T.C.
BAHÇEŞEHİR UNIVERSITY

FACULTY OF ENGINEERING AND NATURAL SCIENCES

DEPARTMENTS OF BIOMEDICAL AND SOFTWARE ENGINEERING

EEG DATA ANALYSIS FOR DETECTING RELAXATION AND CONCENTRATION MOODS

Capstone Project Proposal

Ilhan Ozan Aral

Bensu Sengel

Can Yılmaz

Muharrem Asan

Advisors:

Dr. Burcu Tunç Çamlıbel
Asst. Prof. Betül Erdoğdu Şakar

ISTANBUL, September 2020

STUDENT DECLARATION

By submitting this report, as partial fulfillment of the requirements of the Capstone course, the students promise on penalty of failure of the course that

- they have given credit to and declared (by citation), any work that is not their own (e.g. parts of the report that is copied/pasted from the Internet, design or construction performed by another person, etc.);
- they have not received unpermitted aid for the project design, construction, report or presentation;
- they have not falsely assigned credit for work to another student in the group, and not take credit for work done by another student in the group.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SEN	Software Engineering
BME	Biomedical Engineering
EEG	Electroencephalogram
ML	Machine Learning
LDA	Linear Discriminant Analysis
SVM	Suppor Vector Machine
RF	Random Forest
DT	Decision Tree
NB	Naive Bayes
AI	Artificial Intelligence
ERP	Event Related Potential
ICA	Independent Component Analysis
EU	European Union

1. OVERVIEW

Electroencephalogram (EEG) is a method used to measure the electrical activity at the peripheral part of brain. These activities are measured by placing number of electrodes across the scalp of the subject. However, it is almost impossible to understand what these measurements mean, if we look at only the raw signal data. Because there are many factors which can affect the signals while performing the measurement. That is why we need to detect these bad signals filter them out and utilize several techniques for processing the signals to make them understandable. EEG sensors do not penetrate the body of the subject, therefore it cannot detect most of the activities that are away from the surface of the scalp. However, since this technique doesn't require invasive operations it is relatively easier to perform on many subjects and yet it can still acquire many useful information about the brain.

1.1. Identification of the need

Due to the limitations of the EEG technique we can obtain very limited information about the brain. However by utilizing several signal processing techniques we can obtain a clean data and by using machine learning techniques, we can make the most sense from the signals we gathered. This is an under-development research topic around the globe. Many researches have been made and many are needed to be made for a better understanding of the human brain. We saw an opportunity in the literature to make our contribution to the field.

1.2. Definition of the problem

Scientists have been doing many types of researches to understand how does the human brain work. That is still not a fully understood concept by science. EEG have been around since decades, but making sense of the signals gathered with this technique gave us little clue about the activities of the brain. Thanks to the latest developments in the field of computer science also new types of techniques have been developed for processing signals. One of the most important development of the field is the machine learning algorithms. Thanks to this method we can make hundreds of thousands of measurements and run a computer program to make sense of these signals in a very short period of time. Without this technology we would have to look at the plotted signals by eye and try to understand each signal one by one to come to a conclusion. Figure 1 shows the different steps of our work.

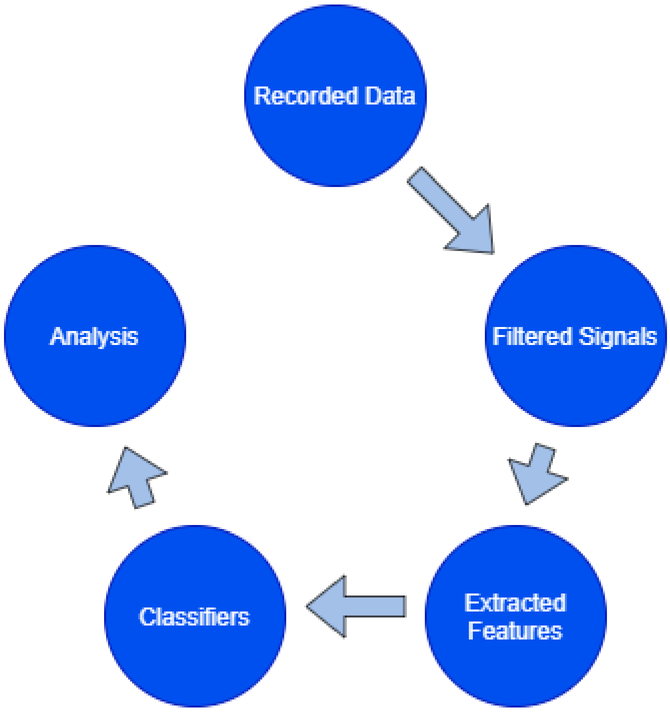


Figure 1 Schematic View of the System

In addition, Figure 2 shows the recorded data that have not been filtered yet. Displayes inly the first 10 seconds of data with 14 different channels.

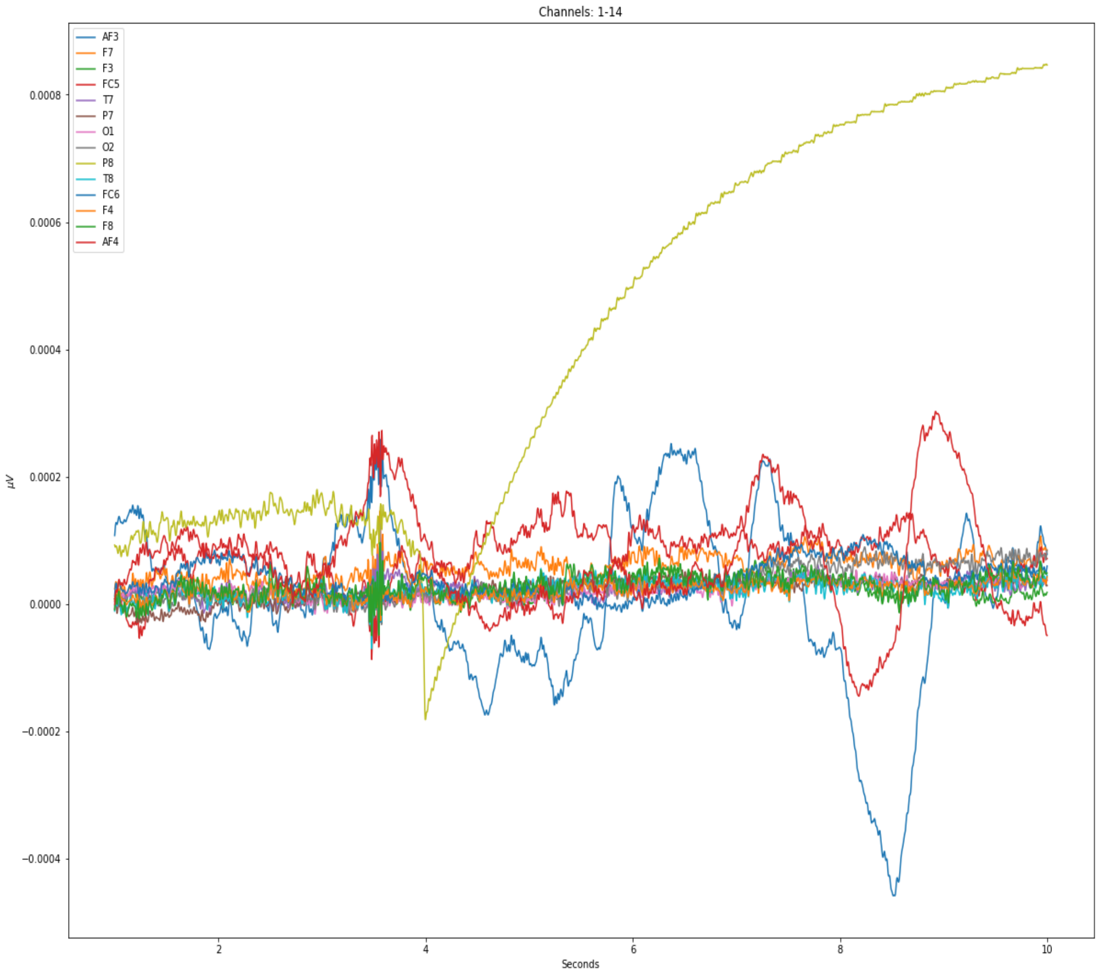


Figure 2 The graph of frequency by seconds

1.3. Standards and constraints

On the physical side of the project, we need to use an EEG sensor for collecting sample data from different subjects for several sessions. Unfortunately, due to the pandemic conditions it is not possible to work with lots of subjects and do tests on them. For this reason, we used already collected and shared data available in the literature. Therefore, physical side of the project will not be evaluated by its effects.

On the software side of the project, it is expected that a program will be created and this program will analyze the data and it will reveal results. This program will only run on computers and its effect on the environment are examined:

- **Environmental Effects:** Computers work with electricity power. The system will not harm the ecosystem.

- **Social Effects:** People can waste time dealing with a lot of data or there may be parts that get overlooked because there is too much data. Since the system can complete the given job within a few seconds, it positively affects the social lives of people.

- **Economic Effects:** Since the system's uptime will shorten and people will not spend too much time analyzing the data, it is anticipated that the budget allocated to research will be used less. Also, some advanced expensive programs will be less needed for EEG analysis.

- **Ethical Effects:** The system will provide the first rule of the code of ethics that is published by IEEE because it aims to decrease the time to spend and increase the rate of success of the human work.

- **Health and Safety:** The project is created to analyze and reveal an answer with an accuracy. It is necessary to find a method with more accuracy than others to decrease the error rate. But the whole system is designed to find the secret inside of the human brain and therefore it is possible to understand that the system increase human health and safety.

