Text

Description automatically generated

**IDENTIFICATION OF MAJOR DEPRESSIVE DISORDER BY ANALYZING EEG SIGNALS**

by

H.M. Thulana Thidaswin

Mohamed Arif Zahir

Sandali Siriwardana

Saad Mohideen

Harees Srinivasan

A Thesis submitted for   
**Undergraduate Degree, Computer Science,**

**IIT, Sri Lanka**

January 2022

# Declaration

We hereby certify that this project report and all the artifacts and research associated with it is our own work and has not been submitted before nor is currently being submitted for any degree program.

|  |  |  |
| --- | --- | --- |
| **Name of the Student** | **IIT number** | **UOW number** |
| Heenatigala Madeenage Thulana Thidaswin | 20200904 | w1838837 |
| Mohamed Arif Zahir | 20200880 | w1838914 |
| Sandali Siriwardana | 20200525 | w1830249 |
| Saad Mohideen | 2019602 | w1810577 |
| Harees Srinivasan | 20200499 | w1830218 |

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

# Abstract

Depression is a primary mental disorder that has become one of the leading causes of death by suicide. Common practices of diagnosing and ranking depression are by using screening tests and questionnaires to patients. These traditional methods are time-consuming, and mistakes are often made. So, it’s essential for an automated method to diagnose and rank depression. This paper proposes using EEG signals for this. EEG signals are emitted by the brain and is a cheap way of identifying brain related issues. Hence, this method becomes a foolproof and inexpensive method of diagnosing and ranking depression and therefore can be utilized in education institutions like schools and universities where it is needed the most. A software called Intellignosis is proposed which contains the necessary feature selection, ranking and classification techniques to identify depression. Initially an introduction is given regarding the topic and research gap which is then elaborated upon in the literature review where prior research and tools and techniques are discussed in detail. It is followed by the methodologies used during the course of this project and then the system requirements specification are discussed in detail which is followed by the social, legal, ethical and professional issues and their mitigations are discussed. Finally, the necessary architecture and designs are given with the concerning parties surrounding the project discussed.

Keywords: Depression, EEG signal, machine learning, diagnosing, ranking

# Acknowledgement

# Table of Contents

[Declaration i](#_Toc91713093)

[ABSTRACT+KEYWORDS ii](#_Toc91713094)

[Acknowledgement ii](#_Toc91713095)

[Table of Contents i](#_Toc91713096)

[List of figures iv](#_Toc91713097)

[Table appendix iv](#_Toc91713098)

[Table of abbreviations v](#_Toc91713099)

[Chapter 1: Introduction 1](#_Toc91713100)

[1.1 - Chapter Overview 1](#_Toc91713101)

[1.2 - Problem Background 1](#_Toc91713102)

[1.3 - Problem Statement 2](#_Toc91713103)

[1.4 - Research Gap The topic required extensive research to be undertaken in the different areas it explored. The different methods of identifying MDD and the techniques used to classify depressed patients that are currently in use needed to be understood to apply the relevant methodologies to a novel application. Extraction of features from an EEG signal report, classification techniques and ranking were the essential components that required research. These methods required a proper understanding of machine learning techniques and algorithms, in order to choose the optimal method to get a high accuracy percentage. 2](#_Toc91713104)

[1.5 - Research Questions 3](#_Toc91713105)

[1.6 - Research Aim 3](#_Toc91713106)

[1.7- Project Scope 4](#_Toc91713107)

[1.7.1 - In-scope 4](#_Toc91713108)

[1.7.2 - Out-scope 4](#_Toc91713109)

[1.8 - Rich Picture Diagram 4](#_Toc91713110)

[1.9 - Objectives 5](#_Toc91713111)

[1.10 - Resource Requirements 7](#_Toc91713112)

[1.11 - Chapter Summary 8](#_Toc91713113)

[Chapter 2: Literature Review 1](#_Toc91713114)

[2.1 Chapter Introduction 1](#_Toc91713115)

[2.2 Major Depressive Disorder 1](#_Toc91713116)

[2.3 Electro-Encephalography (EEG) Signals 2](#_Toc91713117)

[2.4 Existing work 3](#_Toc91713118)

[2.4.1 Binary Discrimination 3](#_Toc91713119)

[2.4.2 Intensity scaling 9](#_Toc91713120)

[2.5 Discussion of limitations 10](#_Toc91713121)

[2.6 Tools and techniques 10](#_Toc91713122)

[2.7 Chapter Summary 12](#_Toc91713123)

[Chapter 3: Methodology 1](#_Toc91713124)

[3.1. Chapter Overview 1](#_Toc91713125)

[3.2. Research Methodology 1](#_Toc91713126)

[3.3. Development Methodology 2](#_Toc91713127)

[3.4 Design methodology 3](#_Toc91713128)

[3.5 Teamwork Breakdown Structure (WBS) 4](#_Toc91713129)

[3.6 Usage of project management and collaboration software in the project 4](#_Toc91713130)

[3.7 Risks and mitigations 6](#_Toc91713131)

[3.8. Chapter summary 7](#_Toc91713132)

[Chapter 4: System Requirements Specification (SRS) 1](#_Toc91713133)

[4.1 Chapter Overview 1](#_Toc91713134)

[4.2. Stakeholder Analysis 1](#_Toc91713135)

[4.2.1. Onion Model 1](#_Toc91713136)

[4.2.2. Stakeholder Descriptions 2](#_Toc91713137)

[4.3. Selection of Requirement Elicitation Techniques/Methods 2](#_Toc91713138)

[4.4. Discussion/ Analysis of Results 4](#_Toc91713139)

[4.5. Use Case Diagrams 11](#_Toc91713140)

[4.6. Use Case Descriptions 11](#_Toc91713141)

[4.7. Functional Requirements (with prioritization) 14](#_Toc91713142)

[4.8. Non-Functional Requirements 15](#_Toc91713143)

[4.9. Chapter Summary 16](#_Toc91713144)

[Chapter 5: Social, Legal, Ethical and Professional Issues 1](#_Toc91713145)

[5.1. Chapter Overview 1](#_Toc91713146)

[5.2. SLEP Issues and Mitigation 1](#_Toc91713147)

[5.2.1 Social issues 1](#_Toc91713148)

[5.2.2 Legal issues 2](#_Toc91713149)

[5.2.3 Ethical issues 2](#_Toc91713150)

[5.2.4 Professional issues 3](#_Toc91713151)

[5.3. Chapter Summary 3](#_Toc91713152)

[Chapter 6: System Architecture and Design 1](#_Toc91713153)

[6.1. Chapter Overview 1](#_Toc91713154)

[6.2. System Architecture Design 1](#_Toc91713155)

[6.3. System Design 2](#_Toc91713156)

[6.3.1. Class Diagram 2](#_Toc91713157)

[6.3.2. Sequence Diagram 3](#_Toc91713158)

[6.3.3. UI Design 4](#_Toc91713159)

[6.3.4. Process flow chart 7](#_Toc91713160)

[6.4. Chapter Summary 8](#_Toc91713161)

[References 1](#_Toc91713162)

[Appendix 5](#_Toc91713163)

# List of figures

1. [Figure 1.1 - Rich picture diagram](#fig1_1) 1
2. [Figure 3.1 - A visual representation of EEG signals 2](#fig3_1)
3. [Figure 3.2 - Example of EEG calculations](#fig3_2) 3
4. [Figure 3.3 - Team Work breakdown structure](#fig3_3) 4
5. [Figure 3.4 - Online platform communication examples](#fig3_4) 5
6. Figure 3.5 - Gantt chart 6
7. [Figure 4.1 - Onion model](#fig4_1) 7
8. [Figure 4.2 - Use case diagram](#fig4_2) 8
9. [Figure 6.1 - System architecture diagram](#fig6_1) 9
10. [Figure 6.2 - Class diagram](#fig6_2) 0
11. [Figure 6.3 - Sequence diagram](#fig6_3) 0
12. [Figure 6.4 - Dashboard screen of UI design](#fig6_4) 0
13. [Figure 6.5 - Results screen on UI design](#fig6_5) 0
14. [Figure 6.6 - System Process Flow chart](#fig6_6) 0

# Table appendix

1. [Table i - Declaration table](#declaration_table) 1
2. [Table 1.1 - Research Objectives table](#table1_1) 1
3. [Table 1.2 - Academic Objectives table](#table1_2) 1
4. [Table 1.3 - Operational Objectives table](#table1_3) 1
5. [Table 1.4 - Resource requirements](#table1_4) 1
6. [Table 2.1 - Prior research summary table](#table2_1) 1
7. [Table 3.1 - Mitigation and risks](#table3_1) 1
8. [Table 4.1 - Stakeholder descriptions](#table4_1) 1
9. [Table 4.2 - Use case descriptions UC - 001](#table4_2) 1
10. [Table 4.3 - Use case descriptions UC - 002](#table4_3) 1
11. [Table 4.4 - Use case descriptions UC - 003](#table4_4) 1
12. [Table 4.5 - Functional requirements table](#table4_5) 1
13. [Table 4.6 - Non-Functional requirements table](#table4_6) 1

