brain Regions and EEG Signals 10](#_Toc202128782)

[2.4.3 10-20 Electrode Placement System 11](#_Toc202128783)

[2.4.4 EEG Frequency Bands 12](#_Toc202128784)

[2.4.5 EEG Pre-processing Methods 13](#_Toc202128785)

[2.5 Mental Activity and EEG 14](#_Toc202128786)

[2.5.1 EEG Variations in Eye States 14](#_Toc202128787)

[2.5.2 EEG Variation in Brain Arithmetic and Relaxation Conditions 15](#_Toc202128788)

[2.6 Brain-Computer Interfaces (BCI) 15](#_Toc202128789)

[2.6.1 Overview of BCIs 15](#_Toc202128790)

[2.6.2 Advantages of EEG for BCIs 16](#_Toc202128791)

[2.7 Related Work 16](#_Toc202128792)

[2.8 Conclusion 19](#_Toc202128793)

[Chapter 3 21](#_Toc202128794)

[3.1 Block Diagram 22](#_Toc202128795)

[3.2 Data Acquisition Procedure 23](#_Toc202128796)

[3.2.1 Participants 23](#_Toc202128797)

[3.2.2 EEG Recording Device and Software 24](#_Toc202128798)

[3.2.3 Electrode Placement 24](#_Toc202128799)

[3.2.4 Experimental Environment and Protocol 25](#_Toc202128800)

[3.2.5 Visualization of Acquired EEG Data 26](#_Toc202128801)

[3.2.6 Data Labeling 27](#_Toc202128802)

[3.3 Data Preprocessing 27](#_Toc202128803)

[3.4 Feature Extraction 27](#_Toc202128804)

[3.4.1 Statistical Features 27](#_Toc202128805)

[3.4.2 Frequency-Domain Features 29](#_Toc202128806)

[3.5 Classification Models 29](#_Toc202128807)

[3.5.1 K-Nearest Neighbors (KNN) 30](#_Toc202128808)

[3.5.2 Decision Tree 31](#_Toc202128809)

[3.5.3 Support Vector Machine (SVM) 31](#_Toc202128810)

[3.5.4 Random Forest 32](#_Toc202128811)

[3.6 Model Training 33](#_Toc202128812)

[3.7 Summary 33](#_Toc202128813)

[Chapter 4 35](#_Toc202128814)

[4.1 Data Segmentation Techniques 36](#_Toc202128815)

[4.2 Comparison with online dataset 39](#_Toc202128816)

[4.3 ICA on Acquired Dataset 40](#_Toc202128817)

[4.4 Discussion 41](#_Toc202128818)

[4.5 Model Complexities 42](#_Toc202128819)

[4.6 Hardware Deployment 43](#_Toc202128820)

[Chapter 5 45](#_Toc202128821)

[References 47](#_Toc202128822)

**LIST OF FIGURES**

Figure 1.1: Block Diagram of BCI System [1] 3

Figure 2.1: EEG Recording Process [2] 10

Figure 2.2: Lobes of Brain [3] 11

Figure 2.3 10-20 Electrode System [4*]* 12

Figure 2.4: Wavelets for EEG Analysis [5] 13

Figure 3.1: Block Diagram 22

Figure 3.2: Subject lying supine during EEG acquisition. 23

Figure 3.3: Contec KT88-2400 EEG acquisition system [19] 24

Figure 3.4: Electrodes positioning for EEG signal [20] 24

Figure 3.5: Raw EEG recording 26

Figure 3.6: EEG data in CSV form 26

Figure 3.7 KNN model architecture [21] 30

Figure 3.8 : Decision Tree model architecture [22] 31

Figure 3.9 : Support Vector Machine (SVM) for linear decision boundary [23] 31

Figure 3.10: Support Vector Machine (SVM) for nonlinear decision boundary [24] 32

Figure 3.11 : Random Forest model architecture [25] 32

Figure 4.1 : Model Accuracy Across EEG Window Sizes and Overlaps (Visual Class) 38

Figure 4.2: Results after ICA for Visual, Mental, and Multiclass Tasks 40

Figure 4.3: Execution time comparison 42

Figure 4.4: Graphical user interface (GUI) developed for Odroid XU4 deployment 44

**LIST OF TABLES**

[Table 2.1 Summary of Neuroimaging Methods [11] 9](#_Toc202129760)

[Table 2.2 Comparison of EEG Pre-processing Methods 14](#_Toc202129761)

[Table 4.1Classification Results using Best Data Segmentation Technique 39](#_Toc202129762)

[Table 4.2 Result Comparison between Online and Custom Dataset 40](#_Toc202129763)

**ABBREVIATIONS**

|  |  |
| --- | --- |
| EEG | Electroencephalogram |
| AI | Artificial Intelligence |
| ML | Machine Learning |
| DWT | Discrete Wavelet Transform |
| EMD | Empirical Mode Decomposition |
| MLP | Multilayer Perceptron |
| SVM | Support Vector Machine |
| CNN | Convolutional Neural Networks |
| ANN  ICA  CSV | Artificial Neural Networks   |  |  | | --- | --- | | Independent Component Analysis  Comma-Separated Values | | |  | |  |  |  | | --- | |  | |

# Chapter 1

**Introduction**

## Introduction

Human brain being the control center of the body controls thoughts, feelings and actions. Neuroscience has long been interested in the understanding of brain activity and was able to make progress in the diagnosis and treatment of neurological disorders. A non-invasive technique of recording electrical activity of the brain, known as electroencephalography (EEG), has become the basis of brain monitoring and analysis.

The increasing use of artificial intelligence (AI) and machine learning (ML) in medical technologies transformed how the physiological data, such as EEG signals, are processed and interpreted. Using these tools, one can categorize brain and eye processes that can help to understand mental and physical conditions and also introduce new inventions in Brain-Computer Interface (BCI) systems. The proposed project is the study of the classification of the EEG data in order to determine not only the brain state (mental relaxation and mental arithmetic) but also the state of the eyes (open and closed) and broadens the field of knowledge and use of the EEG analysis.

## Project Background

EEG classification using machine learning is an important part of a developing field of biomedical signal processing known as Brain-Computer Interface (BCI). The principle is to capture the brain signals, process and interpret their meaning.

This research is aimed to identify brain and eye status with the help of EEG signals. It begins with obtaining EEG data on human subjects in a controlled condition where they engage in the act of mental relaxation and mental arithmetic and interchanging the condition of keeping the eyes open and the eyes closed. The signals are used to extract features that are then input into a machine learning classifier that is trained to classify between these states. Once trained, the system categorizes new signals, and determines the activity associated with it. It involves the analysis of different features and classifiers in order to optimize the solution. Moreover, the project focuses on effective implementation, and the final aim will be to make the system work on an embedded edge device to recognize the brain and eye states in real-time.

## A BCI System

A Brain-Computer Interface (BCI) is a communication channel that allows the brain to communicate with external devices, through interpretation of neural signals. BCIs can be used in medical, educational, and gaming sectors among others. BCIs are applied in medical environments to help persons with motor impairments, to track cognitive activities, and to create neurorehabilitation systems.

In this project, EEG signals are recorded by the BCI system during tasks which require different brain and eye conditions. The system aims at categorizing these states in order to improve the knowledge of the cognitive processes and physical conditions. The uses are mental health monitoring systems, cognitive workload systems, adaptive learning systems, and assistive technologies to support those with motor impairments.

A diagram of a person's head

Description automatically generated

Figure .: Block Diagram of BCI System [1]

## AI, ML, and the Medical Field

Machine learning and artificial intelligence have revolutionized the healthcare sector and allowed the use of data-based methods in diagnostics and treatment. In the application to EEG signal analysis, ML algorithms help to detect the pattern and successfully categorize the brain and eye states. The technologies minimize the amount of manual work involved in the analysis and allow tracking in real-time.

This project uses machine learning models to study EEG data. The system is trained by having the models learn on data recorded on subjects carrying out certain tasks and hence the model learns to distinguish states like mental relaxation, mental arithmetic, eyes open, and eyes closed. The combination of ML shows its prospects to improve neurological studies, patient management, and individualized healthcare systems. The use of these models on edge devices also highlights their use in portable, real-time applications.