1.4. Literature Review

Humans have been doing research about the brain for a long time, but still do not fully understand the brain. In these studies, an EEG device is generally used to examine the brain, but although this process has a cost and margin of error, the research can take a long time.

Machine Learning (ML) is a system that can learn as structural functions and search algorithms and constructs that can make predictions on data. These systems will have positive effects in terms of cost and time as they can replace people. Nowadays, with the increase of people's research in this field, the place of ML in our Daily life is increasing. For this reason, in this Project, using ML with EEG, a system will try to predict people's emotional states. In ML, appropriate algorithms should be chosen depending on the projects. Classification algorithm should be used in this Project because the desired result in the project is to understand whether the person is in a focused state or in a relaxed state.

[13] For example, in order to understand whether a person is happy, angry, tired or relaxed, the ratios of arousal and valence values are examined. LDA and SVM methods were used for these two values. [1] In another study, people were watched some videos and asked to fill out a questionnaire containin arousal, valence, familiarity and dominance values fort he situations they experienced while watching and used 5 different classifiers,it was tried to predict exactly which emotions the person felt. [6] In another study, people were asked 6 questions. Answers to these questions will be answered positively or negatively and each answer represents a feeling about the person. As a result of this proces, they want to guess what the heart rate is by looking at the emotion. Researchers use the classifier methods of classification together with the deep learning methods to test. As a resultm they choose the most accurate result. [14] In another study, people are shown a video and at the end of the video they are

asked to give a value between 1 and 9 for the arousal, valence, dominance and liking values they feel. The methods used are SVM and RF because it is stated that these methods can learn quickly and correct predictions are made with small data.

In this proposal, [2] 4 sessions of 180 seconds each were held with 30 people. Each session was held at different times. During the first minute of the session, the person is asked to be confortable, their eyes closed and seated on a chair. On the second minute, the person encounters random pictures or questions on the screen. At the last minute, the person is asked to relax again and his/her eyes are closed. Later, these datas are cleaned and extracted features by BME students. In this way, it is ready before ML is implemented. Ready data is received by the SEN students and goes to the cleaning phase. Then, a total of 21600 seconds of data is divided into test and train parts. The methods to be used are SVM, RF, DT, LDA and NB methods. Because some of them make quite accurate results and some are very popular. Then the results are compared with each other and the appropriate method is selected. Finally, the test data is processed and the project is made ready.

1.5. Physical architecture

Figure 3 Emotiv Epoc Flex 32-Channel

On the physical side of the project, we use an EEG device manufactured by EMOTIV. The model we are going to use is Emotiv FLEX 32-Channel. This device is carefully prepared and each electrode is placed to the right position on the Emotiv EasyCap for neglecting false placement errors. Color coded wires gets connected to the controller on its' each sides. Then the cap is placed to the subject carefully and conductive gel is applied to each electrode one by one. This process usually takes 3 to 5 mins. After checking everything is wired up correctly, we can connect the controller to the computer. By using the EmotivPRO software we can check if the device is working as expected. Now we can start our session depending on the research and log the data of the session in ".edf" file format.

On the software engineering side of the project, it starts with a preparation process after receiving the data with extracted features. In this process, the data are changed and fragmented to work on a model. Then, the models to be worked on are put into the training process one by one. A prediction is provided with the trained data and all models go through the same process. As a result, a model with best accuracy is selected and the system becomes ready for prediction.

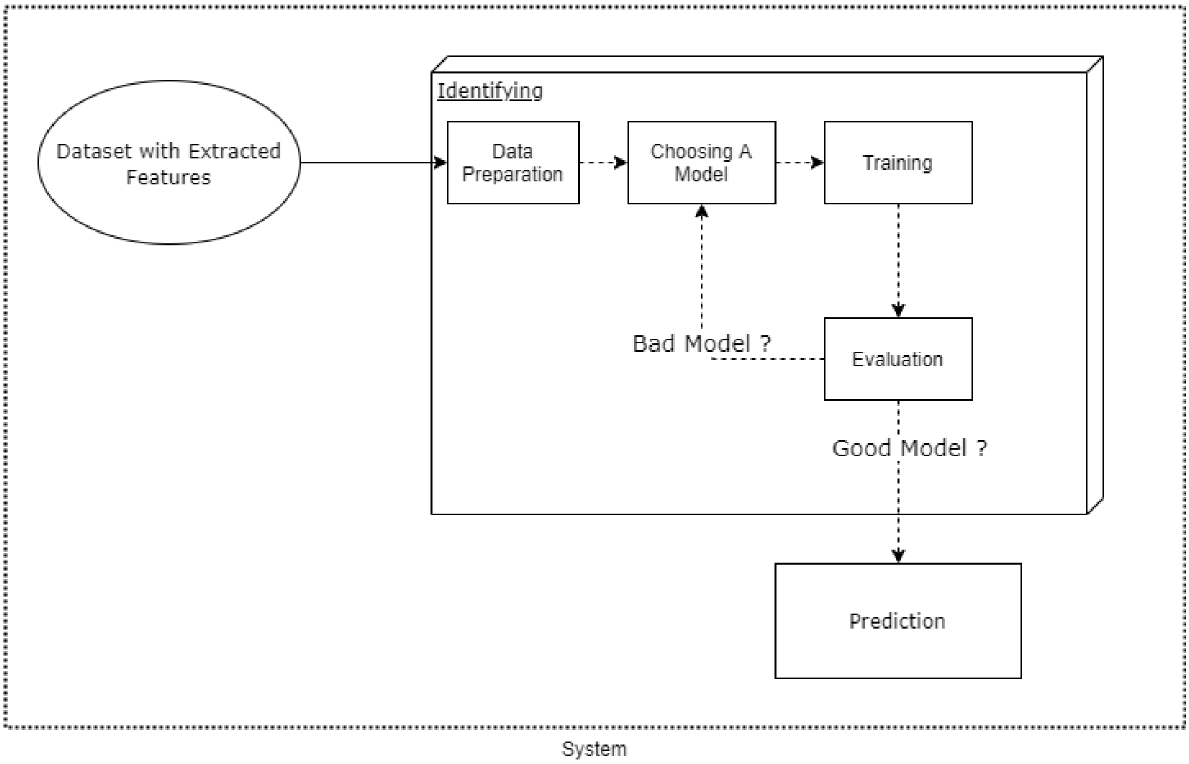


Figure 4 Physical Structure of the System

2. WORK PLAN

2.1. Work Breakdown Structure (WBS)

1. System
    - 1.1. Data
        - 1.1.1. Data Gathering
        - 1.1.2. BME-Preprocessing
            - 1.1.2.1. Event Markers
            - 1.1.2.2. Plotting the Raw Data
            - 1.1.2.3. Filtering the Signal
            - 1.1.2.4. ICA Decomposing
            - 1.1.2.5. Examination by Eye
    - 1.2. Software
        - 1.2.1. Dataset Preprocessing
            - 1.2.1.1. Data Visualization
            - 1.2.1.2. Data Preprocessing
                - 1.2.1.2.1. Data Cleaning
                - 1.2.1.2.2. Data Transformation
        - 1.2.2. Data Splitting
            - 1.2.2.1. Coding
        - 1.2.3. Modeling
            - 1.2.3.1. Searching Models
            - 1.2.3.2. Assembling and Coding
            - 1.2.3.3. Evaluating
        - 1.2.4. Accuracy Improvement
            - 1.2.4.1. Testing and Coding
    - 1.3. Interface
        - 1.3.1. Design
            - 1.3.1.1. Layouts
            - 1.3.1.2. Screens
        - 1.3.2. Implementation of Coding
        - 1.3.3. Testing

Figure 5 Work Breakdown Structure of the System

2.2 Responsibility Matrix (RM)

Tasks	Ozan	Bensu	Can	Muharrem
Data Gathering	C	C	-	-
Event Markers	S	R	-	-
Plotting the Raw Data	R	S	-	-
Filtering the Signal	R	S	-	-
ICA Decomposing	S	R	-	-
Examination by Eye	R	S	-	-
Preprocessing	-	-	R	-
Data Splitting	-	-	R	S
Modeling	-	-	R	S
Acc. Improvement	-	-	S	R
Design	-	-	S	R
Implementation	-	-	R	S
Testing	-	-	S	R
Planning	S	R	R	S
Reporting	S	S	S	R
Integration	S	S	R	S