# Table of abbreviations

ADHD - Attention deficit / Hyperactivity Disorder

ANN - Artificial Neural Networks

ARM - Auto Regressive Method

ASD - Autism Spectrum Disorder

BDI - Beck’s Depression Inventory

CC - Centered Correntropy

CD - Correlation Dimension

CNN - Convolutional Neural Network

DFA - Detrended Fluctuation Analysis

DT - Decision Trees

DWT - Discrete Wavelet Transform

EEG - Electroencephalography

EM - Eigenvector Method

EWT - Empirical Wavelet Transform

FD - Fractal Dimension

FFD - Fuzzy Fractal Dimension

FFNN - Fuzzy Function Neural Network

FFT - Fast-Fourier Transform

FFT - Fast-Fourier Transform

FuzzyEn - Fuzzy Entropy

GA - Genetic Algorithm

H - Hurst’s exponent

HAM-D - Hamilton Depression rating scale

HFD - Higuchi Fractal Dimension

HOS - Higher Order Spectra

KFD - Katz’s Fractal Dimension

kNN - k Nearest Neighbor

LDA - Linear Discriminant Analysis

LLE - Largest Lyapunov Exponent

LR - Logistic Regression

MCFS - Multi-Cluster Feature Selection

MDD - Major depressive Disorder

MLP - Multilayer Perceptron

NB - Naïve-Bayesian

NCA - Neighborhood Component Analysis

OOD - Object oriented design

PID-5 - Personality Inventory DSM-5

RF - Random Forest

RQA - Recurrence Quantification Analysis

RWE - Relative Wavelet Energy

SLEP - Social, Legal, Ethical problems

SSADM - Structured systems analysis and design methodology

SVM - Support Vector Machines

SampEn - Sample Entropy

TFD - Time Frequency Distributions

WT - Wavelet Transform

# Chapter 1: Introduction

## 1.1 - Chapter Overview

This chapter aims to provide the reader with a befitting introduction to the problem that is being addressed, what attempts have been made to tackle the problem in the past, i.e., previous research and gaps determined, the problem addressed and the research gaps and scope that is to be covered. Finally, based on adequate research and forethought, the project, academic and operational objectives will be expounded on. Finally, based on adequate research and forethought, the project, academic and operational objectives will be expounded on.

## 1.2 - Problem Background

Major Depressive Disorder (MDD), a common mental illness that plagues over 280 million people over the world, inclusive of 5.0% in adults and 5.7% among individuals over 60 (WHO, 2021), and is drastically different from a common waiver in emotions or a temporary negative emotional reaction to an event.

In terms of diagnosis of MDD, excluding the physical tests and blood tests, generally a psychiatric evaluation is needed to interpret behavior, history and to analyze symptoms (National Institute of Mental Health, 2018). Two-thirds of cases of depression are estimated to be undiagnosed in the U.S alone and 65.9 percent rates of misdiagnosis (Williams, Chung and Muennig, 2017), the need for an accessible, convenient and a reliable objective method to detect MDD, as early as possible, is more important now than ever.

The stigma and discrimination on the topic of depression, and mental illness in general is still widespread (APA, 2020). What this ultimately points towards, is that the group most likely to undergo depression is also the most likely to not seek aid or diagnosis. This social pressure on young adults is an instrumental factor that accords to drastically high amounts of undiagnosed depressed patients.

As the 4th highest leading causes of death in young adults (aged 15 to 29), and with over 700,000 suicide cases every year (WHO, 2021), depression is a mental illness that, despite many efforts in media and education to diminish, has proven to be a calamitous problem in the 21st century.

## 1.3 - Problem Statement

With Depression on the rise and undiagnosed/misdiagnosed patients commonplace, a need for an objective, accurate manner of detecting depression as well as scaling the severity as early as possible, especially in regard to the age groups of 15-29 in order to facilitate an accessible, inexpensive manner to detect depression for young adults is in need.

## 1.4 - Research Gap The topic required extensive research to be undertaken in the different areas it explored. The different methods of identifying MDD and the techniques used to classify depressed patients that are currently in use needed to be understood to apply the relevant methodologies to a novel application. Extraction of features from an EEG signal report, classification techniques and ranking were the essential components that required research. These methods required a proper understanding of machine learning techniques and algorithms, in order to choose the optimal method to get a high accuracy percentage.

In general, symptoms of Depression must be present for at least two week before a diagnosis can take place (National Institute of Mental Health, 2018), and this time frame does not take into account the time that the patient takes to actually decide to go for a real diagnosis. Which might not be suitable for a patient who requires immediate action. This project aims to reduce that time to a maximum of a day (for a safe margin of technical difficulties) or a minimum of 10 minutes. This way doctors and specialists can aid the patient much faster and traverse towards a cure.

While a few previous research attempts have explored identifying depression using EEG signals, this project aims to further the research on identifying the level of depression using the Beck Depression Inventory (BDI-II), a rating scale that will be used to standardize the rating.

## 1.5 - Research Questions

The following research questions were used to identify the necessary research components that were required for the successful implementation.

Q1) What are the techniques used to classify a healthy and depressed individual?

Q2) What are the techniques to extract relevant features from a set of EEG signals?

Q3) What are the techniques and methodologies used to identify the level of a depressed individual?

## 1.6 - Research Aim

This research would contain multiple aims to fulfill different aspects that the final application would need.

Firstly, in order to design the project a background on EEG signals and an elucidation on each of the different types of features that can be extracted would be a must. With the scale for rating already defined, suitable datasets would be needed for the implementation.

To implement the program, a sound understanding of Machine learning algorithms is needed to decide on a fit type that satisfies the specifications of the project and yielding a satisfactory accuracy percentage.

After the program is implemented, an evaluation would be made according to the accuracy percentage achieved in comparison to other projects similar to this, the level of depression defined precisely, and the practicality of the program (regarding the interface and ease of use).

## 1.7- Project Scope

This project deals mainly with the analysis of EEG signals to provide a detailed report of the patient’s mental state.

### 1.7.1 - In-scope

1. This program is aimed to be used in schools and universities to identify depression among students and make the concerned parties aware about a disorder.
2. The features that are included is the ability to differentiate depressed patients from healthy controls and test the level of depression using the BDI Scale.

### 1.7.2 - Out-scope

This process is solely a tool to help students accurately uncover their depression symptoms and find out the severity of the disorder. Due to time constraints imposed on the project the aspects out of scope are:

1. This project method deals only with the usage of EEG signals for detection; however the collection of EEG signals is left to the standardized and low cost processes available around the country and the world and it is preferred that schools and universities purchase such EEG equipment to help students assess their mental status.
2. Practices and actions that need to be taken after diagnoses are also not provided since they are best handled by a human, especially psychiatric doctors.

## 1.8 - Rich Picture Diagram

Diagram

Description automatically generated

Figure 1.1 shows the Rich picture diagram

EEG signal acquisition will simply be the direct input to the computer. As the EEG setup and hardware is out of scope, this step would be the acquisition of signals to the program. The signal preprocessing would be the extraction of the relevant features and then automatically choosing the right variables the application needs in order to move to the next step. The output is after the algorithm has successfully classified the patient's mental state and accurately determines the level of depression if the patient is detected with depression.

## 1.9 - Objectives

The aim is to identify MDD using EEG signals to provide the level of depression the individual possesses the target group of 15-29.

Table 1.1: Research Objectives

|  |  |
| --- | --- |
| **Research Objectives** | RO1: To identify MDD using EEG signals. |
| RO2: Extraction of features |
| RO3: Level identification of depression. |

Table 1.2: Academic Objectives

|  |  |
| --- | --- |
| **Academic Objectives** | AO1: Identification of a research gap. |
| AO2: Utilizing a principle of computer science in order to further the      research and create a successful research project in the given time. |

The research gap referred to in Table 1.2 is the utilization of EEG signals in identifying depression with machine learning algorithmic principles to create a program aimed towards the target group in the span of one academic year.

Table 1.3: Operational Objectives

|  |  |
| --- | --- |
| **Operational Objectives** | OO1: The program will have to be accessible to the target group and to universities and students. Therefore, storing the program in a cloud-based service would be recommended but due to time constraints, the application would be run offline. |
| OO2: The interface and the overall ease of use should be another point of focus. Students and teachers alike should be able to operate the program with ease by the use of intuitive interfaces, simple program design and architecture and vomiting much use of jargon or other complex terminology that may confuse the user. This way even users without a sound base knowledge on the topic of depression or EEG signal data would be able to utilize the program and still get a comprehensible result. |
| OO3: The program should yield a satisfactory level accuracy along with a precise level of MDD detected in order for the program to completely achieve its main objective. |

## 1.10 - Resource Requirements

The resources needed to work on the research project are highlighted below.

Table 1.4: Resource requirements

|  |  |
| --- | --- |
| **Hardware Requirements** | The hardware that will suffice would be a personal computer with decent performance (a safe benchmark of measure would be the level of performance of individual components of the respective computer or laptop)   * A minimum range of an Intel i5 processor, around 8 gigabytes of RAM, and the necessary ports needed to input the EEG signals. * The recommended specifications of an **intel i7 processor**, **16 gigabytes of RAM**, **a** **dedicated GPU**, or higher. A good performance benchmark like this is important since machine learning requires decent processing power and multitasking. |
| **Software Requirements** | * **Python** would have to be installed as the mechanic learning algorithms are coded in Python. * **Matlab** in order to read datasets. * **Machine learning software** would also be recommended, popular examples would include IBM machine learning software, Anaconda, Tensorflow, and so on. Based on UI, usability, and integrations with other libraries a suitable software can be chosen along with **libraries**. |
| **Data Requirements** | * The dataset to diagnose depression was obtained from a public domain called Figshare. The dataset was created by Wajid Mumtaz and consisted of 34 depressed patients and 30 healthy subjects. * A dataset to identify the level of depression was obtained from OpenNeuro. This dataset was created by James F. Cavanagh. |

## 1.11 - Chapter Summary

Depression is a disorder faced by many young individuals around the world owing to a multitude of reasons. This chapter highlights how this project aims to create a program to identify MDD among individuals using EEG signals procured from their test and use it to rank the level of depression in order to begin treatment at an early stage. The scope for this project is simply implementing. The interpretation part of the process and the actual acquisition of signals from the EEG equipment are best left to a human specialist.