Through EEG, a number of new applications have emerged to be utilized in the real life situation like the neuro feedback systems. This new procedure can be used to treat different ailments like Attention Deficit Hyperactivity Disorder (ADHD) and autism. [10]

## Motivation

The need to work on this project is associated with the need to solve problems of people with motor impairment and healthcare professionals. The patient who is paralyzed usually needs a hands-free system to communicate with the machine and the healthcare system needs effective monitoring systems. A user-friendly automated system capable of categorizing brain and eye conditions based on EEG signals has the potential to change the face of assistive technologies, making patients more independent and lessening the burden of caregivers.

Moreover, the project will be able to unleash the potential of portable EEG-based systems in a wide variety of domains including stress management, adaptive learning environments, and cognitive workload monitoring. This work aims at offering cost-effective and scalable solutions to personalized healthcare, mental health assessment, among other applications through real-time classification of states on edge devices.

## Problem Statement

The EEG signals are complex and variable making it very difficult to classify the states of the brain and eye. The existing techniques have a high computational burden, which makes them impractical to use on edge devices in real-time. Further, the lack of automated and portable systems poses a problem of access to persons with disabilities and overwhelmed healthcare providers.

This project is dealing with such challenges in the following way:

* Designing an effective mechanism of collecting and processing EEG signals when the brain and the eyes are engaged in specific activities.
* The use of machine learning in labeling brain and eye states with high accuracy.
* Use of system on an embedded edge device in real-time application.

This project aims to improve EEG monitoring with the help of a simple, automated system that can control a device without the need to use hands and reduce the burden on healthcare workers by allowing paralyzed patients to manage a device without assistance.

## Objectives

This project will have the following objectives:

* To categorize EEG data into particular states of the brain and eyes by machine learning models.
* To use the trained model on edge device to perform real-time analysis.
* To determine the accuracy and efficiency of the system to distinguish between cognitive and physical states.
* To help develop assistive technologies in people with motor impairments.

## Scope of the Project

The aim of this project is to classify EEG data of ten subjects who perform four activities that have been predetermined, including mental relaxation with eyes open, mental relaxation with eyes closed, mental arithmetic with eyes open, and mental arithmetic with eyes closed. The machine learning models are made lightweight and highly efficient, and thus they are compatible with edge devices. Examples are the medical technologies like assistive technology to paralyzed patients and automated monitoring of mental and physical conditions.

# Chapter 2

**Literature Review**

## Introduction

The brain of a human being is a complex organ that regulates body functions, thinking and feelings. The electrical activity of the brain is a measurable and analyzable phenomenon that occurs due to the ionic current flows in the neurons and allows one to get an insight into brain functioning. Non-invasive electroencephalography (EEG) is a popular brain signal investigation tool because it has the advantage of measuring the electrical activity of the brain on a real-time basis.

Recent developments in machine learning have led to classification of particular brain states (e.g. arithmetic and relaxation) and body states (e.g. open and closed eyes) using EEG measurements. These applications have broad implications to neuroscience, medicine and human-computer interaction. EEG-based systems specifically have been of great use in cognitive workload measurement, early detection of neurological conditions, and interaction models of assistive technology.

This chapter gives a detailed literature review starting with the way the brain works, the methods of signal acquisition and the importance of EEG in Brain-Computer Interface (BCI) systems. The talk also gets specific to EEG signal properties, electrode position, bands and patterns in certain brain and eye states.

## Brain Function and Signals

The brain is the centre of control of the body, it contains billions of neurons all connected together and form a complex network. The network enables the production and transmission of electrical signals that enable everything including movement and cognition. The basic components of the brain and the nervous system, neurons, produce these signals by communicating with each other chemically and electrically, creating rhythmic patterns called the brainwaves.

The electrical signaling between neurons is done by means of action potentials. During an action potential, the ions such as sodium and potassium move through the membranes of the nerve cells, creating the voltages that can be measured on the scalp.

Such rhythmic patterns of coordinated neuronal activity can be recorded and observed in order to learn about various states of the brain. This kind of activity does not only provide support to the higher levels of cognition, but also controls the involuntary processes like breathing and heart rate. Populations of neurons in brain regions send coordinated signals to make recognizable patterns that occur in association with functional states.

These patterns have been comprehended and have resulted in diagnostic and research instruments that allow neuroscientists to monitor neural behavior through time.

## Brain Signals and Signal Acquisition Methods

### Brain Signals

Neuronal activity is in the form of brain signals that are electrochemical in nature. They are important determinants of motor control, cognition and states of mind. Produced by the neural networks interaction, rhythmic oscillations are divided into several brainwave frequency bands, such as delta, theta, alpha, beta, and gamma waves. These oscillations act as biomarkers of many physical and psychological conditions, ranging between relaxation and mental load.

Brain signal studying technologies are:

* EEG (Electroencephalography): Non-invasive, electrical changes on the scalp in terms of voltage fluctuations due to ionic flows.
* fMRI (Functional Magnetic Resonance Imaging): Offers spatial detail of the activity of the brain through monitoring blood flow but poor time resolution.
* MEG (Magnetoencephalography): It records magnetic signals caused by the electrical activity of neurons.

EEG is one of them and has now emerged as a powerful method of monitoring the moment-to-moment changes in the brain activity because of its low cost, ease of operation, and online monitoring properties.

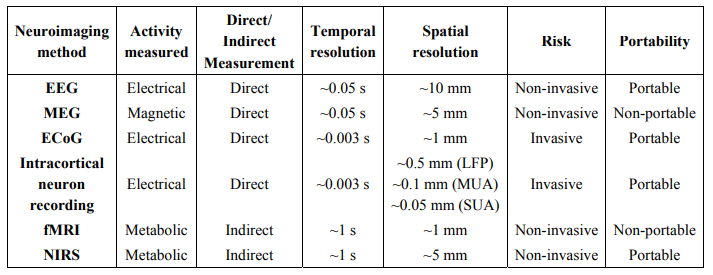
### Signal Acquisition Methods

Processing and classification are based on the signal acquisition stage. It is reliable in that it tries to reduce noise and collect quality data. EEG acquisition procedures fall into the following categories:

* Invasive Techniques:
* Electrocorticography (ECoG): Surgically implanted electrodes are used to supply high-resolution data. It is invasive, which restricts its application to clinical practice. [11]
* Deep Brain Stimulation: It is mostly applied in research or treatment of neurological disorders such as Parkinson disease. [11]
* Non-Invasive Techniques:
* EEG: EEG is transportable and affordable; it records widespread cortical activity through electrodes that are attached to the scalp. Electrode caps and dry electrodes are being advanced to be more efficient and easy. [11]
* MEG: Good temporal resolution at the expense of being much more costly and cumbersome than EEG. [11]
* fNIRS (Functional Near-Infrared Spectroscopy): It is a light-based measurement of changes in oxygenation in the brain, which supplements the shortcomings of EEG. [11]

The method of acquisition is application-specific; EEG is used where a quick and low-cost system is desired, such as in Brain-Computer Interfaces. Table-2.1 gives the EEG recording methods.

Table . Summary of Neuroimaging Methods [11]



## Electroencephalography (EEG)

### Overview of EEG

Electroencephalography (EEG) is the recording and detection of electrical activity in the brain by use of electrodes placed on the scalp. Voltage variations measured in EEG are the aggregate synaptic activity of thousands of neurons. EEG is also temporally very high resolving (milliseconds), which is invaluable in dynamic monitoring of brain-states. Moreover, EEG is non-invasive, which increases its usage in health surveillance, BCI applications, and cognitive science studies.

EEG signals are taken in various channels as each channel is related to a given location on the scalp. The characteristics based on EEG, either in the time-domain (e.g. average amplitude, variance) or frequency-domain (e.g. power spectral density) are the basis of differentiating various brain states. It is important to manage artifacts, because muscle activity, environmental interference, or electrode failure can corrupt signals. The process of EEG recording is displayed in Figure-2.1.