Table 1 Responsibility Matrix

2.3. Project Network (PN)

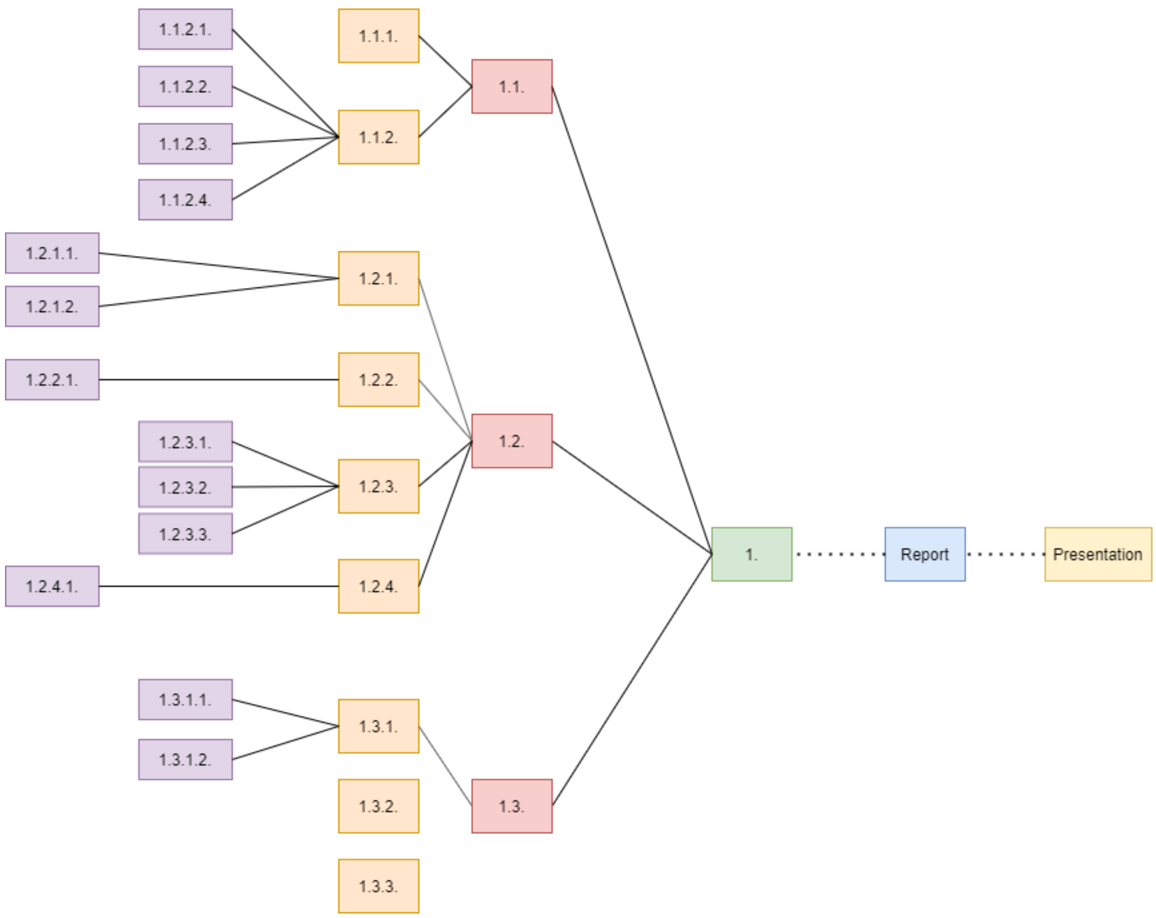


Figure 6 Project Network of the System

2.4. Gantt chart

Week	1	2	3	4	5	6	7	8
Ozan	1.1.1.			1.1.2.2.		1.1.2.3.		
Bensu			1.1.2.1.				1.1.2.4.	
Can	1.2.1.1.	1.2.1.2.		1.2.2.	1.2.3.2.	1.2.3.3.		1.3.2.
Muharrem	1.3.1.1.		1.3.1.2.	1.2.3.1.		1.2.4.1.		
Integration								
Verification								
Documentation								
Presentation								

Week	8	9	10	11	12	13	14	15
Ozan	1.1.2.5							
Bensu		1.1.2.5						
Can	1.3.2.							
Muharrem		1.3.3.						
Integration								
Verification					1			
Documentation						Draft	Report	
Presentation							Draft	Presentation

Figure 7 Timeline of the Project

2.5. Risk Assessments

Failure Event	Probability	Severity	Risk Level	Plan of Action
Low accuracy	Likely	Major	VERY HIGH	Selecting another model or changing splitted data
Lack of data	Possible	Moderate	MEDIUM	Finding new datas or creating random datas
Improper or incomplate training	Possible	Major	HIGH	Wrong coding, fixing the issues

Table 2 Risk Assessments of the Project

3. DESIGN PROCESS

3.1. Biomedical Engineering

3.1.1 Task clarification

During the EEG signal gathering sessions, we need to prepare the device with care. After the preparation we need to put it on the subject and dispense conductive gels to each electrode carefully. We need to check if everything works properly by observing the signals from the EmotivPRO software interface. We proceed with the designed session and save the recordings. When the session is over we need to remove the cap and clean all the electrodes using alcohol for the next session. After all the sessions are over and the kit is cleaned we need to take apart each component and pack them according to the procedure and store the device in the required place.

After the physical tasks are completed we will proceed on the software side. We will use both Matlab and Phyton for processing the signal and detecting the moods of the subjects. Two groups will create algorithms; one group using Matlab and the other using Phyton. Both groups are responsible of the following steps:

-Detect and remove artifacts in raw data
-Filter raw data based on frequency range of interest
-Perform time-frequency analysis

-Divide the clean signal into 4 bands (alpha, beta, theta, delta)
-Visualize the data by plotting
-Develop an algorithm to detect moods

After both of the groups complete the software side of the project, we will compare both of the approaches and report accordingly.

3.1.1.1 Definition of the problem and objectives

Our objective during the sessions are to keep the sensor well connected and make sure each electrode is in contact with the sculp of the subject. We need to keep the session room clean, quiet and dark (depending on the aim of research) to keep away any sort of disturbances which may affect the signals.

3.1.1.1 Requirement specification

i. Technical requirements (the functional and performance requirements of the product)

Preparation of functional analysis tool (FAST) stating the problem in terms of functions and is intended to identify functional and performance requirements.

ii. Requlatory requirements (governmental regulations, international standards, or product recommendations)

All medical devices must be approved for (CE) marking by a Notified Body before they are released to market. Devices fall into certain type categories. Broadly speaking, for each market area and for each type of device there are internationally and nationally agreed requirements for which the device must be designed to conform.

There are three types of regulatory requirements standards:

1. Horizontal standards - general regulations that apply to all medical devices sold and used in a particular market area. Examples for Europe: Medical Device Regulations, EN ISO 9000 Quality Management Systems Standards, EN ISO 14971 Application of risk management to medical devices. In the USA, the Code of Federal Regulations is the master reference source for regulatory information.

2. Semi-horizontal standards - particular to a family of devices, such as electrical devices, and regulating certain aspects of the devices such as toxicity, material, or safety. Example: BS EN 60601-1 Medical electrical equipment - General requirements for safety

3. Product-specific or vertical standards - apply to a particular device or piece of equipment. Examples: Rehabilitation Engineering and Assistive Technology Society of North America (RESNA) standard WC93-1991 sets maximum wheel chair dimensions, BS EN 60601-2-16 specifies particular requirements for the safety of haemodialysis equipment.

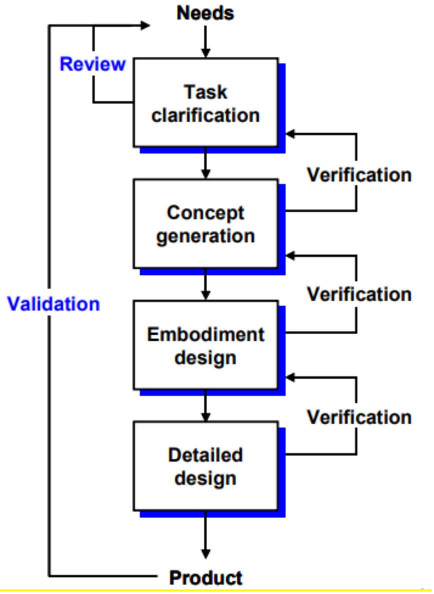
iii. Device Classification

Medical devices are classified into groups by their potential risk and the dangers they pose to users and/or patients. The classification of the device determines the level of intervention by the relevant regulatory authorities in controlling the design and production of the device.

Class 1 - low risk healthcare devices, for example, spectacles, hospital beds and tongue depressors

Class 2 - medium risk devices The EU regulatory system splits Class 2 devices further into 2a and 2b, with the latter being higher risk. Class 2a covers short-term invasive devices, for example, endoscopes and hypodermic needles. Class 2b covers devices that impart energy in a potentially hazardous way, for example, X-ray machines.

Class 3 - high-risk devices This relates to long term invasive devices that come into contact with the central nervous system or the circulatory system.



3.1.2 Concept generation

In the brain signals obtained using EEG, signals with lower frequencies compared to the signals that normally occur in the brain and that is not easily read by some external conditions are obtained. To purify these data from artifacts and to be easily readable/understandable, with the commands in MATLAB / EEGLAB, these are it is done with easily executable commands such as filtering data, applying ICA, removing bad channels. For this project, one of the possible solution for the problem found by us is to provide continuity by using MATLAB algorithms to reduce these processes to a briefer time.

3.1.3 Detailed design

The software program we use for the first step of the project is MATLAB.

The data we are working on is EEG data, and it has been measured with Emotiv Epoc +.