The research gap encountered were the techniques that were required to answer the research questions mentioned above. The objectives of the research were to create a program that would help educational institutions identify depressed students and take immediate action regarding it. Finally, the software, hardware and data requirements highlight the necessities in order to carry out the project successfully.

# Chapter 2: Literature Review

## 2.1 Chapter Introduction

This chapter provides a descriptive analysis of the existing work on EEG signals in identifying depression. A variety of approaches will be discussed, and their limitations will be reviewed at the end. This chapter will also provide an in-depth analysis of the research this report is based on and how it could overcome the challenges faced by the previous studies on the topic. The essential tools and techniques to be used are also analyzed in detail. Prior to discussing these in length, a brief overview of the necessary components of the research will be given for clearer understanding on the subject.

## 2.2 Major Depressive Disorder

Unlike mental disorders like ASD, ADHD, Schizophrenia and such, depression is a word that is loosely used to describe low moods and is often interchangeably used with stress and anxiety. But clinical depression is a much more serious mental issue. The American Psychiatric Association defines nine major symptoms out of which five are enough to diagnose a patient with MDD (Nemeroff, 1998).

Although diverse technologies have been implemented for the identification and treatment of major mental disorders, automated diagnosis of depression remains in a gray space. The research into this is an ongoing process and doctors continue to use traditional methods on patients. Screening, occasional questionnaires are the methods doctors follow to make conclusive diagnosis.

Screening is a process where doctors ask a set of standardized questions from patients to look for specific symptoms in a patient’s thought and behavior. General blood tests and physical e

xaminations are also done to diagnose physical wellbeing, but doctors mainly look for the overall mental state of the patient. This often leads to incorrect diagnosis. No two depressed patients are the same and each person displays different ways of withdrawal from a healthy mental state. This makes the diagnosis of depression a difficult process (Fulghum, 2008). It is essential to identify the mental state correctly and rapidly for effective treatment.

While the research on the biology of depression is far from completion, sources have already proven a lot regarding the subject of the physical effect on the brain in MDD patients.

One of the most common findings in this topic would be that the hippocampus, a part of the brain that plays a major role in long term memory and learning (Anand and Dhikav, 2012) is comparatively smaller in patients with MDD than healthy individuals. (Sheline et al, 1999)

Regarding EEG signal detection, Depressed patients display a hemispherical asymmetry in comparison to individuals without depression. (Davidson, 1998) and more specifically in the alpha band waves (8-13Hz) in the mid-frontal region (Henriques and Davidson, 1991), (Gotlib, 2010).

(Allen et al, 2004) A study shows that it has proven to be a consistent reliable manner of detection and therefore the resting frontal electroencephalographic (EEG) alpha asymmetry serves as a trait marker in EEG signal detection for depression.

EEG signals in depressed people also tend to be “less complex” and more predictable compared to control groups, as a decrease in the functionality of the brain in MDD patients correlates to less interaction with the environment (Nandrino, 1994).

## 2.3 Electro-Encephalography (EEG) Signals

The human body emits a variety of electric signals that are essential for bodily functions. Out of these, the signals that are emitted by neurons during exchange of information are called brain signals. Electro-encephalography is a widely used technique that is used to capture these signals. A set of electrodes is placed on the scalp for a period of time and the ionic activity of neurons are captured and displayed through a monitor. The output of an EEG test contains a structure of waves with differing amplitudes. They are a combination of four different wave structures: Gamma, Beta, Alpha, Theta and Delta. They are within the range of 0.1 – 100Hz or more (Kumar and Bhuvaneswari, 2012).

Throughout the years the usage of EEG signals has increased well beyond the scope of medicinal applications owing to their low cost and easy access. EEG signals are experimented with brain-computer interactions and neuromarketing (Lai et al., 2018). In the medical field, a few of the uses of EEG signals are the diagnosis of seizures, head injuries, brain tumors, stroke, and dementia. Therefore, it is obvious that EEG could be used for the detection of mild mental disorders like depression as well.

## 2.4 Existing work

### 2.4.1 Binary Discrimination

Machine learning is not new to the field of healthcare. It has been used in keeping smart records, medical imaging and diagnosis, drug discovery and development, and the treatment of diseases. However, there is still reluctance towards using machine learning to diagnose mental illnesses. This reluctancy stems from the lack of widely accepted research.

Using EEG signals provides a way to overcome this issue. As described in the previous section, EEG signals are used extensively to identify problems related to the brain. Using machine learning to interpret a depressed patient’s EEG signals to diagnose and classify his/her depression could prove to be essential in mental healthcare.

The necessary processes when diagnosing depression using EEG signals are feature extraction and classification. Feature extraction refers to the process of reduction of features in a dataset to be processed by a machine learning model (Guyon and Elisseeff, 2006). There are several methodologies for the extraction of features from EEG signals. They include Time Frequency Distributions (TFD), Fast-Fourier Transform (FFT), Wavelet Transform (WT), Eigenvector Method (EM) and Auto Regressive Method (ARM) (Al-Fahoum and Al-Fraihat, 2014).

Feature classification is about using algorithms to predict an outcome. In this case, there will be two types of outcomes: a binary outcome (whether a patient is depressed or not) and a multi outcome (level of depression). Some examples of the classifiers that could be used are Decision Trees (DT), Logistic Regression (LR), Support Vector Machines (SVM), Naïve-Bayesian (NB) (Kotsiantis, 2006).

Although many feature extraction methods and classifiers exist, it is necessary to identify the most suitable techniques for the present scenario to obtain a highly accurate prediction. The following few paragraphs describes the most relevant literature that have been used to research on this topic. Following that, a comparison chart is presented summarizing the given literature.

In 2012, Hosseinifard, Moradi and Rostami studied 45 unmedicated depressed patients and 45 normal subjects using non-linear features and was able to obtain an accuracy of 83.3 percentage. The study used Detrended Fluctuation Analysis (DFA), Higuchi Fractal (HF), Correlation Dimension (CD), and Lyapunov Exponent as the features that were extracted from the 4-band EEG signal. k-Nearest Neighbor (kNN), Linear Discriminant Analysis (LDA) and Logistic Regression (LR) were used as classifiers. Genetic Algorithm (GA) was used as the feature selection technique. The study says that other classification techniques were also tested such as Support Vector Machine (SVM) with non-linear kernel and Naïve Bayes (NB) but the results with LDA, LR and kNN were far superior.

A paper by Puthankattil and Joseph (2012), uses Relative Wavelet Energy (RWE) to extract features and LDA and SVM were used as classifiers to give an overall accuracy of 98.11 percent. The study consisted of normal and depressed patients out of whom 16 were female and 14 male and the same number and sex of normal subjects were used.

Another study conducted by Acharya et al (2015a) mainly focused on using non-linear methods for computer-aided diagnosis of depression. This study used feature extraction techniques such as Fractal Dimension (FD), Recurrence Quantification Analysis (RQA), Higher Order Spectra (HOS), sample and approximate entropy, Largest Lypanov Exponent (LLE), Hurst’s exponent (H), and DFA. The classification algorithms used by the author are SVM, DT, and kNN. The ‘ten-fold cross validation’ method was used to select the best classifier which would use the least number of features with higher accuracy. This paper was mostly a comparison between different methods of feature extraction and classification, and the SVM classifier yielded the highest accuracy of 98.5 percent.

A follow-up paper to the above was published by Acharya et al (2015b) which provided a novel technique to identify depression using EEG signals. The methods used in this research was the same as above, but the classifier used was only SVM. The data was collected from 15 normal and 15 depressed patients. The author claims to have an accuracy percentage of 98 percent. But the limitation in this study is the number of test cases and the lack of diversity among the subjects. The author also mentions a possible extension of this study to identify the intensity of depression which is precisely the goal of this paper.

Bairy, Niranjan and Puthankattil (2015) used Discrete Wavelet Transform (DWT) to extract features such as skewness, energy, kurtosis (the degree of sharpness of a particular curve (Oxford Reference, 2021), standard deviation, mean and entropy. The classifier used was SVM with the highest classification accuracy of 88.92 percent. The data for this study was collected from the Psychiatry College, Calicut, Kerala, India and 2400 depression data and 2159 normal data were obtained.

A paper by Mantri et al (2015), analyses EEG signals, mainly focusing on linear features. 13 depressed patients and 12 controls were used as the dataset for this experiment. The Fast-Fourier Transform (FFT) method was used as the feature extraction technique and a combination of SVM and Artificial Neural Networks (ANN). The study claims an accuracy percentage of 84 percent.

Mohan et al (2016) used ANN to classify depressed and normal subjects. FFT was used as the feature extraction technique. The study used a dataset of 53 normal and 63 depressed subjects and obtained an accuracy of 95 percent.