A diagram of a person's brain

Description automatically generated

Figure .: EEG Recording Process [2]

### Brain Regions and EEG Signals

EEG records activity that is generated in different parts of the brain and this is divided into four major lobes:

* Frontal Lobe: Controls executive functions, emotion and voluntary movements and reasoning. There is an association between complex mental functioning and enhanced beta and gamma activity in this area.
* Parietal Lobe: Combines sensory data and plays an important role in proprioception and visuospatial logic.
* Temporal Lobe: Memory, control of emotions and hearing. Long-term memory recall has been attributed to high theta activity.
* Occipital Lobe: It deals with visual processing of information and has a high alpha activity that shows a relaxed state and less visual attention.

The functional separation of this allows regional EEG analysis, where classification is concentrated on areas of interest in the task. The details of lobes of

brain are depicted in Figure-2.2.

A diagram of the brain

Description automatically generated

Figure .: Lobes of Brain [3]

### 10-20 Electrode Placement System

One of the most important points in EEG acquisition is the International 10-20 System, which is intended to give a standardized electrode positioning, Figure-2.3 depicts the detailed 10-20 electrode placement system. The system used to name electrodes is founded on underlying brain areas and lateralities. Examples include:

* F3, F4: Signalizing front lobe activities on the left and right hemispheres.
* T7, T8: bilateral capture of temporal activities.

Electrode placement would allow the same data to be collected in different people and so would allow the reproducibility of studies and strong comparisons in BCIs.

A diagram of the brain

Description automatically generated

Figure . 10-20 Electrode System [4]

### EEG Frequency Bands

EEG frequencies indicate certain mind and body conditions:

* Delta Waves: This occurs when one is in deep sleep and unconscious. [6] [11]
* Theta Waves: Detected when one is meditating and in creative activities. [6] [11]
* Alpha Waves: This type is displayed in relaxation and it is predominant in the occipital recordings when the eyes are closed. [6] [11]
* Beta Waves: Show mental alertness and analytical thinking, which are eminent during arithmetic tasks. [6] [11]
* Gamma Waves: Indicates high-order thinking e.g. problem-solving and perception. [6] [11]

These frequencies and topographic distributions provide information about the dynamics of the brain, helping with tasks such as event detection, focus analysis and classification of particular states. Figure-2.4 indicates the EEG wavelets.

A close-up of several blue waves

Description automatically generated

Figure .: Wavelets for EEG Analysis [5]

### EEG Pre-processing Methods

EEG signal analysis Pre-processing is a necessary step in the analysis to eliminate noise and artifacts. It makes the signals clean and prepares them to be processed further. The most important pre-processing techniques are briefly described below:

**1. Artifact Subtraction (AS)**

Artifact Subtraction estimates and removes noise (e.g. eye movements) by supposing linearity between signals and artifacts. Though easy to implement and interpret, it can unwillingly eliminate actual EEG components. [6]

**2. Principal Component Analysis (PCA)**

PCA is computationally efficient in the sense that it diminishes the dimension by isolating noise components. But it suffers noise which covers the signal spectrum. [6]

**3. Blind Source Separation (BSS)**

BSS, and, in particular, Independent Component Analysis (ICA), decomposes signals into independent sources. It is very precise and flexible yet computationally intensive and independence-based. [6]

**4. Wavelet Transform (WT)**

Wavelet Transform is a time-frequency decomposition of signals; it can deal with non-stationary signals. But it needs the careful choice of the wavelet basis. [6]

**5. Spatial and Adaptive Filtering**

The techniques enhance signal-to-noise ratio by eliminating the noise sources of interest. They are flexible and can be applied in real time at the expense of intensive familiarity with the nature of noise. [6]

**Comparison of Pre-Processing Methods**

The comparison of EEG pre-processinng methods is summarized in Table-2.2

Table . Comparison of EEG Pre-processing Methods

|  |  |  |
| --- | --- | --- |
| Method | Advantages | Disadvantages |
| Artifact Subtraction (AS) | Simple, interpretable | May lose genuine EEG components |
| Principal Component Analysis (PCA) | Computationally efficient | Limited for overlapping spectral noise |
| Blind Source Separation (BSS) | High accuracy, versatile | Computationally intensive, assumes independence |
| Wavelet Transform (WT) | Effective for non-stationary signals | Requires careful wavelet selection |
| Spatial and Adaptive Filtering | Versatile, real-time capable | Requires detailed noise knowledge |

## Mental Activity and EEG

### EEG Variations in Eye States

EEG records different variations in open and closed-eye conditions:

* Closed Eyes:

This is linked to an increase of the alpha activity and largely in the occipital regions. Denotes states of relaxation [7].

* Open Eyes:

This is reduced in alpha amplitude and elevated in gamma as a result of increased involvement. [7]Opening eyes decreases theta activity all over [8].

These variation become valid inputs to models that categorize states of physical systems. Such systems as driver alertness systems, for example, use this to anticipate the onset of sleep.

### EEG Variation in Brain Arithmetic and Relaxation Conditions

EEG changes are obtained by performing different mental tasks:

* Arithmetic Tasks: They involve increased beta and gamma in the frontal electrodes. Heightened coherence between the frontal-parietal regions is a sign of inter-regional coordination. The frontal, occipital and temporal lobes are more involved in the execution of mental arithmetic task [10].
* Relaxation: This is associated with dominant alpha waves particularly the posterior region, which shows less mental workload. Relaxation is associated with alpha waves and theta waves are observed when a person is asleep. [9]

Studies of EEG variations that are state-specific are still a source of development of adaptive BCIs, cognitive tests, and workload monitoring systems.

## Brain-Computer Interfaces (BCI)

### Overview of BCIs

Brain-Computer Interfaces (BCIs) are devices which allow direct communication between the brain and other external devices by skipping the normal neural pathways. These systems convert brain activity into computer-operating commands that can operate a computer, a prosthetic, or other machines, providing radical potentials in medical care, recreation, and other areas. BCIs are especially effective in people with motor impairments as they offer a new method of communication and interaction with the world.

BCIs use brain signals, which they capture, process to identify and utilize meaningful patterns and utilize the patterns to run external applications. A BCI is highly dependent on the recording method of the brain activity, and a wider range of methods exist, including invasive (such as intracortical electrodes) to non-invasive (such as EEG).

**2.5.2 Role of EEG in BCIs**

One of the most popular methods to develop a BCI is electroencephalography (EEG) because it is non-invasive, cheap, and has a high temporal resolution. EEG entails fixing electrodes on the scalp to detect electrical activity produced by the firing of the neurons in the brain. These signals are measured as oscillatory signals or waveforms that can be analyzed to identify particular patterns that are related to the motor intentions, sensory stimuli or cognitive state.

### Advantages of EEG for BCIs

EEG happens to be the most appropriate method of BCI because of the following reasons:

1. **Non-Invasive Recording:** EEG does not involve any surgical procedures hence it is safer and more convenient to the user. [11]
2. **High Temporal Resolution:** EEG measurements are real-time making them essential in responsive BCI applications.
3. **Cost-Effectiveness:** EEG systems are relatively cheap and easy to implement as compared to invasive methods.
4. **Portability:** The EEG technology has also improved such that there are portable systems, which are allowing the use of BCIs not only in laboratories.

## Related Work

The development of EEG in preprocessing, feature extraction and classification has made it one of the dominant modalities in neurology and machine learning. The innovations are aimed at enhancing accuracy, adaptive learning, and cross-subject performance, which is crucial to real-life implementations. This part summarizes the research that has been done in the given studies, and it focuses on the progress in EEG signal processing in different applications.