3.1.4 Data Gathering

Following dataset has been collected and shared online with the public for scientific purposes. This dataset gave us the opportunity to test our signal processing methods before working with any real life subjects. Information about the dataset are as follows:

Title: EEG dataset of Fusion relaxation and concentration moods
DOI: 10.17632/8c26dn6c7w.1
By: Ahmed Albasri
URL: https://data.mendeley.com/datasets/8c26dn6c7w/1
Data type: EDF
Device: EMOTIV EPOC+

Description of this data

1. **Aim:**
This dataset aims to provide open access of raw EEG signal to the general public. We believe that such fusion of human moods (Relaxation & concentration) shall increase scientific transparency and efficiency, promote the validation of published methods, and foster the development of new algorithms. In addition, publishing research data is becoming more important as public funding agencies are moving towards open research data requirements.
2. **Scenario:**
The proposed scenario adapted to acquire the brain EEG signals in two different mental status. First while subjects in a relaxed mood, and second in concentration mood. Both of these cognitive stimuli considers as self-induced motivation. The recording period continues till three minutes for each session, as follows:
-In the first minute, the subject is asked to relax and sit on a handed chair with eye open looking at a black screen computer of about 40cm far. Until hearing beep sound.
-In the second minute, a random picture appear on the screen contain a question or some different objects. The subject is asked to solve the problem or to find common relation links all these objects together.
-In last minute, the subject is asked to close his/her eyes and relax again until the beep sound.
3. **Sessions:**
Fore sessions were recorded for each subject. Such that, first two sessions are done on the same day with 1-2 hours interval, and remaining sessions are done after 2-3 days in the same way. The reason behind this separation is to avoid medium term influences that may subjects have. Each session continues for three minutes. The total recording time for each subject equal to 720 seconds. A small program designed to control the timing and recording procedure of the sessions.
4. **Numbering system:**
The numbering system is formatted to include both subject enrollment number and trials. First four characters represent the subject number, where last three characters represent the session record number. For example (S001E03) indicate 1st subject and 3rd recording session.
5. **Artifacts:**
In this experiment, we notice that some subjects accidentally generated internal artifacts. Therefore we intentionally continue recording their brain signals to provide more realistic condition to the experiment and also provide a role for the artifact removal techniques in the pre-processing phase.
6. **Data recording:**
EEG raw data recorded using EMOTIV EPOC+ device with 14 channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF42), plus References in the CMS/DRL noise cancellation configuration P3/P4 locations. The signals were sampled with 250 SPS.
7. **Sample space:**
The sample space consists of 30 participants (56.6% male and 43.3% female) with ages of 18-40 years. The subjects do not suffer(ing/ed) from any brain problems (mentally or physiologically). 33% of the subjects were smokers and 3% of them were alcoholics. All the subjects are well educated and have at least B.S degree.

3.1.5 BME - Preprocessing

For processing the gathered data we used a Matlab plugin software specifically developed for EEG signal processing called EEGLAB. We used the software versions “Matlab - R2020b” and “EEGLAB - 2020.0”.

EEGLAB contains many functions needed to analyze, process and visualize signals. Screenshot of the plugin can be seen in the figure. After importing the dataset to EEGLAB the interface shows important information about the dataset being used. We can utilize the “Tools” section for performing several operations on the signals or utilize the “Plot” section for different visualization options .

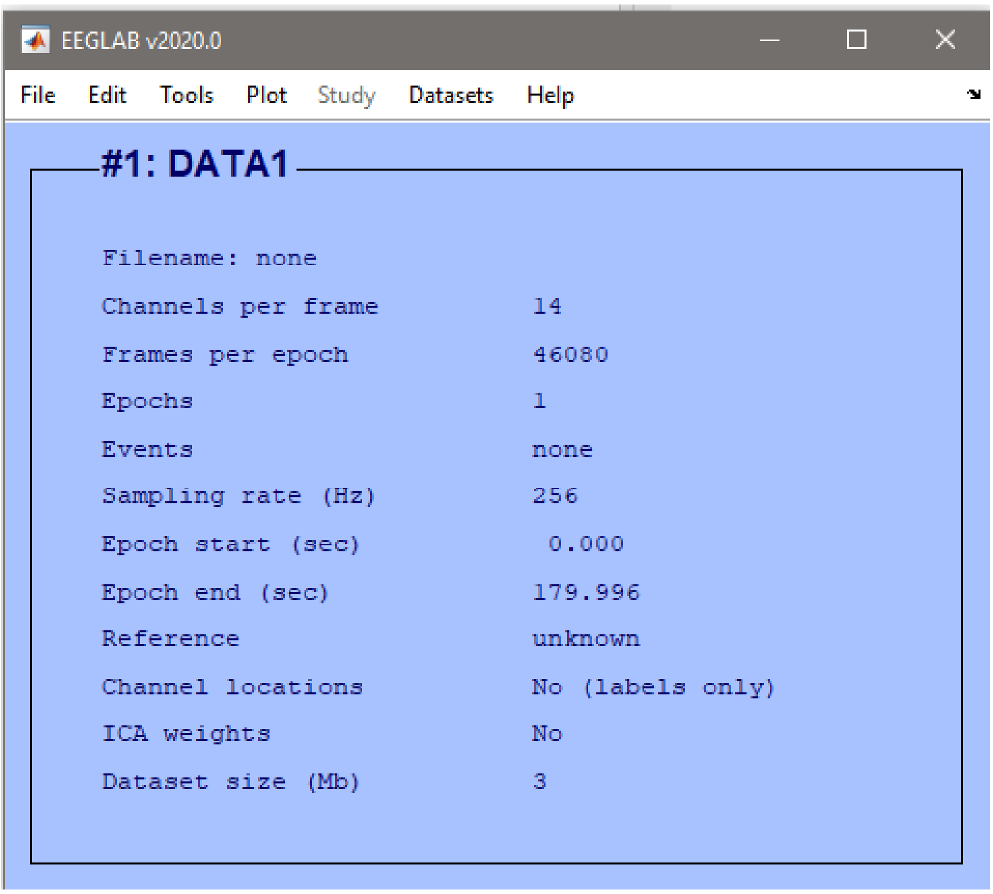


Figure 8 Screenshot of the interface after importing the raw dataset

We also need to import the channel locations for topographic visualization option. After importing channel locations the interface will change its' "Channel locations status" to "Yes". We still need to make sure that the channel locations have been imported right by comparing the topographic plot with the device manual.

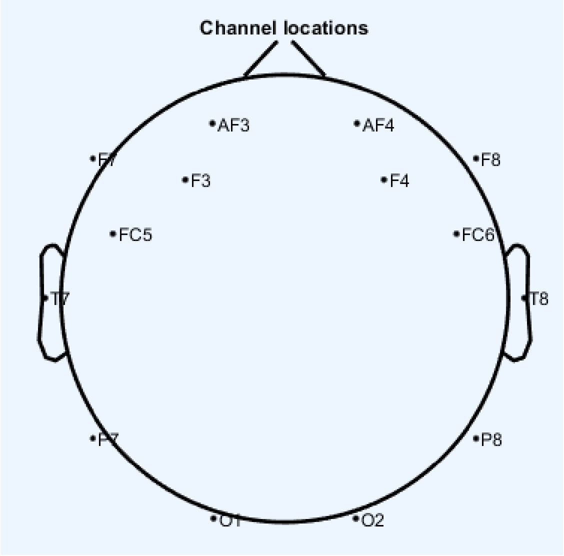


Figure 9 Figure 7. Channel locations plot for EEGLAB

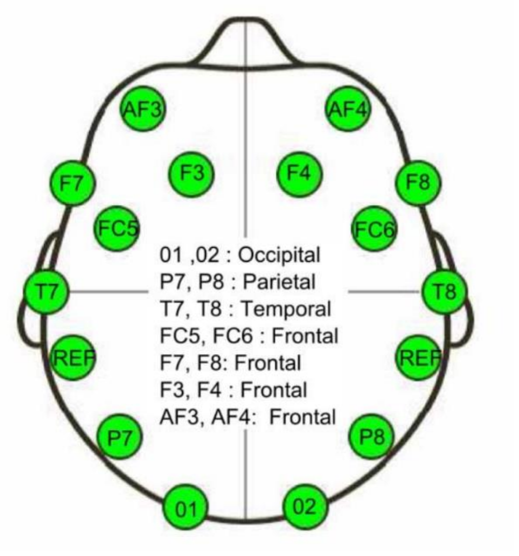


Figure 10 Channel locations of Emotiv Sensor

3.1.5.1 Plotting the Raw Data

Initially we plot all the channels in one visual for first observations about the quality of the signal.

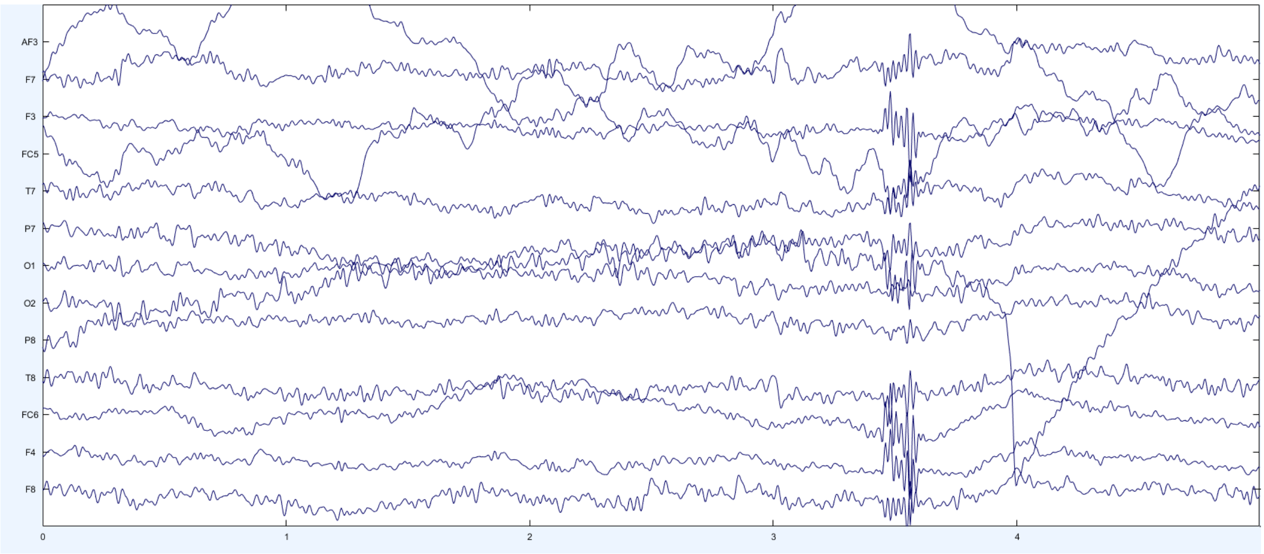


Figure 11 Raw data of 5 seconds

By changing the scale rate we can differentiate the channels and observe the signals better