In a paper by Mumtaz et al (2017), 33 MDD patients and 30 healthy controls participated in a study to diagnose depression using machine learning methods applied to EEG signals. In this paper, a rank-based feature selection method was utilized and LR, SVM, and NB models were used as classifiers out of which SVM provided the highest accuracy of 98.4 percent.

As a pervasive approach to the problem, Cai et al (2018), extracted a combination of both linear and non-linear features using DWT and used four classifiers SVM, kNN, Classification Trees, and ANN out of which kNN showed the highest accuracy percentage of 79.27 percent. For this study a database containing 92 depressed patients and 121 normal controls was constructed.

A study conducted by Mahato and Paul (2020) used Multi-Cluster Feature Selection (MCFS) as the feature selection technique and SVM, LR, NB and DT as classifiers for the diagnosis of depression. The dataset acquired was from a publicly available dataset which consisted of Eyes-closed EEG of 30 MDD patients and 30 normal subjects. The EEG signal bands that provided a higher classification accuracy in this research was the Alpha2 band with an accuracy percentage of 88.33%.

A study conducted by Saeedi, Saeedi and Maghsoudi (2020) extracted a combination of both linear and non-linear features using GA. The linear features consisted of the five common frequency bands and non-linear feature were sample and approximate entropy which were applied to wavelet-packet coefficients. The algorithms used for classification were SVM, MLP, and kNN. The aim of the study was to use an enhanced kNN algorithms for classification and the study claims to have an accuracy percentage of 98.44 percent. The dataset used in this paper consisted of EEG signals from 34 depressed patients and 30 normal subjects. The conclusion of the study was the fact that the gamma band of EEG signals is the most important feature when classifying depressed patients.

A study by Eralemir et al (2020) used 30 depressed patients to obtain the EEG signals and features were selected using Continuous Wavelet-Transform and kNN was used as the classifier. The authors claimed to have achieved an accuracy of 91.3 percent from this study.

Akbari, Sadiq and Rehman (2021) claimed an accuracy of 99.05 percent but using Centered Correntropy (CC) and Empirical Wavelet Transform (EWT) as feature extraction techniques and using kNN and SVM as classification techniques. The dataset was collected from 22 depressed and 22 normal patients.

A study by Aydemir et al (2021) proposed a novel depression detection using melamine patterns. Melamine patterns were to generate features which resembled the molecular structures of DNA. The study used Neighbourhood Component Analysis (NCA) to select the features and kNN and SVM as classifiers. The dataset used consisted of 34 MDD patients and 30 healthy controls. This study claimed to have a highest classification accuracy of 99.11 percent.

In the above literature, regardless of the features and feature extraction/selection techniques, the classifiers kNN and SVM seem to produce the highest accuracy. But a paper by Cukic et al (2020) says that the successful classification of normal and depressed subjects should be attributed to the feature extraction methods rather than the classification techniques. This paper uses Higuchi’s Fractal Dimension (HFD) and Sample Entropy (SampEn) as non-linear features of the EEG and uses seven algorithms, namely, Multilayer Perceptron (MP), LR, SVM (both linear and polynomial kernel), DT, Random Forest (RF), and NB. The average accuracy among the classifiers were in the range between 90.24 to 97.56 percent. Out of these SampEn had the highest track record.

The table given below summarizes the prior research described above.

Table 2.1: Summary of research findings

|  |  |  |  |
| --- | --- | --- | --- |
| **Papers** | **Feature extraction/selection techniques** | **Classification algorithms** | **Highest Accuracy** |
| Hosseinifard et al (2012) | Genetic Algorithm | kNN  LDA  LR | 83.3 % |
| Puthankattil and Joseph (2012) | Relative Wavelet Energy | LDA  SVM | 98.11 % |
| Acharya et al (2015a) | Fractal Dimension  Recurrence Quantification Analysis  Higher Order Spectra  Sample and Approximate entropy  Largest Lypanov Exponent  Hurst’s exponent  Detrended Fluctuation Analysis | SVM  DT  kNN | 98.5 % |
| Bairy, Niranjan and Puthankattil (2015) | Discrete Wavelet Transform | SVM | 88.92 % |
| Mantri et al (2015) | Fast-Fourier Transform | SVM  ANN | 84.0 % |
| Mohan et al (2016) | Fast-Fourier Transform | ANN | 95.0 % |
| Mumtaz et al (2017) | Rank-based feature selection method | SVM | 98.4 % |
| Cai et al (2018) | Discret Wavelet Transform | kNN | 79.27 % |
| Mahato and Paul (2020) | Multi-Cluster Feature Selection | SVM  LR  NB  DT | 88.33 % |
| Saeedi et al (2020) | Genetic Algorithm | kNN | 98.44 % |
| Eralemir et al (2020) | Continuous Wavelet-Transform | kNN | 91.3 % |
| Akbari et al (2021) | Centered Correntropy  Empirical Wavelet Transform | kNN  SVM | 99.05 % |
| Aydemir et al (2021) | Neighbourhood Component Analysis | kNN  SVM | 99.11 % |
| Cukic et al (2020) | Higuchi’s Fractal Dimension  Sample Entropy | SVM  MP  LR  DT  RF  NB | 97.56 % |

The above literature explained the different features, feature selection and extraction techniques as well as the classifiers that have been used in prior research to diagnose depression using EEG signals. But a common limitation of all the above papers was the inability to calculate the intensity of the depression in the patients. This severity scaling is essential during early diagnosis because it helps the concerned parties obtain a deeper understanding and will help in quicker decision-making regarding the patient’s mental health.

### 2.4.2 Intensity scaling

In traditional methods of diagnosing depression, doctors use several techniques to identify the level. Some such methods include Beck Depression Inventory (BDI), CES-D scale, Hamilton Depression (HAM-D) rating scale, and SPSI-RTM. Out of these the BDI is consistent compared to the other scales and is generally preferred by both clinicians and patients (Cusin, 2009). The BDI consists of 21 questions with 4 answers each that contain a certain set of points which are summed up at the end of diagnose the severity of a MDD patient’s illness (Beck et al, 1961).

Prior work on identifying the level of depression is very limited and only a handful of papers were found with regard to this.

Mohammadi, Hajian and Moradi (2019) used a Fuzzy Function Neural Network (FFNN) and SVM as a classifier and Fuzzy Entropy (FuzzyEn), Katz’s Fractal Dimension (KFD), and Fuzzy Fractal Dimension (FFD) as features. This study used BDI to discriminate the level of depression into four categories. The dataset used contained 60 participants with different levels of depression and the highest accuracy was obtained at 90 percent.

A recent paper by Mahato et al (2020) discusses the possibility of using the HAM-D scale for discriminating the level of depression. A combination of both linear and non-linear features (SampEn and DFA). The classifier used for both discrimination and severity scaling was SVM and the accuracy obtained respectively were 90.26 percent and 75.31 percent. The dataset was collected contained 24 depressed patients and 20 normal subjects which were collected from Central Institute of Psychiatry, Jharkhand, India.

A unique approach is taken by Zhang et al (2020) by using EEG signals to directly identify the intensity of depression in a patient rather than diagnosing binarily. The paper uses LSBoost regression and Fourier features to identify the levels. The scaling is based on the Personality Inventory for DSM-5 (PID-5) where 4 points are used to identify the level. The features that were used were alpha asymmetry and HFD. The data was obtained from 73 participants.

Kang, Kang and Lee (2021) used the Beck Depression Inventory to predict the scores of depressions by using a regression model. The EEG signals were imaged using the deep-asymmetry method and a two-dimensional CNN model was trained. The dataset was obtained from a public domain which consisted of 122 college students along with their BDI scores.

## 2.5 Discussion of limitations

The literature described above clarifies that the approaches that can be used to diagnose depression from EEG signals are diverse. Although many papers claimed to have an accuracy of over 90 percent, direct implementation of the methods required deeper understanding of the technologies and methods, expensive devices, and professional understanding of the subject. But due to the lack of availability of such resources in many countries, it is essential to make a simplistic method that could be easily made available in schools and universities to look after the mental health of students. Although using this machine learning technique is not a replacement for a doctor’s diagnosis, it could prove to be helpful in a learning environment where the concerned parties would be made aware of such a problem.

## 2.6 Tools and techniques

The classifiers that produced the highest accuracy in all the papers describes above were kNN and SVM. k-Nearest Neighbour (kNN) was developed by Fix and Hodges in 1951 and is generally based on the distance between a test sample and a training sample, hence it is a powerful algorithm that can used for classification (Peterson, 2009). Support Vector Machine (SVM) is also a supervised machine learning algorithm that is powerful at pattern recognition and has become popular for classification and regression (Noble, 2006). During the implementation of the machine learning component of the application, these two algorithms could be used for classification and regression of EEG signals.

EEG signal features could be divided into two categories as linear and non-linear features. Linear features include peak, standard deviation, skewness, variance, etc. Non-linear features include HFD, SampEn, Largest Lyapunov Exponent (LLE) and so on and so forth. Most of the literature described above chose to use non-linear features because of the non-linear patterns of brain signals and therefore claimed to have obtained a higher overall accuracy. Therefore, feature selection of EEG signals needs to be carefully considered during implementation.