**1. Recognition of Eye Movements Using Brain Computer Interface and Random Forests Based on EEG-Based [12]**

This paper has designed braincomputer interface (BCI) system to recognize six eye movements: left, right, open, closed, up, and down. EEG was measured in the Emotiv EPOC flex device with 32 saline electrodes arranged in the international 1020 system. The sampling frequency of signals was 128 Hz. Power spectral density analysis based on particular frequency bands was used to extract features. Random forest (RF) classifier performed better than the rest of the models with an accuracy of 85.39%, which proves its applicability in multi-class EEG classification in assistive technologies. The narrowness of the sample can be viewed as a limitation of the proposed study. Although a small sample can be considered, the suggested approach demonstrates a satisfactory classification quality with more than 85 percent of the accuracy of a multiclass classification task

**2. Mental Stress Quantification Using EEG Signals [9]**

This study examined the detection of stress based on the EEG signal recorded in 20 participants that performed arithmetic tasks of different difficulty. BrainMaster 24E system and seven active electrodes [FP1, F3, F7, Fz, FP2, F4 and F8] and one reference A1 were placed on the earlobes at 256 Hz sampling frequency to record data on the frontal cortex. EEG was decomposed using wavelet transform. Mean of absolute value of wavelet coefficients, average power and energy were extracted and used as features. The classification was done using SVM classifier and 10-fold cross-validation. This system recorded accuracies of 94%, 85% and 80% of low, medium and high stress levels respectively illustrating the potential of EEG in mental stress measurement.

**3. Classification of Relaxation and Concentration Mental States with EEG [13]**

The accuracy of the classification of relaxed and concentrated states of mind is researched in the paper.The experiments are performed in two repetitions, involving 7 and 10 participants.The EEG device used in the experiments has a headband that will allow attaching two sensors to the FP1 and FP2 points. Mean absolute amplitude and band power were extracted. Using SVM and BPNN classifiers, the study achieved accuracies of 79.3 % for SVM and 74.7 for BPNN for the metric α + β + γ, 4 Hz.

**4. EEG Signal Classification for Mental Stress During Arithmetic Tasks Using Wavelet Transform and Statistical Features [14]**

The present research was an extension of the stress detection research, which compared the statistical features including mean, variance, and skewness with the wavelet-based features. EEG data is taken of physionet ATM. The research analyzes EEG of 11 subjects in their occipital (O2) part of the brain. Statistical features were extracted by using DWT. NCA model was effective in the selection of the dominant four features in order to maximize classification accuracy. The proposed technique produced the best results for the Ensemble Subspace KNN classifier, with accuracy of 77.3%, sensitivity of 91% and specificity of 64%.

**5. EEG Based Mental Arithmetic Task Classification Using a Stacked Long Short Term Memory Network for Brain-Computer Interfacing [10]**

TThe research paper has adopted a deep learning methodology based on stacked long short-term memory (LSTM) networks in mental arithmetic classification. The PhysioNet database was used as source of EEG signals, which comprised 36 participants and 22 channels. The interconnected ear reference electrodes were allotted to all the electrodes. Each channel had a sample rate of 500 Hz. To increase the data, the window-based segmentation was used. The LSTM network learned features automatically and displayed an accuracy of 91.67% indicating its effectiveness on sequential EEG data.

**6. Comparison of SVM and ANN for Classification of Eye Events in EEG [15]**

The current paper compared the support vector machines (SVM) and the artificial neural networks (ANN) in classifying eye events like blinks, closures, and movements in the EEG signal is recorded using Biopac MP-36 system, and the two electrodes were placed in FP1 and F3 region The sampling frequency is configured to 200 samples per second. Such features were kurtosis and spectral entropy. Although SVM had better specificity, ANN performed better in generalization. The SVM and ANN gave a maximum of 90.8% and 86.8% classification accuracy respectively.

**7. Classification of Mental Arithmetic and Resting-State Based on Ear-EEG [17]**

The paper has discussed the application of ear-EEG as a small version of the scalp EEG in the classification of mental arithmetic and resting states. EEG recording was done with 7 healthy participants, 31 electrodes were used referenced to FCz and the sampling rate was set at 1000 Hz. The data of EEG was analyzed by dividing the entire brain region into four regions of interest (frontal, central, occipital, and ear area) and comparing their EEG characteristics and the performance of classification. Each region of interest was analyzed on event-related (de)synchronization in delta and alpha bands. The achieved classification accuracies of the whole scalp, frontal, central, occipital and ear area are 88.9%, 72.6%, 76.7%, 82.6% and 75.6 %, respectively demonstrating the feasibility of ear-EEG for portable BCIs.

**8. EEG Mental Arithmetic Task Levels Classification Using Machine Learning and Deep Learning Algorithms [16]**

This paper was a comparison of machine learning and deep learning algorithms in the classification of the level of difficulty of tasks in mental arithmetic. This research relied on the publicly available EEG data, the recordings were taken on 36 subjects using Neurocom EEG 23-channel device at a frequency of 500 Hz. The characteristics applied are skewness, kurtosis, mean, standard deviation and variance. The accuracies achieved on machine learning classifier are KNN-91%, SVM-89% and Decision Tree-65%. The acquired accuracy of two deep learning models is ANN-96.80% and LSTM-94%

**9. Real-Time Mental Arithmetic Task Recognition From EEG Signals [18]**

In this paper, an experiment involving ten participants (aged 22-30) in performing relaxation and mental arithmetic and EEG signals recorded with a 14-channel Emotiv device at 128 Hz is mentioned. Examples of features extracted are the power spectrum density (PSD), autoregressive model coefficients (AR), statistical measures, and a new fractal analysis technique, the Generalized Higuchi Fractal Dimension Spectrum (GHFDS). A dimensionality reduction was done using Principal Component Analysis (PCA). The Support Vector Machine (SVM) with RBF kernel was used to perform classification. The system has a classification accuracy of 97.87% when all 14 channels are used and 97.11% when only four channels (F8, F3, AF3, O2) are used proving the efficacy of the combined feature method and the possibility of practical and portable neurofeedback systems.

## Conclusion

This thesis has covered how machine learning can be used to identify brain activities based on EEG signals which is an avenue of research on Brain-Computer Interfaces (BCIs). The project has shown the potential of EEG-based systems to be used in real-time by developing a system that could differentiate effectively between different mental and physical states, as well as eye states (open and closed).

The study emphasizes the significance of a good EEG preprocessing technique and the significance of feature extraction and classifier selection in the attainment of high accuracy in classification. Also, the fact that the trained models can be deployed on embedded edge devices highlights the viability of portable solutions to real-time neurofeedback systems. These developments open the door to more available technologies that are designed to suit a variety of needs, such as assistive technologies in the case of people with motor impairments and tools that help monitor cognitive workload.

Future improvements of this work can lead to the integration of larger datasets, the enhancement of the stability of the classifiers to different groups of people, and the investigation of other neural states. The combination of deep learning models and hybrid feature extraction techniques also pose attractive avenues to improve the performance of the classification and applicability in real-time.

To sum up, the paper confirms the compatibility of the EEG technology and machine learning in the development of innovations to fill the gap between neuroscience and real-life applications.

# Chapter 3

**Methodology**

## Block Diagram

A diagram of a process

AI-generated content may be incorrect.Figure 3.1 shows the block diagram that gives a broad overview of methodology followed in this study under EEG-based brain activity classification using machine learning. It presents the repetitive and consecutive steps which take part, including initial data collection, model deployment, and system optimization.

Figure .: Block Diagram

The experiment requires collection of data, which is achieved through obtaining EEG signals of the participants under the conditions specified beforehand in the experiment. Raw EEG data is the next subject to feature extraction under which statistical and frequency-based properties are extracted. These characteristics are used as inputs at the stage of developing the model, including training and testing a range of machine learning algorithms.

Testing is done to assess the performance in the real world after developing the model. Part of the user input and the results of the tests are used in the decision on whether the model fits the targets of the desired performance. In case of lack of conformance of the model, the pipeline returns to the stage of data collection to improve the model and retrain it.

After the model meets target specifications, hardware selection is followed and then the model is deployed. The deployed system must be tested at hardware level and any final corrections in the form of optimization and finalization stage should be done to make the system reliable and usable.

## Data Acquisition Procedure

### Participants

The EEG data is obtained using 10 healthy adult participants who gave their informed consent before the study commenced. The recording was done on an individual basis in a controlled laboratory setting with minimal interference by the environment.

All the participants were positioned in the supine position on a flat surface and instructed to recline comfortably, as this posture was chosen to minimize muscular tension and reduce artifacts due to head, neck, and spinal motion. The subjects were asked to sit still and relax during the recording process. Such precautions played a critical role in the stability of the signals and consistency between sessions.