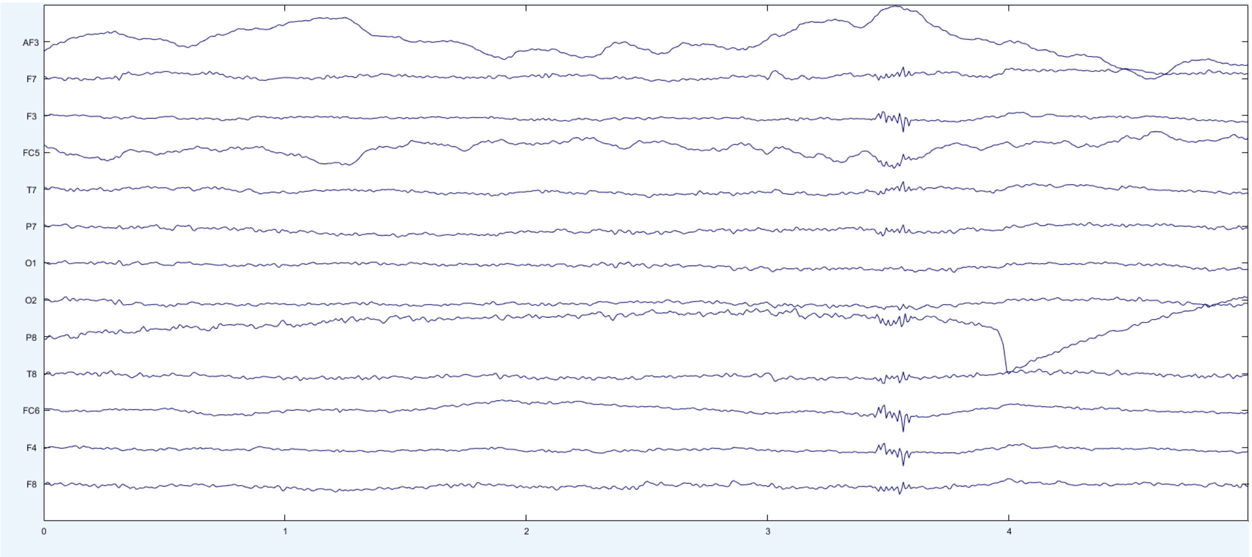


Figure 12 Raw data 5s sample with edited scale

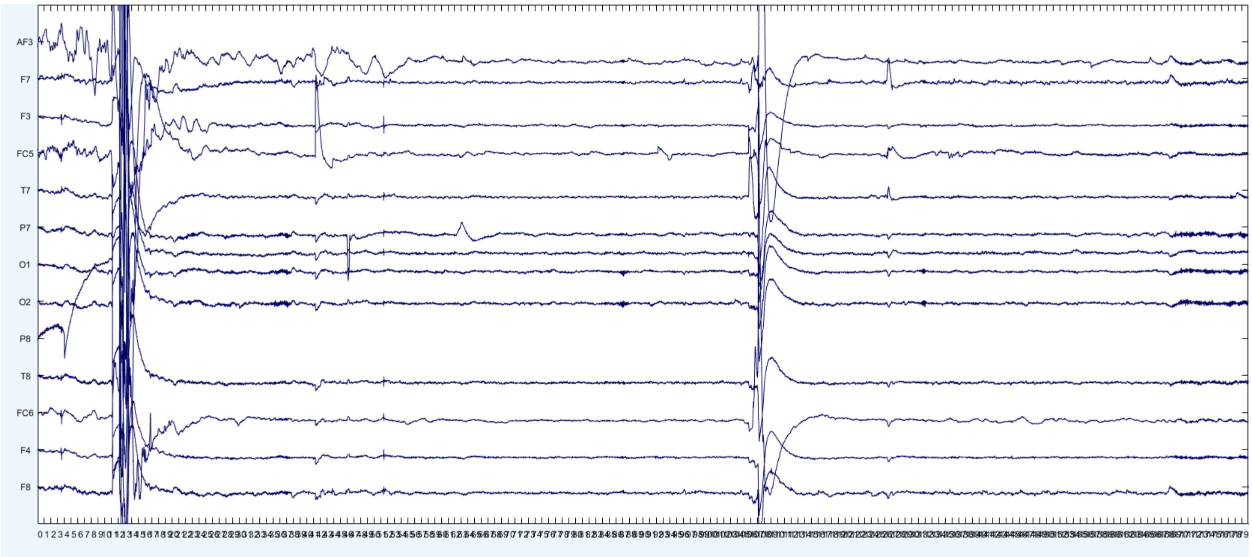


Figure 13 Raw data 180s sample

3.1.5.2 Filtering the Signal

We generated low and high pass filters using the EEGLAB tools. Since the brainwave signals are generally in between 0.5 Hz and 50 Hz we set the low end limit to 0.5 Hz and the high end limit to 50 Hz. The graphic of the filter can be seen in the figure by it's magnitude and its phase.

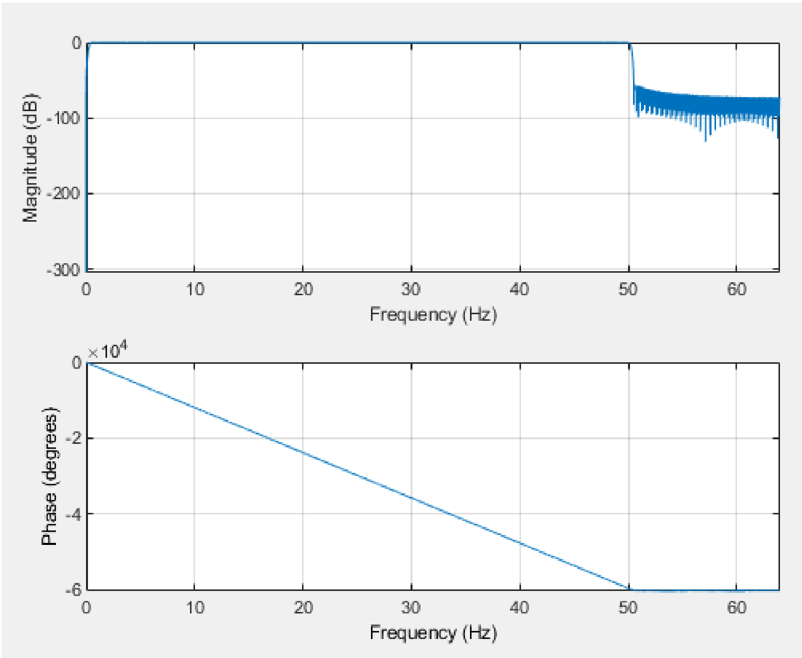


Figure 14 Filtered Data

3.1.5.3 ICA Decomposing

Figure shows the plot of the last operation of preprocessing called ICA**.**

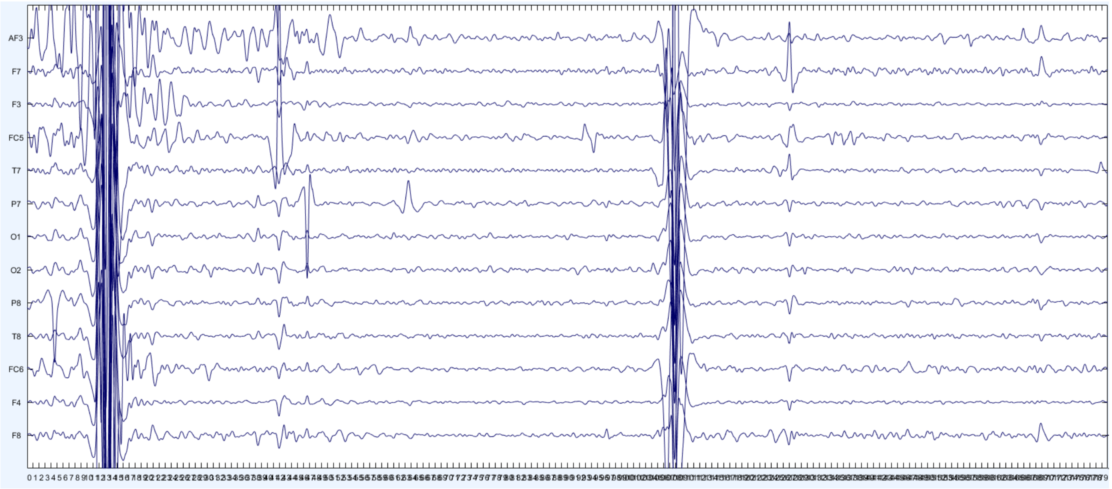


Figure 15 ICA decomposed data

To minimize the statistical dependency between certain components that are subsequently added to the signal, ICA offers a linear transformation. After applying ICA to EEG signals and comparing it with old EEG data, it has been shown that it significantly cleans the background noise and separates the signals as desired.

3.1.5.4 Examination by Eye

In the figure we can clearly see some disturbances in all of the channels at the same time. These events cause us to suspect possible eye blinking artifacts in the signals. To examine that we plot the topographic visuals for further analyzing.