In the literature review presented above, a multitude of feature extraction techniques were described but a recurring technique that was seen was the FFT algorithm. This algorithm is used to describe the relationship between time and frequency domain features of discrete signals (Nussbaumer, 1981). Since EEG signals are discrete signals and mainly deal with the time and frequency, this algorithm would be ideal for feature extraction.

Channel selection is also necessary for the detection of EEG signals. There are five types of channels that can be selected in EEG signals: frontal, central, parietal, occipital, and temporal. Prior research used only frontal channels such as Fp1, Fp2, F3, etc., but certain papers such as Kang, Kang and Lee (2021) used channels from all five types. Channel selection is also a necessary component during implementation.

The Beck Depression Inventory (BDI) is to be used as the standard to calculate the level of depression which contains six categories of depression.

Datasets for the project were obtained from publicly available domains. A dataset consisting of 30 healthy subjects and 34 MDD patients was available on [www.figshare.com](http://www.figshare.com) which was uploaded by Wajid Mumtaz on 23.11.2017. This dataset was created at the Universiti Sains Malaysia, Malaysia by Mumtaz et al. (Mumtaz, 2016). Another dataset to identify the intensity of depression was obtained from [www.openneuro.org](http://www.openneuro.org) which was uploaded on 15.01.2021 by James F. Cavanagh. This data was created from the University of Arizona, USA.

MATLAB has been chosen as the application to run the machine learning models due the availability of around 30 EEG based toolboxes. Python is the language that is the most suitable for machine learning because of its simplicity and easily readable code format and is therefore chosen as the primary language for the application. The application is to be designed as an offline desktop application that is to be used in computers and therefore the frontend and interface is to be developed using Flutter and/or JavaScript.

## 2.7 Chapter Summary

Using EEG signals as depression detection instruments have been under discussion for a long time and the contemporary literature surrounding the topic propose multiple methods and techniques to diagnose depression. In addition, predicting the intensity level of depression using predefined scales, although scarce have also been discussed. This chapter has also discussed the limitations of prior researches and describes how they align with the research gap of this paper. The possible tools and techniques that could be used are also discussed in detail.

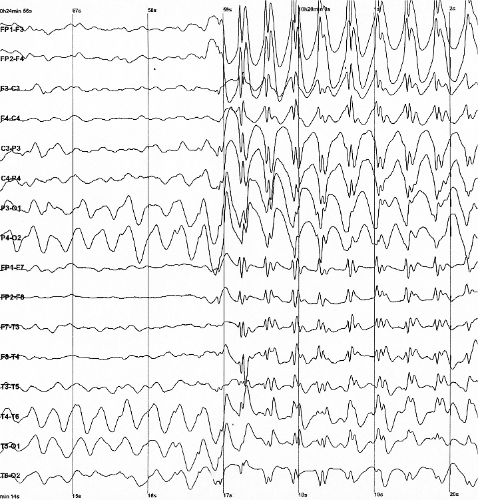
# Chapter 3: Methodology

## 3.1. Chapter Overview

This chapter focuses on the technicalities of the project, namely the methodologies followed, the work breakdown structure, tentative methods to follow during implementation and how the team coordinated and delegated work during the project. The risks and liabilities faced during the project are discussed and how they were mitigated are also explicitly discussed.

## 3.2. Research Methodology

EEG signals often come in the form of crests and troughs having both negative and positive values and the EEG machine (device used to measure EEG signals) often display the waves numerically or graphically.

Diagram

Description automatically generated

Figure 3.1 shows a visual representation of how the EEG machine displays EEG signals

Figure 3.2 shows how an EEG machine calculated EEG signals

The lines produced by the EEG machine do not pr

ovide much information, however, the peak and low voltages do provide good information. Using EEG signals to diagnose a medical or physiological disease is highly statistical as the voltage itself does not give any symptoms of any disease. The people with a certain disease which is proven by a qualified doctor is taken as a subject (depression, for instance). Countless tests are run on the subject and then compared with a healthy person to obtain valid information, therefore quantitative research methodology is being used here as it relies on statistical and numerical data. A lot of measurements will be taken to diagnose the illness (Watson, 2015).

## 3.3. Development Methodology

As for the development methodology, the waterfall development model is being implemented as requirements must be researched upon and made clear before proceeding into the next phase. Each step/plan must be executed and completed fully before proceeding therefore it increases efficiency within the team and speeds up the process as all the members are focused on one particular task at a time. Each phase is also often frozen/untouched when it's completed therefore it's a little time consuming but gets the work done (Balaji and Sundararajan, 2012).

A few advantages of this system would be:

1. Very simple to use and understand
2. Easy to manage as all the processes are frozen after its done
3. It saves a lot of time as all the processes can phases can be completed simultaneously
4. It is very helpful for beginners as everything is simple

A few disadvantages of this system would be:

1. Once you reach the latter part of the project it is difficult to amend something from the early stages
2. High amount of risks
3. A lot of time would be consumed as every step and phase is well documented as going back into the previous stages is denied as they are frozen

(Chandra, 2015)

## 3.4 Design methodology

The design methodology best suited for this project is OOD (Object oriented design) as this design is efficient and time saving due to its characteristics such as inheritance, encapsulation and more (Bansiya and Davis, 2002). Compared to SSADM (Structured systems analysis and design methodology) OOD is more efficient and more modern as SSADM is very old and uses structures like data dictionaries, decision tables and more outdated technology.

A few advantages of OOD compared to SSADM are

1. OOD has high reusability as the same instance or object can be used multiple times anywhere in the program (GeeksforGeeks, 2020)
2. It refines and extends the design using incremental or iterative technique. (GeeksforGeeks, 2020)
3. Makes the software rich in quality as redundancy is reduced and the coding style is improved (Aladib, 2015)
4. By using abstraction, data could be hidden well increasing the overall security (Aladib, 2015)

A few drawbacks of this system are

1. The learning curve is a bit steep for new developers as OOD is rather complicated than SSADM as it uses case diagrams, state chart diagrams, etc. (GeeksforGeeks, 2020).
2. The programs using OOD tends to be more complex than SSADM (Aladib, 2015)
3. OOD programs are quite large in size therefore the program would be slower than others (BrainKart, n.d.).

## 3.5 Teamwork Breakdown Structure (WBS)

Diagram

Description automatically generated

Figure 3.3 shows the teamwork breakdown structure

## 3.6 Usage of project management and collaboration software in the project

Due to the pandemic our team movement and efficiency was affected drastically due to travel restrictions and health guidelines, so all communication was conducted through online means. The reason we did not even try to meet up physically was because each team member lived far away from each other, even in separate countries. Therefore, it wasn’t ideal to meet up physically. WhatsApp was used as the main means of communication as it is a simple way to pass on messages and media. A lot of planning happened on this platform which contributed a lot to our project.

Graphical user interface, text, application

Description automatically generatedThe Google collaboration platform proved to be a major help for this project. Google calendar was used in organizing virtual conferences and meets as it has functions like screen sharing etc. Google meet was built in to google calendar therefore it was used for better accessibility. Gmail and Google drive were used to share and save large documents. In addition to Google Meet, the Zoom platform was also used for to access remote control.

Figure 3.4 shows an example of how Google Calendar and Google Meet were used

For project management two main platforms were used one being Trello and the other being Clickup. Initially, Trello was used but the team quickly switched to Clickup and it became the software that was used for the majority of this project as it had many features such as setting deadlines for easier work management and assigning tasks to relevant members. Google scholar, IEEE Xplore, O’Reilly Library, UoW’s online library, IIT’s online library and other sample theses helped a lot to this project with a lot of information since libraries were not available due to the pandemic. Google Docs was used to write/edit the SRS document due to its functions of collaborative writing where multiple people can edit the same document. The final draft of the SRS was written in MS Word because of the plethora of features it had.

## 3.7 Risks and mitigations

Table 3.1: Risks and mitigations

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Severity** | **Frequency** | **Mitigation Plan** |
| Validity of the data set | Very severe as incorrect data might lead to false conclusions | Low depends on the the data set | More research into finding a valid data set |
| Power cuts | Very severe as this was a major cause for meetings being rescheduled or cancelled | Very high due to our country’s situation | Investing on a generator but that is not a valid point |
| Finding credible data | Not that much of an effect | Low as a lot of research papers were used | Looking for more credible resources and not getting information from unknown websites. |
| Time constraints | Moderate as work started to pile up the last moment | Low as the work cleared | Work ahead of time |
| Complex theoretical components | Very severe because a lot of new things had to be learnt by reading papers and online journals | Very high as the learning curve was very steep | Research more into that particular subject and if possible, consult experts on this matter |
| Network problems | Severe as most of the work and communication is done online | High as its not reliable | Working at off peak hours although it is not an option for most |
| Communication difficulties | Severe as a lot of time is taken to explain concepts and theories online | High as almost everything is done online | Meeting physically |

## 3.8. Chapter summary

This chapter discusses how the team came up with their research methodology, design methodology and development along with the risks which the team had to undertake. It also demonstrates how the work was distributed and what software and tools were used in this project.

# Chapter 4: System Requirements Specification (SRS)

## 4.1 Chapter Overview

SRS, or System Requirements Specification is the part of this document that describes the features of the software that is being developed. It revolves around the different elements that delineates the intended functionality required by this research project. It also identifies the relevant stakeholders for this project.