Figure 3.2 illustrates the experimental setup for a single participant during EEG data collection.

A person with his eyes closed lying on a pillow with wires attached to his head

AI-generated content may be incorrect.

Figure .: Subject lying supine during EEG acquisition.

### EEG Recording Device and Software

EEG data were recorded through the Contec KT88-2400, a commercially available 24 channel EEG recording platform known to have clinical-level accuracy and signal integrity. Real-time signal visualization, recording management, and data storage were also possible at the same time using EEG24 proprietary software.

.

Figure .: Contec KT88-2400 EEG acquisition system [19]

### Electrode Placement

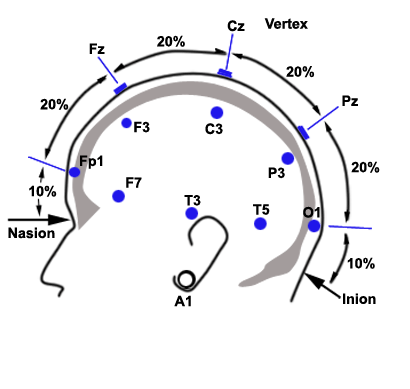
Electrode placement was based on the international 10-20 system, a universal standard of reference of the electroencephalography electrode positioning. This convention locates the electrodes with known cranial landmarks, thus enabling the standard and repeatable coverage of the scalp and related cortical areas.

Figure .: Electrodes positioning for EEG signal [20]

Twelve electrodes were selected to target brain areas involved in attention, visual processing, and cognitive load. The specific electrodes used included:

* Frontal: Fp1, Fp2, F3, F4, F7, F8
* Temporal: T3, T4, T5, T6
* Occipital: O1, O2

Electrode-to-skin conductivity was enhanced using ten20 conductive paste in the brain dynamics evaluation. The electrodes were pressed hard to create a good electrode-skin contact and to reduce noise artifacts. This standardization of ten20 paste on all ten participants made the data compare and consistent in data recordings of EEG.

### Experimental Environment and Protocol

At the recording stage, each subject was requested to perform four discrete mental states relaxed with the eyes open, relaxed with the eyes closed, mental arithmetic with the eyes open and mental arithmetic with the eyes closed. The conditions were 5 minutes each, and the total recording time per subject was 20 minutes. No comments or corrective information were given when carrying out the tasks to provide a sustained state of mind and a spontaneous cognitive reaction.

In the case of mental arithmetic condition, the participants were asked to perform subtraction continuously without using paper or any other external factor. They were told to take a four-digit number (which was randomly given to them, e.g., 7821) and to subtract a two-digit number (e.g., 37) therefrom, and then to repeat the procedure (e.g., 7821 37 = 7784, 7784 37 = 7747, etc.).

The reason this task was chosen is that it consistently can induce cognitive load, activation of working-memory, and maintenance of mental effort all of which are known to cause different and measurable patterns in EEG.

### Visualization of Acquired EEG Data

To have a better idea of the EEG data obtained during the experiment, this section contains two visual representations of the data. These visualizations indicate the temporal waveform structure of the signals as well as the data format used to store and process the data.

**A graph of a graph

AI-generated content may be incorrect.**Figure 3.5 represents raw EEG recordings of one of the subjects. The plot displays continuous signals that are recorded by twelve channels recorded from twelve channels, each corresponding to a specific electrode location

Figure .: Raw EEG recording

A table with numbers and letters

AI-generated content may be incorrect.Figure 3.6 is a snapshot of the EEG data in CSV form (Comma-Separated Values) which was the format used for preprocessing and feature extraction. In this representation, columns represent individual EEG channels and rows are time-sampled values measured at a constant sampling rate.

Figure .: EEG data in CSV form

### Data Labeling

Each EEG segment was manually labeled based on the task being performed by the participant during the recording. The EEG acquisition software produced separate data files corresponding to each of the four mental activities, each of which was assigned the label of the activity class.

## Data Preprocessing

To enhance signal quality further and to remove artifacts like eye blinks, power line noise and muscle movement, Independent Component Analysis (ICA) was used. ICA can break down multichannel EEG data into statistically independent components and thus allow them to detect and eliminate non-neural artifacts. This step significantly enhances the signal-to-noise ratio and ensures that subsequent analyses are based on physiologically meaningful signals.

The recorded signals were initially segmented into overlapping time windows of 5 seconds duration with an 80% overlap between successive segments. This segmentation strategy increases the number of training instances, thus improving statistical strength and time resolution.

## Feature Extraction

Feature extraction is a pivotal step in the transformation of raw EEG data into meaningful numerical representations that can be fed into machine learning models. We extracted both statistical features and frequency-domain characteristics.

### Statistical Features

Statistical measures were computed for each segmented EEG window to capture the basic amplitude and shape characteristics of the signal. The features included:

**1. Mean (μ)**

The mean represents the arithmetic average of all the signal values in a segment.

**2. Variance (σ²)**

The variance quantifies the overall spread of the signal values around the mean.

**3. Standard Deviation (σ)**

The standard deviation is the square root of the variance and provides a direct measure of the dispersion of the signal. It reflects how concentrated the signal values are around the mean.

**4. Skewness (γ₁)**

Skewness is a measure of asymmetry of the distribution of the signal. A skewness close to zero indicates a symmetric distribution whereas positive or negative values indicate a right or left skewed distribution respectively.

**5. Kurtosis (k)**

Kurtosis is a measure of the tailedness or peakedness of the distribution of the signal. A kurtosis value greater than one means that the distribution has heavy tails and sharp peaks whereas low kurtosis implies a flat distribution.

**6.Mean Absolute Deviation**

The mean absolute deviation (MAD) is another statistical measure used to quantify the dispersion of a signal. Unlike the standard deviation, which is based on squared deviations from the mean, MAD is calculated as the average of the absolute differences between each signal value and the mean. This makes MAD less sensitive to extreme outliers and provides an intuitive measure of how spread out the signal values are around the mean.

Mathematically, the mean absolute deviation is defined as:

### Frequency-Domain Features

To quantify the spectral content of EEG signals more robustly, the **Welch method** was employed for estimating the **power spectral density (PSD)**. This method is widely used in EEG analysis because it reduces the variance of the spectral estimate compared to the traditional periodogram by averaging modified periodograms obtained from overlapping segments of the signal.

In practice, the EEG signal was divided into overlapping segments, each windowed (commonly with a Hamming or Hann window) to reduce spectral leakage. The periodogram (the squared magnitude of the FFT) was calculated for each segment, and then these periodograms were averaged to produce the final PSD estimate.

Once the PSD was obtained, the **band power** for each of the standard EEG frequency bands (Delta, Theta, Alpha, Beta, and Gamma) was computed by integrating the PSD values over the corresponding frequency range:

The EEG signals were also decomposed into standard brainwave bands using digital bandpass filters. Spectral content of the signals was described by means of calculating the energy on each frequency band. The following bands were used:

* Delta (0.5–4 Hz)
* Theta (4–8 Hz)
* Alpha (8–13 Hz)
* Beta (13–30 Hz)
* Gamma (30–100 Hz)

These frequency bands are associated with various cognitive and neural activities. For instance, alpha waves are commonly observed during relaxed wakefulness, while gamma waves are linked to attention and cognitive load.

## Classification Models

The classification phase in this research was aimed at differentiating among four different states of mental and cognitive states based on supervised machine learning algorithms trained on the extracted EEG features. Four popular and popularly established types of classifiers were finally applied and compared in terms of performance. These classifiers, K-Nearest Neighbors (KNN), Decision Tree, Support Vector Machine (SVM), and Random Forest, have different complexities of the model and decision strategy, which, when combined with different constraints (computational and interpretability), allows an unbalanced evaluation. These models were chosen because of their effectiveness in the analysis of bio-signals, especially in the classification of cognitive state using EEG.