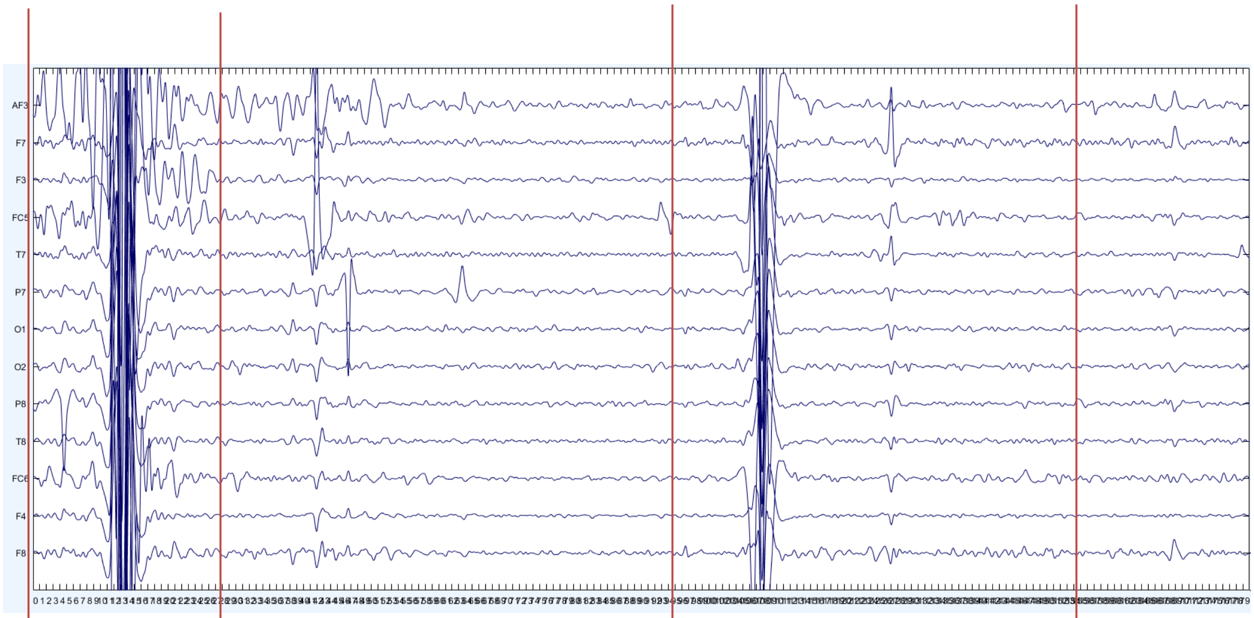
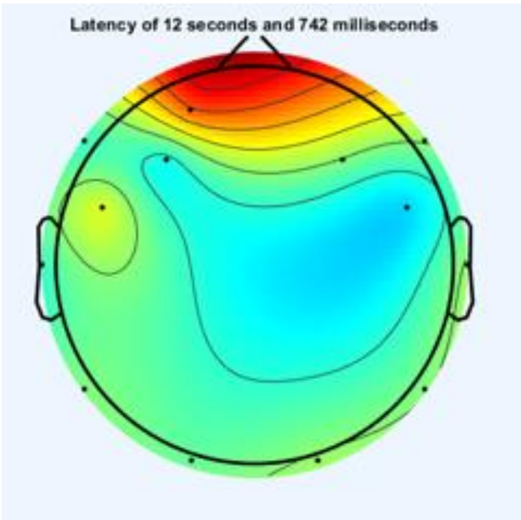
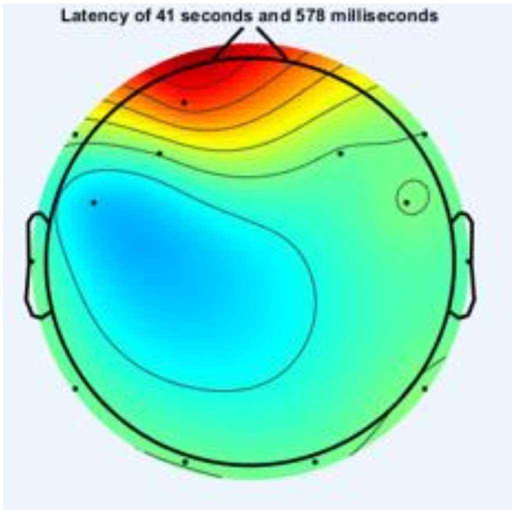


Figure 16 Eye blink expressed in seconds

In the topographic visuals we can see that the signals from the frontal part of the brain is much more active compared to the rest of the brain. This brings us to the conclusion that these events are related to the eye blinking of the subject. Therefore, we reject these signals.





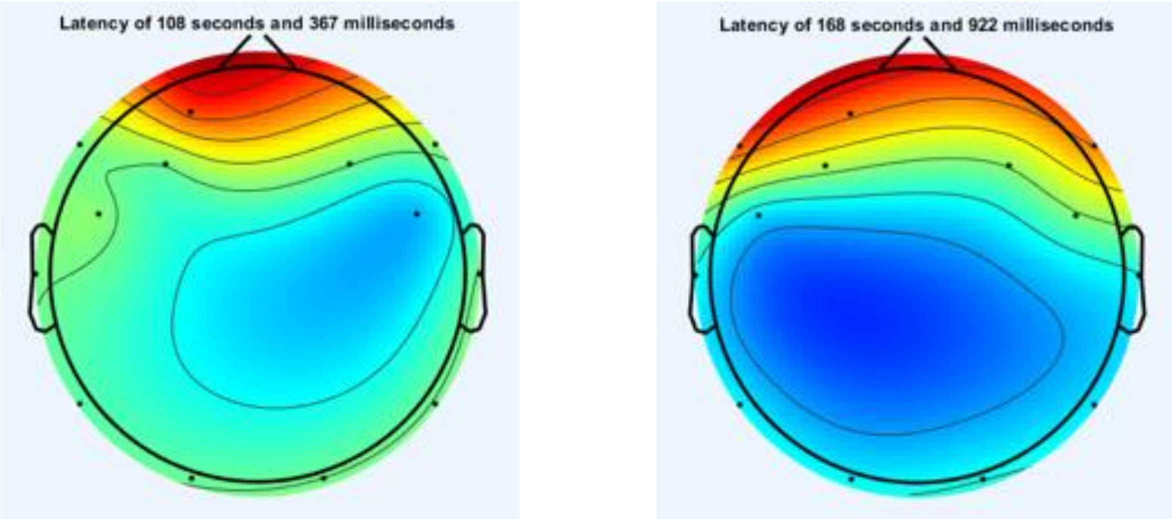


Figure 17 Eye blink shown in topography

3.2. Software Engineering

3.2.1. User Interfaces

The user will have the option to select a database in the file menu and then there will be four frames to adjust the data until it is detect process.

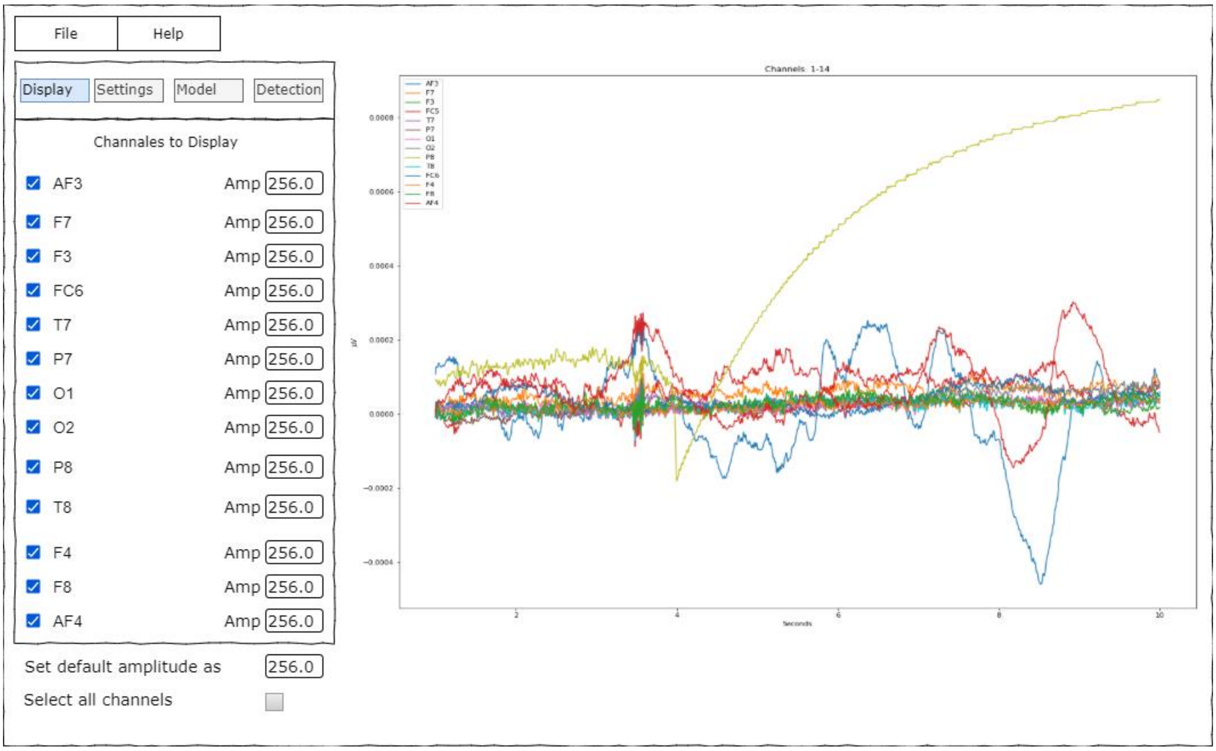


Figure 18 Screen Design of Display Frame

An example of the system's main interface in the figure. Users can display the related data's visual graph from the display frame.