## 4.2. Stakeholder Analysis

### 4.2.1. Onion Model

Diagram

Description automatically generated

Figure 4.1 shows the Onion model of the SRS

### 4.2.2. Stakeholder Descriptions

Table 4.1: Stakeholder descriptions

|  |  |
| --- | --- |
| **Stakeholder** | **Viewpoint** |
| **Functional beneficiary** | |
| System Admins | Will operate the program with the use of EEG detection hardware to help the users and maintain the system as a whole |
| **Social beneficiary** | |
| Public/media | Helps shape the users attitude towards the product |
| **Negative Stakeholders** | |
| Competitors | Intends to create a system that performs/outclasses the said product/system |
| **Regulatory** | |
| Regulators | Wants developers’ system to achieve expectations |
| **Experts** | |
| Mentor/s | Will support developers to produce the desired product to a satisfactory level |

## 4.3. Selection of Requirement Elicitation Techniques/Methods

Requirement elicitation and its different techniques is a cardinal step in determining the requirements that are needed in a software that satisfies its users’ needs and wants, ultimately deciding if it leads to a successful software project or not. Improving elicitation methods and the amount of participation it takes in a software project cycle, improves the prospect of getting closer to users' needs (Hickey and Davis, 2004).

There are several elicitation methods that can be used reliably, these could include:

1. Brainstorming  
   Brainstorming is a technique that involves a group and is aimed to generate new ideas to solve a problem at hand. While brainstorming is not recommended for crucial decisions, it does serve as a great introduction to tackling a problem utilizing ideas from different group members and improving on it as a whole. It is generally open minded and innovative (Mushtaq, 2016).
2. Document Analysis/review

Using existing documents, research papers and past work, not only can true and tried requirements be identified, but also ways to improve or research gaps can be established and built upon. This process, while sometimes long and arduous, may prove to be helpful to any system especially like a research project as this thesis revolves around. Multiple papers and documents would have to be read in order to get to a conclusive decision (Karppinen and Moe, n.d.).

1. Questionnaires

A survey can allow for multiple responses of the same set of questions from different respondents quickly. This also means that the respondent can take their own time to answer the questions, allowing for quantitative data over qualitative. Too many open-ended questions may garner less or no responses, hence why it is better to abide by a select form or multiple-choice questions (Courage and Baxter, 2005).

1. Interviews  
   Interviews are an instrumental tool in any elicitation process. An interview is between two or more people where questions are inquired in order to get insight into a topic. It can be closed (with set questions beforehand which allows for easy analyzation) or open (no predefined agenda and open questions that allow for more ideas and questions to be generated along with brainstorming) or a combination of both which is quite common.

The overall success of an interview would generally come down to two factors, the type of questions asked and the expertise of the interviewee (Suhaib and Iqbal, 2014).

1. Prototyping

Prototyping can help bridge the gap between what stakeholders expect and what the developers think they have to build. By creating a prototype, not only do misconceptions clear, but a better idea about the final product is also given even if the second party is not aware of the technical aspects behind it. Not only do the developers get recommendations on how to improve the system, but also positive features become highlighted and can be refined (Mushtaq, 2016).

## 4.4. Discussion/ Analysis of Results

With the exception of the prototype, all the other requirement elucidation techniques mentioned will be expounded on how they were implemented and a discussion of results of each of them.

1. Brainstorming

With the use of online platforms like Google Meet and Zoom, the team designated for the project brainstormed different innovative requirements like MDD detection and its level that could prove to be useful to the project, such as saving the users results in a profile and database and the possible requirement of sending the data to a neurologist.

1. Document Analysis/review

Extensive research was carried out and after many research papers, MDD detection and its level were two requirements that were identified as the main research component. Using EEG signals as the source and machine learning algorithms MDD and its level can be identified. This was expounded on in Chapter 2.  The feature selection and classification algorithms were also highlighted.

1. Questionnaires

A questionnaire was released on 20th December 2021 and emailed to the students of the University of IIT and others. The questionnaire was open for a span of 3 weeks and the results yielded 83 responses and are shown below.

**Question 1 and 2:**  
Age group and current level of education

Chart, pie chart

Description automatically generated**Results:**

Chart, pie chart

Description automatically generated

**Conclusion:**

Target audience is the majority of the age of 18 - 22 and attends a university.

**Question 3 and 4:**Identifying signs of depression in university students

Chart, pie chart

Description automatically generatedChart, pie chart

Description automatically generated**Results:** Chart, pie chart

Description automatically generated

**Conclusion:**

While it isn't a complete identification of MDD, these signs are general telltale symptoms of depression and even if not are still dangerous. Over 62 percent of the respondents have thought of self-harming thoughts.

**Question 5 and 6:**

Discovering if students have opened-up to people and who.

**Results:** Chart, pie chart

Description automatically generated

Chart, pie chart

Description automatically generated

**Conclusion:**  
While over 50 percent of respondents opened-up to people, they were mostly friends (60.3%) and only a small percentage to parents and doctors (and none to teachers). This indicates that the group of people that can actually help and take action (doctors, parents and teachers) were not the majority of who people opened up to.

**Question 7 and 8:**Reasons to why respondents did not open up to people and if they wished they received support.

**Results:** Chart, pie chart

Description automatically generated Chart, pie chart

Description automatically generated

**Conclusion:**

Many different reasons were listed (with the highest being that they did not want to bother others with their own personal problems). However, a large proportion of respondents answered yes to “Did you wish that you or your friends received some form of support during this time”. This further extrapolates the fact that despite students undergoing problems and possibly MDD, they are demotivated to open up to people who can help them and still wish they received support.

**Question 9:**

Finding out if students believe objective tests like EEG can help identify MDD in early stages can help.

**Results:** Chart, pie chart

Description automatically generated **Conclusion:**

While it is simply an opinion, respondents believe a project like this can help.

**Question 10 & 11:**Identifying how respondents perceive schools’ and universities' attention to mental health of students, and how immediately they believe action should be taken by them.

**Results:** Chart, pie chart

Description automatically generated Chart, waterfall chart

Description automatically generated

**Conclusion:**

Only 6 percent of respondents claimed that universities and schools not only cared about students' mental health but also appointed councilors and supported them, while the rest of the 94% of the respondents believed that either universities and schools simply didn't care or don't have enough resources to help students. They also indicated in the scorn question that action should be taken quite urgently since a disproportionate number of responses were towards 4 and 5 (the scale of 1 to 5 on how urgent they felt it was).

This questionnaire as a whole proves how the target group has signs of depression and how it is a real time problem and students do not get enough support from universities and schools.

## 4.5. Use Case Diagrams

Chart, bubble chart

Description automatically generated

Figure 4.2: Use case diagram

## 4.6. Use Case Descriptions

Table 4.2: Use case description 001

|  |  |  |
| --- | --- | --- |
| **Use Case Name** | MDD Detection | |
| **Use Case ID** | **UC-001** | |
| **Description** | Uses ML algorithm to detect MDD in the user | |
| **Priority** | Critical | |
| **Primary Actor** | User | |
| **Pre-Conditions** | EEG cap worn and connected to relevant PC/laptop | |
| **Trigger** | Actor action | |
| **Main flow** | **Actors** | **System** |
| 1.User loads up application  2.User chooses MDD detection option | 1.Verify EEG Connection  2.Display Results |
| **Exception flow** | **Actors** | **System** |
| 1.User loads up application 2.User chooses MDD detection option | 1.Asks user to connect EEG cap and hardware |
| **Post Conditions** | Program displays results of the test | |

Table 4.3: Use case description 002

|  |  |  |
| --- | --- | --- |
| **Use Case Name** | Login | |
| **Use Case ID** | UC-002 | |
| **Description** | Login and new profile create options | |
| **Priority** | Desirable | |
| **Primary Actor** | User | |
| **Trigger** | Actor action | |
| **Main flow** | **Actors** | **System** |
| 1.User loads up application  2.User chooses login option | 1.Verify password  2.Allow access |
| **Exception flow** | **Actors** | **System** |
| 1.User loads up application 2.User chooses login option | 1.Displays incorrect password |
| **Alternate flow** | **Actors** | **System** |
| 1.User loads up application 2.User chooses new profile option | 1.Creates new profile in database |
| **Post Conditions** | Program logins in the user and saves results under the relevant profile | |

Table 4.4: Use case description 003

|  |  |  |
| --- | --- | --- |
| **Use Case Name** | Database | |
| **Use Case ID** | UC-003 | |
| **Description** | Save and display results from a profile | |
| **Priority** | Desirable | |
| **Primary Actor** | User | |
| **Pre-Conditions** | User logged in/created new profile | |
| **Trigger** | Actor action | |
| **Main flow** | **Actors** | **System** |
| 1.User loads up application  2.User chooses Display past results options | 1.Displays past results in a graph |
| **Exception flow** | **Actors** | **System** |
| 1.User loads up application 2.User chooses Display past results options | 1.Asks user to log in/create new profile |
| **Alternate flow** | **Actors** | **System** |
| 1.User loads up application 2.User chooses Display past results options | 1.Displays error if there are no past results  2. Prompts user to go to MDD detection option |
| **Post Conditions** | Program displays results in a graph along with dates and other relevant details | |

## 4.7. Functional Requirements (with prioritization)

* Critical – The requirements that are critically needed in the successful completion
* Desirable – The requirements that can add value, but are not required immediately
* Luxury – The requirements that would add luxury to the system