### K-Nearest Neighbors (KNN)

A diagram of a diagram of a diagram

AI-generated content may be incorrect.K-Nearest Neighbors algorithm is an instance-based, non-parametric learning algorithm, which classifies the data point depending on the majority label of its k nearest neighbors in the feature space. The Euclidean (or other) distance between the test point and each training set point is computed in the case of a given test point. The closest k samples are picked and the test instance is given the most prevalent class of these neighbors.

Figure . KNN model architecture [21]

KNN is simple to implement and interpretable, but it can be computationally expensive for large datasets, as it requires calculating distances for each new prediction.

### Decision Tree

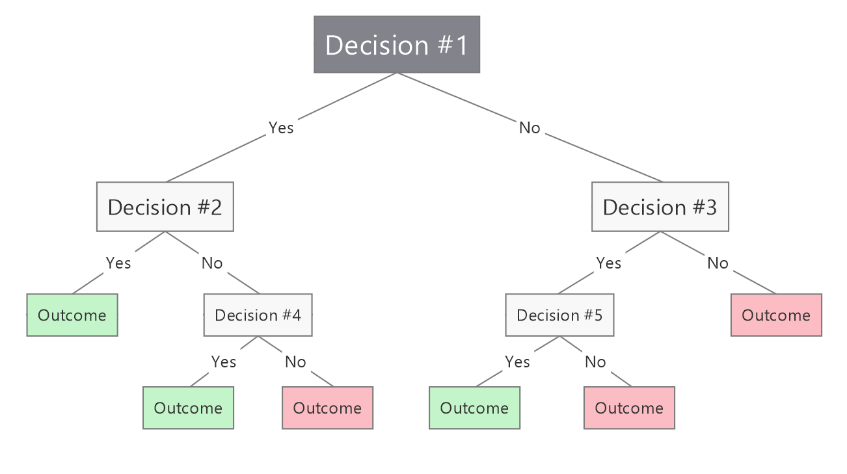
A Decision Tree is a tree type of classifier that recursively divides the feature space into mutually exclusive regions, using attributes that maximize information gain or minimize such measures of impurity as Gini Index or entropy. Internal nodes represents decisions formed on the value of an attribute whereas the leaf nodes are class labels.

Figure . : Decision Tree model architecture [22]

### Support Vector Machine (SVM)

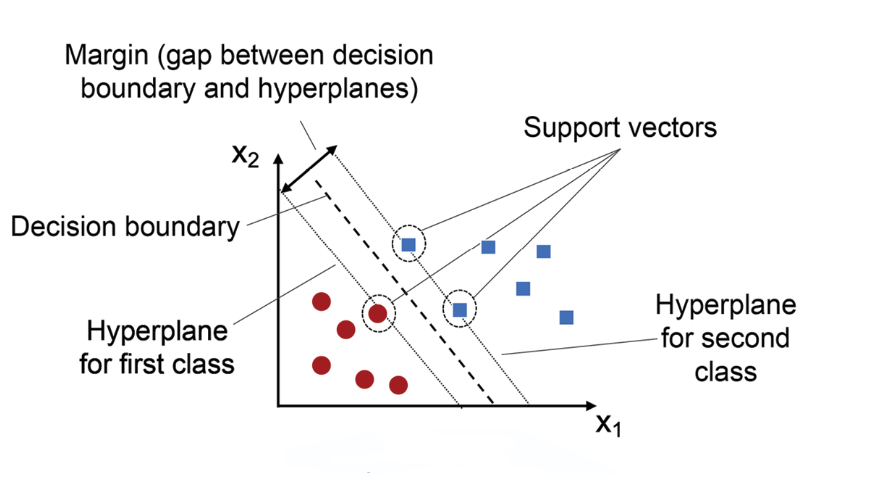
The Support Vector Machine (SVM) is a powerful supervised learning algorithm that aims to find an optimal hyperplane that maximizes the margin between data points belonging to different classes.

Figure . : Support Vector Machine (SVM) for linear decision boundary [23]

This hyperplane is the decision boundary of linearly separable data. In non-linear problems, SVM uses a trick called the kernel trick, which transforms the input features into a higher-dimensional space where the model seeks a separating hyperplane in the transformed space.

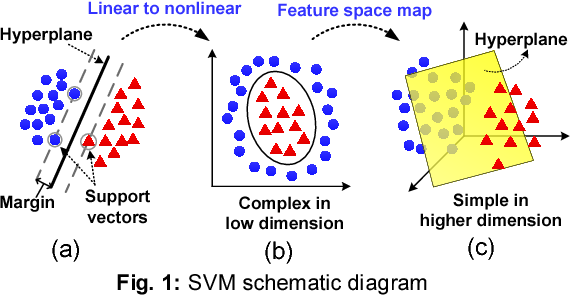


Figure .: Support Vector Machine (SVM) for nonlinear decision boundary [24]

### Random Forest

A diagram of a tree

AI-generated content may be incorrect.Random Forest is an ensemble classification method that constructs several decision trees in the training process and combines their results to enhance generalization and minimize overfitting. It works by performing bootstrap sampling (random sampling with replacement) to create several training subsets, with each training subset being used to grow an individual tree. A prediction is made by each tree and in a classification problem, the output is resolved by majority vote over all the trees.

Figure . : Random Forest model architecture [25]

## Model Training

The implementation and evaluation of the model were performed in Python, and the most notable libraries used are Scikit-learn to perform machine learning, pandas and MNE to manipulate EEG data. A 5-fold cross-validation technique was taken to make sure that the estimates of the model performance are robust and can be generalized. The data was divided into five subsets; four of the subsets were used to train and the fifth one to test the data in each fold and rotating among all combinations.

## Summary

The thesis proposes a machine learning algorithm of brain activity state classification based on the EEG signal. EEG data was recorded in 10 normal adult subjects who were required to perform four cognitive and rest tasks that were Relaxed with Eyes Open, Relaxed with Eyes Closed, Mental Arithmetic with Eyes Open, and Mental Arithmetic with Eyes Closed. EEG was measured with Contec KT88-2400, 10 electrodes were put on the frontal, temporal, and occipital area in accordance with the international system of 10-20.

During the preprocessing step, the EEG signals were split into 5-sec overlapping windows (80 percent overlap) to enhance the temporal resolution of the data. Independent Component Analysis (ICA) was used to improve signal quality and eliminate non-neural artifacts (including those introduced by eye blinks, muscle activity, and environmental noise). The EEG signals are decomposed into statistically independent components using this technique and thus the noise and the artifacts can be removed without eliminating interesting neural activity.

Based on the cleaned and segmented signals, many statistical features (mean, variance, standard deviation, skewness, kurtosis) and frequency-domain features (energy at delta, theta, alpha, beta, gamma band) were calculated. Four supervised classifiers were trained and tested using these features, namely K-Nearest Neighbors (KNN), Decision Tree, Support Vector Machine (SVM), and Random Forest. The evaluation of model performance was conducted through 5-fold cross-validation, which guarantees an unbiased robust assessment.

The performance of the trained models was checked by analyzing the classification results to see whether it corresponds to the pre-defined performance requirements. After a good performance was obtained the methodology provides the basis that can be used in future hardware implementation where the models can be incorporated into real-time EEG monitoring systems and can be further enhanced and tested.

# Chapter 4

**Results**

## Data Segmentation Techniques

For the acquired dataset, classification was performed by organizing the EEG recordings into two distinct classes: the Visual class, consisting of "Relaxed with Eyes Open" and "Arithmetic with Eyes Open," and the No-Visual class, comprising "Relaxed with Eyes Closed" and "Arithmetic with Eyes Closed".

Various segmentation windows and overlaps were tested through a hit-and-trial method to determine the one that would give the highest classification accuracy. We began with a window of 1-second (no overlap) and then we added windows of increasing size and tested different percentages of overlap. In particular, the following techniques were used:

* 1-second window: no overlap
* 2-second window: no overlap, 50% overlap
* 3-second window: no overlap, 33% overlap, 66% overlap
* 4-second window: no overlap, 25% overlap, 75% overlap
* 5-second window: no overlap, 20% overlap, 80% overlap
* 6-second window: no overlap, 33% overlap, 66% overlap

For each configuration, five classifiers Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) were applied to evaluate classification performance. The used features were Standard Deviation and Mean Absolute Deviation. 80-20 split is used for training and testing. Initial training and testing were conducted using the visual vs. no-visual binary classification setup. The results, as illustrated in Figure 4.1, show that classification accuracy generally improved with longer window lengths and higher overlap values. The 5-second window with 80% overlap consistently yielded the best results across most classifiers, with KNN and Random Forest achieving the highest accuracy rates. This suggests that longer windows offer richer temporal information, while higher overlap increases the number of training segments, enhancing the model's learning capacity.