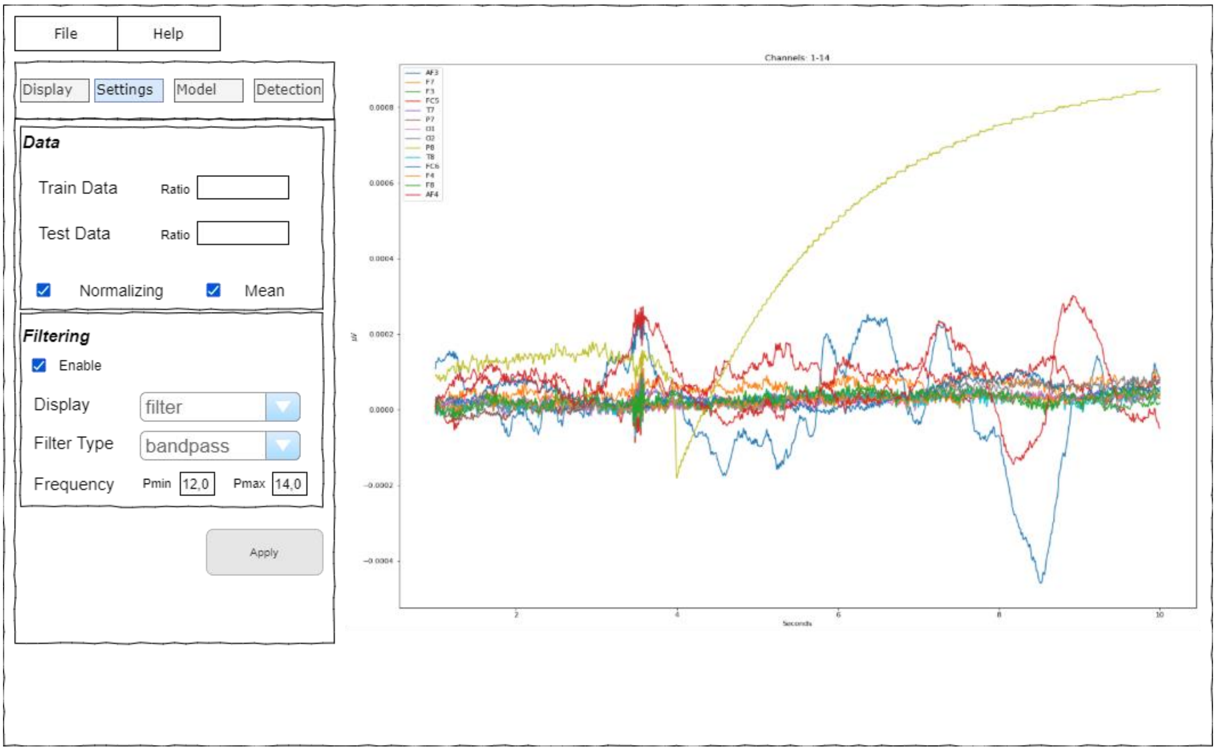


Figure 19 Screen Design of Settings Frame

Settings frame is to adjust the datas and filter methods with the apply button to save the changes.

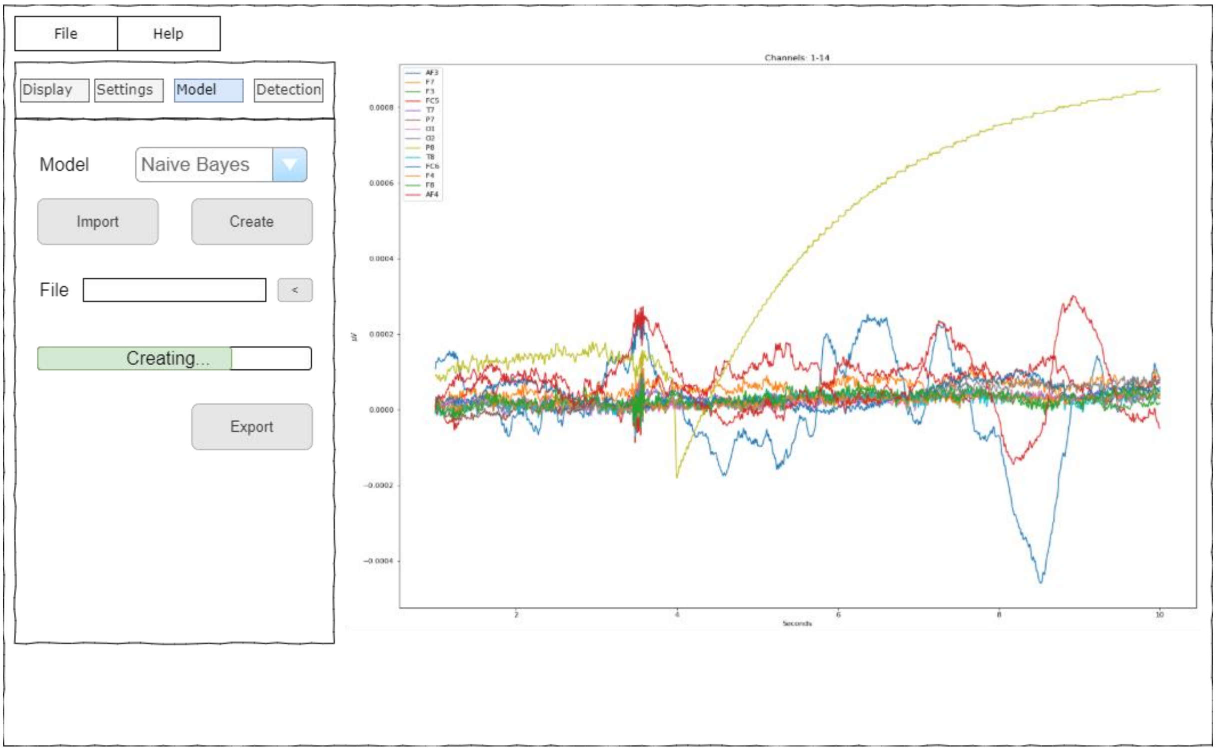


Figure 20 Screen Design of Model Frame

In the model frame, users can select the preferred model type to create a new model. Also there are buttons to import another model or export to save the model.

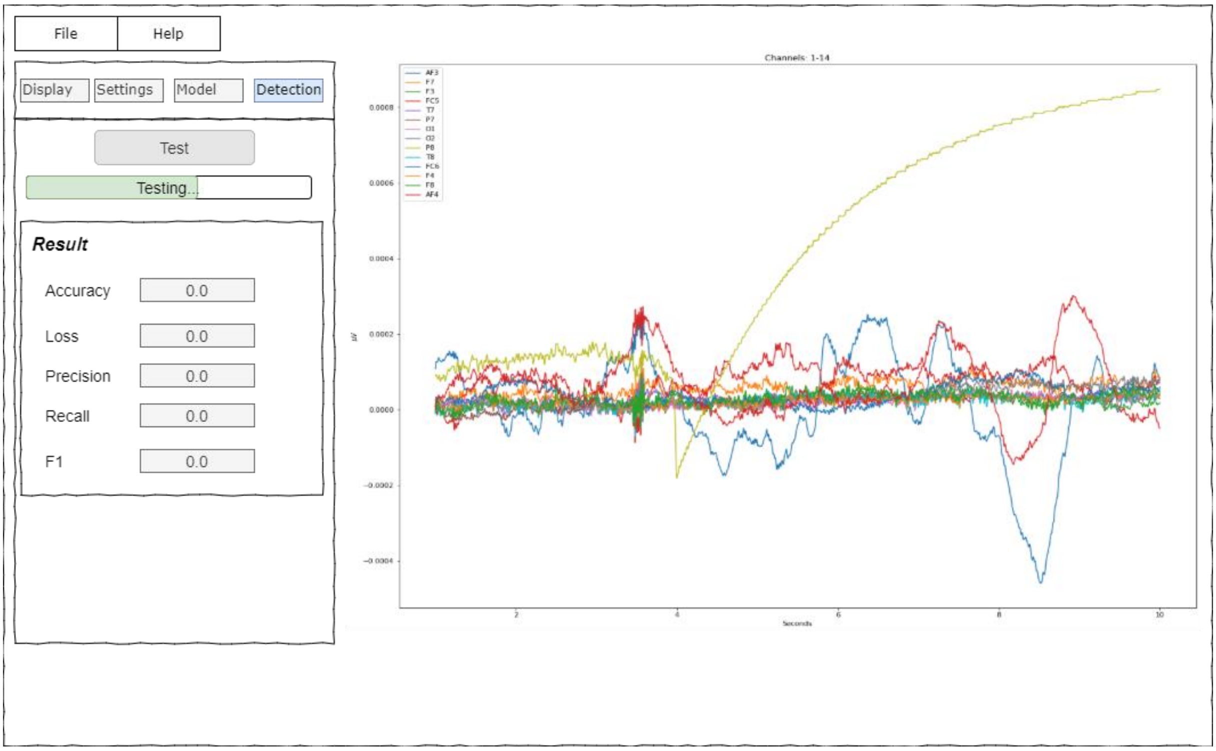


Figure 21 Screen Design of Detection Frame

After selecting a model, in the detection frame of the system, users can test it. The result will be displayed below the result container in the image.