Table 4.5: Functional requirements

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Requirements List** | **Priority Level** | **Description** |
| FR1 | User should be able to discover presence of MDD (and export results) | Critical | Along with the EEG Cap and hardware, the user should be able to have the basic function of detecting whether MDD is present or not as this is the most basic functionality of this program |
| FR2 | User should be able to see the level of MDD | Critical | While FR1 has already been implemented in previous research projects, the level of MDD is a feature that has not been explored as much. As the main focus of the research gap in this project this functional requirement is equally as important as FR1 |
| FR3 | User should be able to login and view previous data of results | Desirable | The ability to create profiles and save each user's data separately and securely can prove to be quite useful as users can come back and view their progress and discover if their condition has improved or worsened. This data can also be quite useful to a professional neurologist if the user wills. |
| FR4 | User can view information on what he/she should do next after getting their results | Luxury | Having a comprehensive set of steps and information on what to do next can guide the user on how to get further help/who to contact in order to get possibly a full diagnosis from an actual neurologist or steps in order to start therapy, etc. |
| FR5 | User should be able to view general facts and myths about MDD, along with real symptoms and signs for more information | Luxury | Having a reliable and accurate set of information on MDD would firstly clear misconceptions about MDD and also help users to gain a decent amount of general knowledge about it. |
| FR6 | The user should be able to send the data directly over to a neurologist of choice | Luxury | Being able to have the ability to directly send the information would save time and help ease the process of getting a full diagnosis. |

## 4.8. Non-Functional Requirements

System qualities would include:

Table 4.6: Non-functional requirements

|  |  |  |
| --- | --- | --- |
| **Requirement** | **Importance** | **Description** |
| Reliability | Very high | Should reliably give consistent results with a high accuracy rate |
| Scalability | High | The program should be available to use to many institutions |
| Maintainability | Medium | Should be able to consistently update and add more data to increase accuracy and or reduce |
| Security | High | Data should be secure and the user/patients details and results should be kept confidential only to be accessed by the system admins and the doctors that the information is sent to |
| Usability | Medium | The UI should be intuitive and relatively easy to use without needing prior knowledge about complex EEG data coding, etc. |

## 4.9. Chapter Summary

In summary, this chapter analyzed and discovered the relevant stakeholders to this project, organized them in an onion diagram, and explained each of them. Next through reliable requirement elucidation techniques the key requirements were identified and additionally organized with importance along with a use case diagram. Finally, under non-functional requirements, system qualities were highlighted along with their importance.

# 

# Chapter 5: Social, Legal, Ethical and Professional Issues

## 5.1. Chapter Overview

When exploring a project of such a scope, especially in the field of mental health, it is essential to take care of social, legal, ethical, and professional aspects. This chapter provides a brief description of each issue faced, comparing them with the relevant BCS code of conduct and explains how each of the issues was mitigated.

## 5.2. SLEP Issues and Mitigation

Software safety is a crucial component during implementation and prototyping of a software. During the lifecycle of software development, a number of obstacles and issues are faced. These issues are categorized as social, legal, ethical, and professional issues (Kornecki et al, 2003).

### 5.2.1 Social issues

There are relatively very few social issues of this project, in fact the identification of MDD does not have any apparent adverse social impact at all and could be extremely useful in the mental health of students and a viable tool for doctors. The BCS code of ethics talks about public interest and how software development projects should work for the greater good without any form of harm to the public. The following pointers talk about potential social threats of this project and how they are mitigated.

1. Leakage of personal data - All the data stored in the platform will only be done so with the due consent of the user and the patient. All personal information regarding the patient and the user will not be shared with a third party unless the patient agrees to do so. Third party generally refers to a doctor or a consultant.
2. Discrimination on the grounds of mental issues - Intellignosis will not be a platform that will encourage cyber bullying. All the details stored in the system will be confidential and will not be shared even to the developers.

### 5.2.2 Legal issues

The BCS code of conduct talks about duties to the relevant authority and since this project is based in Sri Lanka, it comes under the local jurisdiction and hence takes special care to adhere to all legal laws and regulations. In addition to this, a main focus was given to data protection laws. The following pointers talk about the legal issues that were faced and the mitigation measures that were put in place.

1. Copyright infringement and plagiarism - All the sources that were used for this report were duly cited and referenced. This includes the datasets that were used, the literature that were reviewed and even online sources that were used to gather information.
2. Leakage of questionnaires and interview data - The project required a survey to obtain information on the opinions of young adults for the usage of this application. None of the email addresses were saved and the responses were kept anonymous to safeguard privacy.
3. Piracy and theft - All the information that was necessary for this project was obtained using legal means. For future implementation, this application is to be delivered as open source. This will reduce the risk of unintentional piracy and malpractice and will help developers improve the application real-time.

### 5.2.3 Ethical issues

In the BCS Code of conduct, professional competence and integrity shows developing your professional knowledge and skills and maintaining awareness of technology developments, procedures, standards that are relevant to your field. Also avoiding disgrace to others’ reputation or employment by malicious actions. The following pointers describe the ethical issues in this research project and how they were mitigated.

1. Unethical survey questions - Before gathering data for the project using an online questionnaire, questions were carefully analyzed considering ethical aspects. Questions that seemed too offensive and unempathetic were removed. After sending the questionnaire to the public, the data was gathered anonymously to respect their opinions and the view on the project.
2. False claims - All the data acquired for this research are properly cited. Ideas by the authors were double checked to avoid false claims, especially regarding facts and statistics.

### 5.2.4 Professional issues

The BCS Code of conduct dictates duty to the profession as encouraging and supporting fellow members. This report was written by the authors with this value and due respect and attribution was given to the members of the computing and medical profession. All the sources were properly cited and referenced, and proper acknowledgement was given. The following pointers describe some of the risks that were faced and the relevant mitigation measures.

1. Disrespect to authors of prior research - Although limitations of prior research were discussed, no offense was meant and they were described in length only at the behest of academic necessity. Other than this, authors and their work were treated with utmost respect.
2. Incorrect dataset attribution - The two datasets obtained for this research paper were created by Wajid Mumtaz and James F. Cavanagh. Their work is properly described and attributed in the relevant places.
3. Improper task delegation and teamwork - All the authors of this paper shared work equally and accomplished the tasks that were assigned to them. A healthy competition was maintained for a quality output, but teamwork was at the core of this research paper

## 5.3. Chapter Summary

This chapter described the social, legal, ethical and professional issues that would be faced during the course of this project and the mitigation strategies that were utilized for a professional outcome.

# 

# Chapter 6: System Architecture and Design

## 6.1. Chapter Overview

This chapter will cover the diagrams for the research topic. The diagrams are System Architecture diagram, class diagrams, sequence diagrams, the UI design, and the process flowchart.

## 6.2. System Architecture Design

During the system design phase, a system architectural design sets the goal architecture, hardware required to support the design, and resolves any performance difficulties during the final implementation (Peters, 2008).

Diagram

Description automatically generated

Figure 6.1 shows the system architecture design

The system architecture design for identifying depression using EEG signals is shown above in Figure 6.1. The system architecture is depicted by the input layer, process layer, and output layer. The input layer specifies that EEG signals from users will be collected and processed. The process layer depicts the EEG signal processing, which extracts features from the shown signal and selects the type of feature. The output layer extracts the feature and classifies the type of depression if it is identified, which is then graded on a scale of tiers (minimal, mild, moderate or severe).

## 6.3. System Design

### 6.3.1. Class Diagram

Class diagrams represent information about a domain in terms of objects arranged into classes and their relationships (Berardi, Calvanese and De Giacomo, 2005).

Diagram

Description automatically generated

Figure 6.2 shows the class diagram

The class diagram for identifying depression using EEG signals is shown in Figure 6.2 above. Classes are described briefly below. Intellignosis (System) receives the user's credentials, verifies them against the data stored in the institution's database, and then displays them. The system then receives the user's EEG signal and determines whether the user has MDD. This is then processed and stored in a table as patient information and in the EEG report. The patient details table contains the user's information as well as the results. The EEG report table contains information about the EEG signal as well as its classification. This will be done for each individual user.

### 6.3.2. Sequence Diagram

The sequence diagram is generally used to depict item interactions in the order in which they happen. Developers, like class diagrams, often believe sequence diagrams are exclusive for them.

(Bell, 2004)

Diagram

Description automatically generated

Figure 6.3 shows the sequence diagram

The sequence diagram for identifying depression using EEG signals is shown in Figure 6.3 above. First, the user logs in with credentials, which are then validated by the system. If the credentials are incorrect, an error message is displayed. If the credentials are correct, it detects the user's EEG signal, which is then analyzed to identify MDD and provides results, including the level of depression. Finally, it processes the results, which the user can view.

### 6.3.3. UI Design

Rapid prototyping is sped up by low-fidelity fabrication techniques, which print intermediate versions of a prototype as quick, low-fidelity previews. Only the final form is printed as a full-scale, high-resolution 3D model. Designers may iterate faster as a result, resulting in a better design in less time (Mueller et al., 2015). Software designers and developers utilize high-fidelity prototype tools to work out interface specifics without committing to a final implementation (Li, Tigwell and Shinohara, 2021). In this document, a high-fidelity prototype will be used.