An important observation from this experiment was that performance began to decline at the 6-second window, even with high overlap. This may be due to the inclusion of irrelevant or redundant information within overly long segments, which could reduce sensitivity to shorter cognitive changes. Additionally, longer windows with large overlaps reduce the number of independent training examples, potentially impacting generalization.

Based on these results, the 5-second window with 80% overlap was selected as the optimal configuration and used for subsequent experiments involving mental task classification and multiclass classification. This consistent setting provided a uniform evaluation environment, allowing fair comparison across classification targets and supporting robust model performance in EEG-based brain activity recognition.

Then we used this technique for binary and multiclass classification on our custom dataset. We performed binary classifications as:

* Visual vs No-Visual (termed as “Visual”)
* Mental Arithmetic vs Rest (termed as “Mental”)

The multiclass classification was performed with classes:

* Arithmetic with Eyes Opened
* Arithmetic with Eyes Closed
* Relaxed with Eyes Opened
* Relaxed with Eyes Closed

The classification results i-e model test accuracies are summarized in Table-4.1

Figure . : Model Accuracy Across EEG Window Sizes and Overlaps (Visual Class)

Table . Classification Results using Best Data Segmentation Technique

|  |  |  |  |
| --- | --- | --- | --- |
| **Models** | **5 Second Window (80% Overlap)** | | |
| **Visual** | **Mental** | **Multiclass** |
| **Logistic Regression** | 86.9 | 65.8 | 61.1 |
| **Decision Tree** | 88.9 | 65.2 | 57.1 |
| **Random Forest** | 97.0 | 94.9 | 94.1 |
| **SVM (linear)** | 87.3 | 66.4 | 63.8 |
| **SVM (rbf)** | 94.2 | 85.7 | 83.0 |
| **KNN** | 97.2 | 93.9 | 92.3 |

## Comparison with online dataset

The online dataset “EEG During Mental Arithmetic Tasks” was used for comparison, which was downloaded from PhysioNet. This data comprises EEG data of several subjects performing mental arithmetic tasks with their eyes closed, i.e., serial subtraction of a two-digit number by a four-digit number (e.g., 3141 42). In each subject, each recording includes separate artifact-free EEG segments of 180 s for resting state and 60 s for mental arithmetic. We used data of 10 subjects for binary classification with two classes, i-e “Relaxed State” and “Mental Arithmetic State”, limited data (60 seconds) was used for “Relaxed State” for classes to be balanced.

For comparison of this dataset with our acquired data, we used limited data (60s) of “Relaxed with Eyes Closed” and “Arithmetic with Eyes Closed” of 10 subjects.

For feature extraction, data is segmented into windows of 5 seconds with 4 seconds of overlap and Standard Deviation and Mean Absolute Deviation are Extracted from these windows. Different Classifiers are trained and tested, and results are summarized in Table 4.2

Table . Result Comparison between Online and Custom Dataset

|  |  |  |
| --- | --- | --- |
| **Models** | **5 Second Window (80% Overlap)** | |
| **Online Dataset** | **Custom Dataset** |
| **Logistic Regression** | 91.3 | 70.1 |
| **Decision Tree** | 85.9 | 83.2 |
| **Random Forest** | 99.0 | 95.8 |
| **SVM (linear)** | 90.8 | 70.7 |
| **SVM (rbf)** | 99.5 | 89.1 |
| **KNN** | 99.5 | 98.9 |

## ICA on Acquired Dataset

Upon comparison, we realized that online dataset showed much better accuracies than our custom dataset. This was due to artifacts present in our data. Independent Component Analysis (ICA) was used as a preprocessing method to improve the quality of the signal and increase the quality of the classification results on the obtained EEG data. This was done by using a MATLAB toolbox called EEGLAB. The cleaned dataset was then segmented with the optimal configuration found above a 5-second window and 80 percent overlap and tested on three classification tasks: visual, mental, and multiclass. Table 4.3 summarizes the post ICA performance of different classifiers.

Table .3: Results after ICA for Visual, Mental, and Multiclass Classification

|  |  |  |  |
| --- | --- | --- | --- |
| **Models** | **5 Second Window (80% Overlap)** | | |
| **Visual** | **Mental** | **Multiclass** |
| **Random Forest** | 98.7 | 99.4 | 99.0 |
| **SVM (rbf)** | 97.6 | 98.3 | 97.5 |
| **KNN** | 98.4 | 99.2 | 98.7 |

The performance of all models improved significantly, especially on the more complex tasks. Random Forest classifier performed perfectly in all three tasks, which means that feature separation was very successful following the removal of artifacts. Support Vector Machine (SVM) and KNN also achieved close to perfection accuracy of 97% and 98%.

These findings indicate the beneficial effect of ICA on EEG classification due to the enhancement of the clarity of the input signal, which subsequently increases the capability of machine learning models to correctly differentiate between cognitive states.

## Discussion

The experimental findings that were derived using the online and the acquired EEG data provide useful information regarding the efficiency of machine learning models in the classification of cognitive brain activity. In general, the results show how important proper feature extraction methods, best segmentation approaches, and signal preprocessing are in attaining high classification accuracy.

By changing the length of the window and the percentage of overlap, it was found that 5 seconds window with 80 percent overlap always gave the best classification results. This setup probably provided the most informative time context and the most training samples, thus enhancing model generalization. The KNN and Random Forest models recorded the highest accuracies in this setup, with values of more than 97%. The good results of KNN in both datasets confirm its stability in the classification of EEG signals, especially when the features are discriminative and the sample size is large.

The other important observation made on the obtained dataset was the effect of the application of Independent Component Analysis (ICA) prior to feature extraction. ICA enhanced the quality of EEG signals mostly by eliminating eye blink, muscle movement, and cardiac artifacts. Consequently, the classification performance of all models was enhanced, and KNN achieved 100 percent accuracy following ICA. This emphasizes the significance of good signal preprocessing in increasing the reliability of the model and decreasing the chance of misclassification because of noise or other irrelevant signal parts.

Moreover, the binary classification task, including the visual/no-visual conditions distinction, was classified with higher accuracy, in general, than the multiclass classification of all four mental states. This is not surprising, since the number of classes is usually associated with less overlap between features and simpler models. Nevertheless, the observation that even multiclass classification was able to reach an accuracy of more than 90% in the most successful configuration indicates that the features that were extracted still carried a high degree of discriminative power even in more complicated situations. Overall, the findings suggest that model performance in EEG classification is strongly dependent on a set of factors: quality of the input signal, the informativeness of the features extracted, and the selection of the classification algorithm. KNN and Random Forest models produced the best results across all tested models with relatively low computational complexity, which makes them appealing options to be used in the future in real-time EEG monitoring systems.

## Model Complexities

In addition to accuracy, execution time, and memory usage, **model complexity** was also evaluated to guide the selection of the most suitable classifier for deployment. The comparative analysis of **model complexities**, including RAM usage, ROM requirements, and execution time for different classifiers, is illustrated in **Figure 4.2**.

A graph with blue bars

AI-generated content may be incorrect.

Figure .2: Analysis of Model Complexities

Model complexity refers to the computational resources and internal structure required by a model to make predictions. It directly affects both the inference speed and the feasibility of running the model on hardware with limited resources (such as embedded systems).

Among the classifiers studied:

* **Random Forest** has moderate complexity due to the ensemble of multiple decision trees; however, each individual tree is relatively shallow, which keeps overall inference fast and memory usage reasonable.
* **SVM** (with an RBF kernel) has higher complexity, as predictions involve computing kernel functions against support vectors, which can increase computation time and RAM usage, particularly as the number of support vectors grows.
* **KNN** exhibits the highest complexity at inference, since it must compare each new sample to the entire training dataset, leading to significant RAM consumption and slower predictions, especially as the dataset size increases.

By evaluating model complexity together with accuracy, execution time, and memory footprint (RAM and ROM), **Random Forest** emerged as the most balanced option, providing high classification accuracy with moderate model complexity and low inference time, making it suitable for deployment on resource-limited real-time EEG systems.

## Hardware Deployment

The model was trained and implemented on a small hardware platform Odroid XU4. This single-board computer has an appropriate ratio of processing power and power consumption. According to the previous analysis, Random Forest classifier was chosen to be implemented because of its high accuracy and low running time.

A graphical user interface (GUI) was created to enable the user to interact with the system. With this GUI, the user is allowed to choose a raw EEG CSV file, and the system automatically processes it. The application will do signal segmentation, feature extraction, and classification based on the deployed model. The brain state class predicted such as visual, mental or relaxed is shown with a confidence score, providing real-time, interpretable feedback.

This application shows how EEG classification models can be practically implemented on lightweight devices, showing the possibilities of portable, real-time brain monitoring systems in non-clinical environments.



Figure .3: Graphical User Interface (GUI) for Odroid XU4 Deployment

# Chapter 5

**Conclusion**

We discussed how machine learning methods can be used to classify the brain activity state using EEG signals, specifically on cognitive and visual tasks. EEG data was obtained not only in an online open-access database but also in real-time with the use of the Contec KT88-2400 device. A standardized experimental procedure was adopted, which included four conditions of tasks: relaxed with eyes opened/closed and mental arithmetic with eyes opened/closed. The 10 20 system was used to place electrodes and ICA was used to improve the quality of the signal by removing the artifacts.

A detailed approach was elaborated, which consisted of EEG signal segmentation into overlapping windows, extraction of statistical and frequency-domain features, and training of several classifiers. Of the tested models KNN and Random Forest were the most accurate in all tests, especially with 5 seconds window and 80 percent overlap. The online dataset supported the validity of classical features, and the obtained dataset showed that segmentation strategy and ICA play a decisive role in high accuracy.

The classification tasks in the acquired dataset were expanded to the mental task and multiclass recognition besides being binary (visual vs. no-visual). ICA also contributed greatly to the improvement in classification performance, where Random Forest achieved 99 percent accuracy in all tasks. Nonetheless, when the complexity of models and execution time were factored in hardware deployment, Random Forest was the most appropriate choice because it had low inference time and high accuracy.

The Odroid XU4 was used to develop a hardware prototype that includes a graphical user interface (GUI) that enables users to feed raw EEG data and get real-time predictions with confidence scores. This effective implementation proves the possibility of brain state monitoring using EEG in a portable device.

Finally, the study provides an end-to-end pipeline of data collection to real-time implementation, which validates the fact that lightweight machine learning models are suitable to classify EEG-based brain states when preprocessing and feature extraction are effective.

# References

[1] Rashid, Mamunur & Sulaiman, Norizam & Mustafa, Mahfuzah & Khatun, Sabira & Bari, Bifta & Hasan, Md Jahid. (2020). Recent Trends and Open Challenges in EEG Based Brain-Computer Interface Systems. 10.1007/978-981-15-2317-5\_31.

[2] [Online]. Available: <https://eastneurology.com.au/eeg-electroencephalogram-brain-wave-tests/>

[3] [Online]. Available: <https://www.brainframe-kids.com/brain/facts-lobes.htm>

[4] [Online]. Available: <https://info.tmsi.com/blog/the-10-20-system-for-eeg>

[5] Bajaj, Nikesh. (2020). Wavelets for EEG Analysis. 10.5772/intechopen.94398.

[6] Huang, Z. and Wang, M., 2021. A review of electroencephalogram signal processing methods for brain-controlled robots. *Cognitive Robotics*, *1*, pp.111-124.

[7] Li, L., 2010, September. The differences among eyes-closed, eyes-open and attention states: an EEG study. In *2010 6th international conference on wireless communications networking and mobile computing (WiCOM)* (pp. 1-4). IEEE.

[8] Barry, R.J., Clarke, A.R., Johnstone, S.J., Magee, C.A. and Rushby, J.A., 2007. EEG differences between eyes-closed and eyes-open resting conditions. *Clinical neurophysiology*, *118*(12), pp.2765-2773.

[9] Al-Shargie, F.M., Tang, T.B., Badruddin, N. and Kiguchi, M., 2016. Mental stress quantification using EEG signals. In *International conference for Innovation in Biomedical Engineering and Life Sciences: ICIBEL2015, 6-8 December 2015, Putrajaya, Malaysia 1* (pp. 15-19). Springer Singapore.

[10] Ganguly, B., Chatterjee, A., Mehdi, W., Sharma, S. and Garai, S., 2020, July. EEG based mental arithmetic task classification using a stacked long short term memory network for brain-computer interfacing. In *2020 IEEE VLSI DEVICE CIRCUIT AND SYSTEM (VLSI DCS)* (pp. 89-94). IEEE.

[11] Nicolas-Alonso, L.F. and Gomez-Gil, J., 2012. Brain computer interfaces, a review. *sensors*, *12*(2), pp.1211-1279.

[12] Antoniou, E., Bozios, P., Christou, V., Tzimourta, K.D., Kalafatakis, K., G. Tsipouras, M., Giannakeas, N. and Tzallas, A.T., 2021. EEG-based eye movement recognition using brain–computer interface and random forests. *Sensors*, *21*(7), p.2339.

[13] You, S.D., 2021. Classification of relaxation and concentration mental states with eeg. *Information*, *12*(5), p.187.

[14] Islam, A., Sarkar, A.K. and Ghosh, T., 2021, July. EEG signal classification for mental stress during arithmetic task using wavelet transformation and statistical features. In *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)* (pp. 1-6). IEEE.

[15] Singla, R., Chambayil, B., Khosla, A. and Santosh, J., 2011. Comparison of SVM and ANN for classification of eye events in EEG. *Journal of Biomedical Science and Engineering*, *4*(1), p.62.

[16] Nirde, K., Gunda, M., Manthalkar, R. and Gajre, S., 2023, March. EEG mental arithmetic task levels classification using machine learning and deep learning algorithms. In *2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP)* (pp. 1-8). IEEE.

[17] Choi, S.I., Choi, G.Y., Lee, H.T., Hwang, H.J. and Shin, J., 2018, January. Classification of mental arithmetic and resting-state based on Ear-EEG. In *2018 6th International Conference on Brain-Computer Interface (BCI)* (pp. 1-4). IEEE.

[18] Wang, Q. and Sourina, O., 2013. Real-time mental arithmetic task recognition from EEG signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(2), pp.225-232.

[19] [Online]. Available: <https://qhdcontec.en.made-in-china.com/product/gnJYCWyHarRB/China-Contec-Kt88-2400-24-Channels-Digital-Electroencephalograph-EEG-Machine.html>

[20] [Online]. Available: <https://www.medicine.mcgill.ca/physio/vlab/biomed_signals/eeg_n.htm>

[21] [Online]. Available: <https://trendspider.com/learning-center/k-nearest-neighbors-knn/>

[22] [Online]. Available: <https://blog.mindmanager.com/decision-tree-diagrams/>

[23] A. Kumar, “Support Vector Machine (SVM) Python Example,” *Data Analytics*, Aug. 03, 2022. <https://vitalflux.com/classification-model-svm-classifier-python-example/>

[24] D. Wang and M. Zhang, “Nonlinear decision boundary created by a machine learning-based classifier to mitigate nonlinear phase noise,” Semantic Scholar, Dec. 03, 2015. <https://www.semanticscholar.org/paper/Nonlinear-decision-boundary-created-by-a-machine-to-Wang-Zhang/e383031d32e3d44fcc6d50b6863dd5c8e145ddcf>

[25] [Online]. Available: <https://www.researchgate.net/figure/Illustration-of-random-forest-trees_fig4_354354484>