Graphical user interface, application, Teams

Description automatically generated

Figure 6.4 shows the Dashboard screen of the UI design

Graphical user interface, text, application

Description automatically generated

Figure 6.5 shows the Results screen of the UI design

### 6.3.4. Process flow chart

Diagram

Description automatically generated

Figure 6.6 shows the process flow chart

The system process flowchart is depicted in Figure 6.6. It represents the system's transition from one process to the next. It can help depict the direction of data flow and how decisions are made to control the events. The arrows depict the direction, rectangles indicate a process, parallelograms show processes taking place and a rhombus conveys a decision branching out to multiple possibilities of where the program can go next.

## 6.4. Chapter Summary

This chapter went over the different architecture and design aspects of the project at hand. The system architecture design, class diagram, sequence diagrams, process flowchart, and user interface design were highlighted and expounded on.

# 

# References

1. Acharya, U.R., Sudarshan, V.K., Adeli, H., Santhosh, J., Koh, J.E.W. and Adeli, A. (2015a). Computer-Aided Diagnosis of Depression Using EEG Signals. *European Neurology*, 73(5-6), pp.329–336.
2. Acharya, U.R., Sudarshan, V.K., Adeli, H., Santhosh, J., Koh, J.E.W., Puthankatti, S.D. and Adeli, A. (2015b). A Novel Depression Diagnosis Index Using Nonlinear Features in EEG Signals. *European Neurology*, 74(1-2), pp.79–83.
3. Akbari, H., Sadiq, M.T. and Rehman, A.U. (2021). Classification of normal and depressed EEG signals based on centered correntropy of rhythms in empirical wavelet transform domain. *Health Information Science and Systems*, 9(1).
4. Al-Fahoum, A.S. and Al-Fraihat, A.A. (2014). Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains. *ISRN Neuroscience*, [online] 2014, pp.1–7. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4045570/ [Accessed 7 Apr. 2019].
5. Aydemir, E., Tuncer, T., Dogan, S., Gururajan, R. and Acharya, U.R. (2021). Automated major depressive disorder detection using melamine pattern with EEG signals. *Applied Intelligence*, 51(9), pp.6449–6466.
6. Bairy, G. M., Niranjan, U.C. and Puthankattil, S.D. (2016). Automated classification of depression EEG signals using wavelet entropies and energies. *Journal of Mechanics in Medicine and Biology*, 16(03), p.1650035.
7. Beck, A.T. (1961). An Inventory for Measuring Depression. *Archives of General Psychiatry*, [online] 4(6), pp.561–571. Available at: https://counsellingresource.com/lib/wp-content/managed-media/beck\_at\_1961.pdf.
8. Cai, H., Han, J., Chen, Y., Sha, X., Wang, Z., Hu, B., Yang, J., Feng, L., Ding, Z., Chen, Y. and Gutknecht, J. (2018). A Pervasive Approach to EEG-Based Depression Detection. *Complexity*, [online] 2018(5238028), pp.1–13. Available at: https://www.hindawi.com/journals/complexity/2018/5238028/ [Accessed 30 Jan. 2020].
9. Chandra, V. (2015). Comparison between Various Software Development Methodologies Related papers Multilevel Security Spiral (MSS) Model: NOVEL Approach Hat em Abu-Hamat t ah Multilevel Security Spiral Model: Novel Approach Hat em S. A. Hamat t a A Comparative Study of Universally Accept ed SDLC Models for Software Development Comparison between Various Software Development Methodologies. *International Journal of Scientific Research in Science and Technology IJSRST*, 131(9), pp.975–8887.
10. Čukić, M., Stokić, M., Simić, S. and Pokrajac, D. (2020). The successful discrimination of depression from EEG could be attributed to proper feature extraction and not to a particular classification method. *Cognitive Neurodynamics*, [online] 14(4), pp.443–455. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7334335/ [Accessed 29 Dec. 2021].
11. Cusin, C., Yang, H., Yeung, A. and Fava, M. (2010). Handbook of Clinical Rating Scales and Assessment in Psychiatry and Mental Health, Current Clinical Psychiatry. In: *Handbook of Clinical Rating Scales and Assessment in 7 Psychiatry and Mental Health, Current Clinical Psychiatry*. Springer Science+Business Media.
12. Eraldemir, S. G., Kilic, Ü., Kaya Keles, M., Demirkol, M. E., Yildirim, E. and Tamam, L. (2020). Classification of EEG Signals in Depressed Patients. *Balkan Journal of Electrical and Computer Engineering*, 8(1).
13. Fay, B. (2018). *Hospital and Surgery Costs – Paying for Medical Treatment*. [online] Debt.org. Available at: https://www.debt.org/medical/hospital-surgery-costs/.
14. Fulghum, D. (2008). *Depression Diagnosis*. [online] WebMD. Available at: https://www.webmd.com/depression/guide/depression-diagnosis.
15. Guyon, I. and Elisseeff, A. (2006). An Introduction to Feature Extraction. *Feature Extraction*, pp.1–25.
16. Hosseinifard, B., Moradi, M.H. and Rostami, R. (2013). Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Computer Methods and Programs in Biomedicine*, 109(3), pp.339–345.
17. Kang, M., Kang, S. and Lee, Y. (2021). Prediction of Beck Depression Inventory Score in EEG: Application of Deep-Asymmetry Method. *Applied Sciences*, 11(19), p.9218.
18. Kornecki, A. (2003). Software Safety - Ethics, Professionalism, and Legal Issues. *Proceedings Of The 21st International System Safety Conference - 2003*, pp.952–961.
19. Kotsiantis, S.B. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, [online] 26(3), pp.159–190. Available at: http://www.informatica.si/index.php/informatica/article/viewFile/148/140.
20. Kumar, J.S. and Bhuvaneswari, P. (2012). Analysis of Electroencephalography (EEG) Signals and Its Categorization–A Study. *Procedia Engineering*, 38, pp.2525–2536.
21. Lai, C.Q., Ibrahim, H., Abdullah, M.Z., Abdullah, J.M., Suandi, S.A. and Azman, A. (2018). Literature survey on applications of electroencephalography (EEG).
22. Mahato, S., Goyal, N., Ram, D. and Paul, S. (2020). Detection of Depression and Scaling of Severity Using Six Channel EEG Data. *Journal of Medical Systems*, 44(7).
23. Mahato, S. and Paul, S. (2019). Classification of Depression Patients and Normal Subjects Based on Electroencephalogram (EEG) Signal Using Alpha Power and Theta Asymmetry. *Journal of Medical Systems*, 44(1).
24. Mantri, S., Patil, D., Agrawal, P. and Wadhai, V. (2015). Non Invasive EEG Signal Processing Framework for Real Time Depression Analysis. SAI Intelligent Systems Conference 2015. London: IEEE.
25. Mohammedi, Y., Hajian, M. and Moradi, M.H. (2019). Discrimination of Depression Levels Using Machine Learning Methods on EEG Signals. 27th Iranian Conference on Electrical Engineering (ICEE2019). pp.1765–1769.
26. Mohan, Y., Chi, S.C., Xin, D.K.P. and Foong, L.P. (2016). Artificial Neural Network for Classification of Depressive and Normal in EEG. 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES). pp.286–290.
27. Mumtaz, Wajid (2016): MDD Patients and Healthy Controls EEG Data (New). figshare. Dataset. <https://doi.org/10.6084/m9.figshare.4244171.v2>
28. Mumtaz, W., Xia, L., Ali, S.S.A., Yasin, M.A.M., Hussain, M. and Malik, A.S. (2017). Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD). *Biomedical Signal Processing and Control*, 31, pp.108–115.
29. Nemeroff, C.B. (1998). *The Neurobiology of Depression*. [online] Scientific Americal Inc. Available at: https://www.gatewaypsychiatric.com/pdf/0698042B.pdf.
30. Noble, W.S. (2006). What is a support vector machine? *Nature Biotechnology*, 24(12), pp.1565–1567.
31. Nussbaumer, H.J. (1981). 4. The Fast Fourier Transform. In: *Fast Fourier Transform and Convolution Algorithms*. Springer-Verlag Berlin Heidelberg, pp.81–111.
32. Peterson, L. (2009). K-nearest neighbor. *Scholarpedia*, [online] 4(2), p.1883. Available at: http://www.scholarpedia.org/article/K-nearest\_neighbor [Accessed 21 Aug. 2019].
33. Puthankattil, S.D. and Joseph, P.K. (2012). Classification Of Eeg Signals In Normal And Depression Conditions By Ann Using Rwe And Signal Entropy. *Journal of Mechanics in Medicine and Biology*, 12(04), p.1240019.
34. Saeedi, M., Saeedi, A. and Maghsoudi, A. (2020). Major depressive disorder assessment via enhanced k-nearest neighbor method and EEG signals. *Physical and Engineering Sciences in Medicine*, 43(3), pp.1007–1018.
35. Tes. (n.d.). *What is EEG and will it be used in schools? (sponsored)*. [online] Available at: https://www.tes.com/news/what-eeg-and-will-it-be-used-schools.
36. www.epilepsy.org.uk. (2020). *EEG tests and epilepsy | Epilepsy Action*. [online] Available at: https://www.epilepsy.org.uk/info/diagnosis/eeg-electroencephalogram.
37. Zhang, S., McCane, B., Shadli, S. and Mcnaughton, N. (2020). *Trait depressivity prediction with EEG signals via LSBoost 5 th*. IEEE.

# Appendix

Each section under the appendix should be named “Appendix A”, “Appendix B” etc.

For this the page numbering should be Uppercase roman numerals