3.2.2. Software Interfaces and Setup

The project is made with Python 3.8 version, because this language has a huge library and made by community of Python therefore the libraries are several and easy to find. Tkinter library is used to create interfaces and layouts for a better user experiences. To analyze the data in the prediction part, a 30-person dataset consisting of 4 sessions for each was used and exported as mdf files after extracting different features of part of the project. To understand the mdf files, “EDFbrowser” application is used. The Python libraries are installed by Anaconda 4.8.3. and Pip 20.2.4. using “conda install” or “pip install” commands. These libraries used were matplotlib, pandas, numpy, tkinter, sklearn to name a few. The example screens for the illustration process before creating the interface part was made by “Draw.io” application. The Project were produces on Windows 10 with 2004 version using 7th generation intel core i7 processor with 6GB RAM.

3.2.3 Actor Glossary

User: A person who has access to the system, knowledge about EEG and machine learning basics.

System: Program that runs when requested by the user.

3.2.4 Use-case Glossary

Use-case Name	Description	Participating Actors
Importing Data	User imports dataset from specific location	User, System
Adjusting Display Settings	User sets the preferred settings	User
Adjusting Settings of the Data	User can manage the data before creating a model.	User
Adjusting Model	User can import, export or create a model and system giving feedback while model is creating.	User, System
Sending a Test Data to the Model	After user sends a request to the system to start the process,system gives a detailed feedback.	User, System

Table 3 Use-case Glossary

3.2.5 Use-case Scenarios

Use-case Name	Importing Data
Use-case Description	User imports dataset from specific location
Actors	User
Pre-Condition	A dataset must be included in the computer
Post-Condition	The Dataset is selected in the system.
Normal Flow	Step1: User opens file menu Step2: User selects import dataset section Step3: File menu opens Step4: User chooses a dataset from computer Step5: System includes dataset
Alternate Flow	Step4: If there is no dataset, user needs to find a new one
Business Rules	Dataset must be valid for the system requirements

Table 4 Use-case Scenarios of Importing Data

Use-case Name	Adjusting Display Settings
Use-case Description	User sets the preferred settings
Actors	User
Pre-Condition	At least one channel must be selected
Post-Condition	System applies the specified settings
Normal Flow	Step1: User clicks the display frame Step2: User selects and adjusts needed settings Step3: User clicks to apply button
Alternate Flow	Step2: Some settings can not exceed their maximum value, user must give a proper value.
Business Rules	Some settings must be valid for the system requirements

Table 5 Use-case Scenarious of Adjusting Display Settings

Use-case Name	Adjusting Settings of the Data
Use-case Description	User can manage the data before creating a model
Actors	User

Pre-Condition	User has to give a proper ratio for the data splitting
Post-Condition	System will have test and train datas before modelling and some filter settings
Normal Flow	Step1: User clicks the settings frame Step2: User selects a ratio for the train and test sections Step3: User selects some options to increase the data quality Step4: If filtering will be used, user adjusted filtering options Step5: User clicks to apply button
Alternate Flow	Step2: If there is an error in setting the ratio values, the user must provide appropriate values. Step4: If there is an error in setting the frequency values, the user must provide appropriate values
Business Rules	Some settings must be valid for the system requirements

Table 6 Use-case Scenarious of Adjusting Settings of the Data

Use-case Name	Adjusting Model
Use-case Description	User can import, export or create a model and system giving a feedback while model is creating.
Actors	User, System
Pre-Condition	User needs to select a model type and a model file
Post-Condition	System will have a model to detection part
Normal Flow	Step1: User clicks the model frame Step2: User selects a model type rom combobox widget Step3: User clicks to create button Step4: System creates a model and gives a feedback when creating process is finished Step5: User clicks export button to extract the model to another file base
Alternate Flow	Step3: If another model wants to be used, user clicks the import button Step4: User clicks to search file button

	Step5: User accesses to the file required for the model Step6: User selects the model Step7: System imports the model
Business Rules	Imported model must be appropriate for the system requirements

Table 7 Use-case Scenarious of Adjusting Model

Use-case Name	Sending a Test Data to the Model
Use-case Description	After user sends a request to the system to start the process, system gives a detailed feedback.
Actors	User, System
Pre-Condition	There must be a valid model before starting the test
Post-Condition	System will give a detailed result of the prediction
Normal Flow	Step1: User clicks to test button Step2: System predicts the result Step3: System sends a feedback
Alternate Flow	Step1: If there is no a valid model than user go previous frames and creates a new one
Business Rules	There must be a valid model

Table 8 Use-case Scenarious of Sending a Test Data to the Model

3.2.6 Use-case Diagram

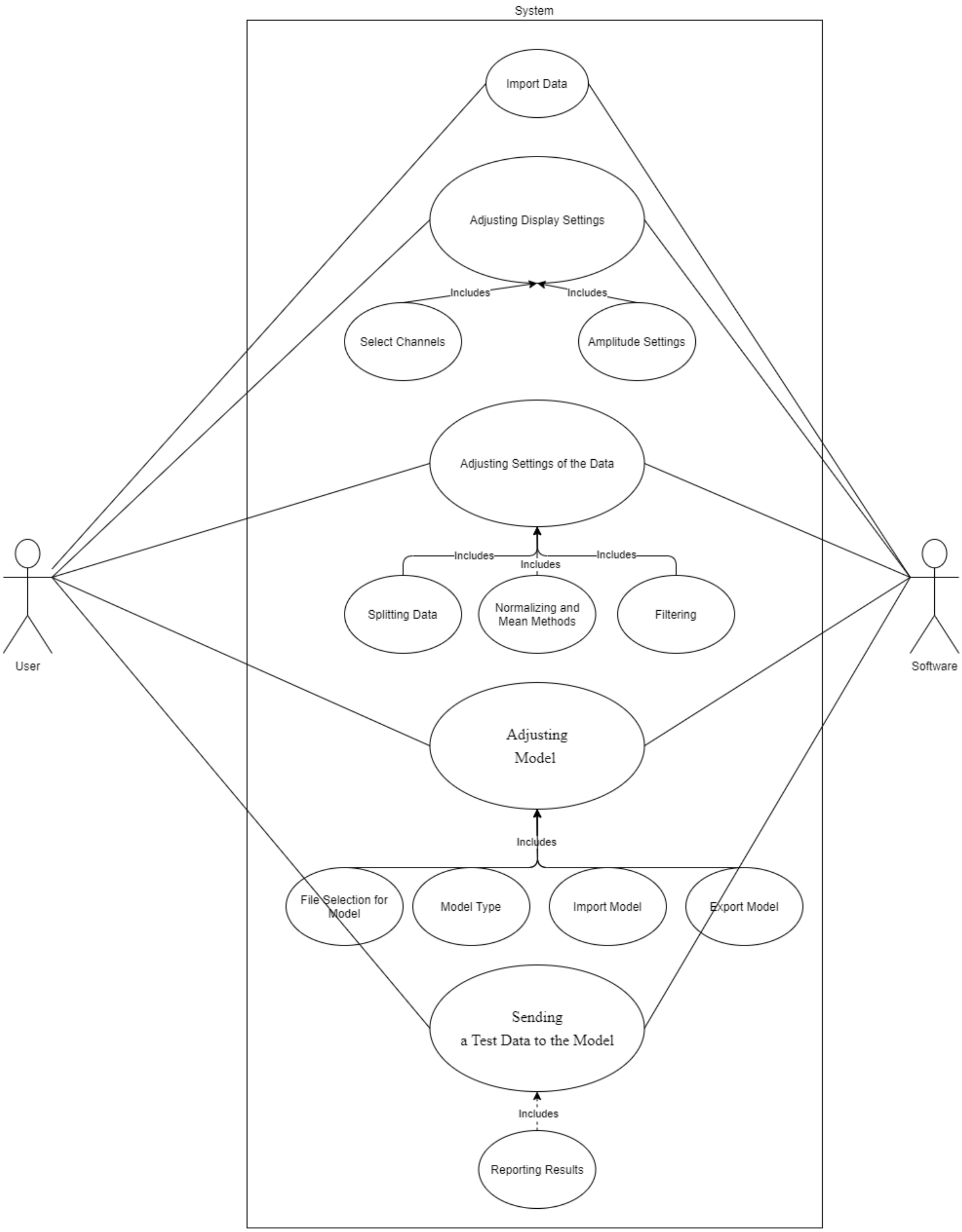


Figure 22 Use-case Diagram

3.2.7 Data Modeling

In this part, try to answer the following questions for your application

- What are the main tasks or functions that are performed by the actor?

The system has 5 functions for user actor to achieve a successful result. These are importing data, adjusting display settings, adjusting data settings, adjusting model settings and testing the model. System actor responds to some of the user actor's requirements for all the above functions.

- What system information will the actor acquire, produce or change?

The user can add a dataset to the system, make changes on the data attributes to be used in the dataset, add or remove this data, and create different models with model types.

- Will the actor have to inform the system about changes in the external environment?

If there is a later dataset change, it may need to be defined to the system again or new additions to some model or filter types may be required.

- What information does the actor desire from the system?

When the user adds data, it may be desirable to see the visual state of the data. Also, model creation can take a long time, and a loading bar symbolizing the formation may be requested by the user. Finally, after the data is predicted, a detailed return is requested.

- Does the actor wish to be informed about unexpected changes?

It is assumed that the user knows the basic information in the steps of machine learning. However, adding new unexpected data manipulation methods or adding new model types may affect user adaptation. The user must be informed about the changes made.

3.2.8 Activity Diagram

Activity Diagram for User and System

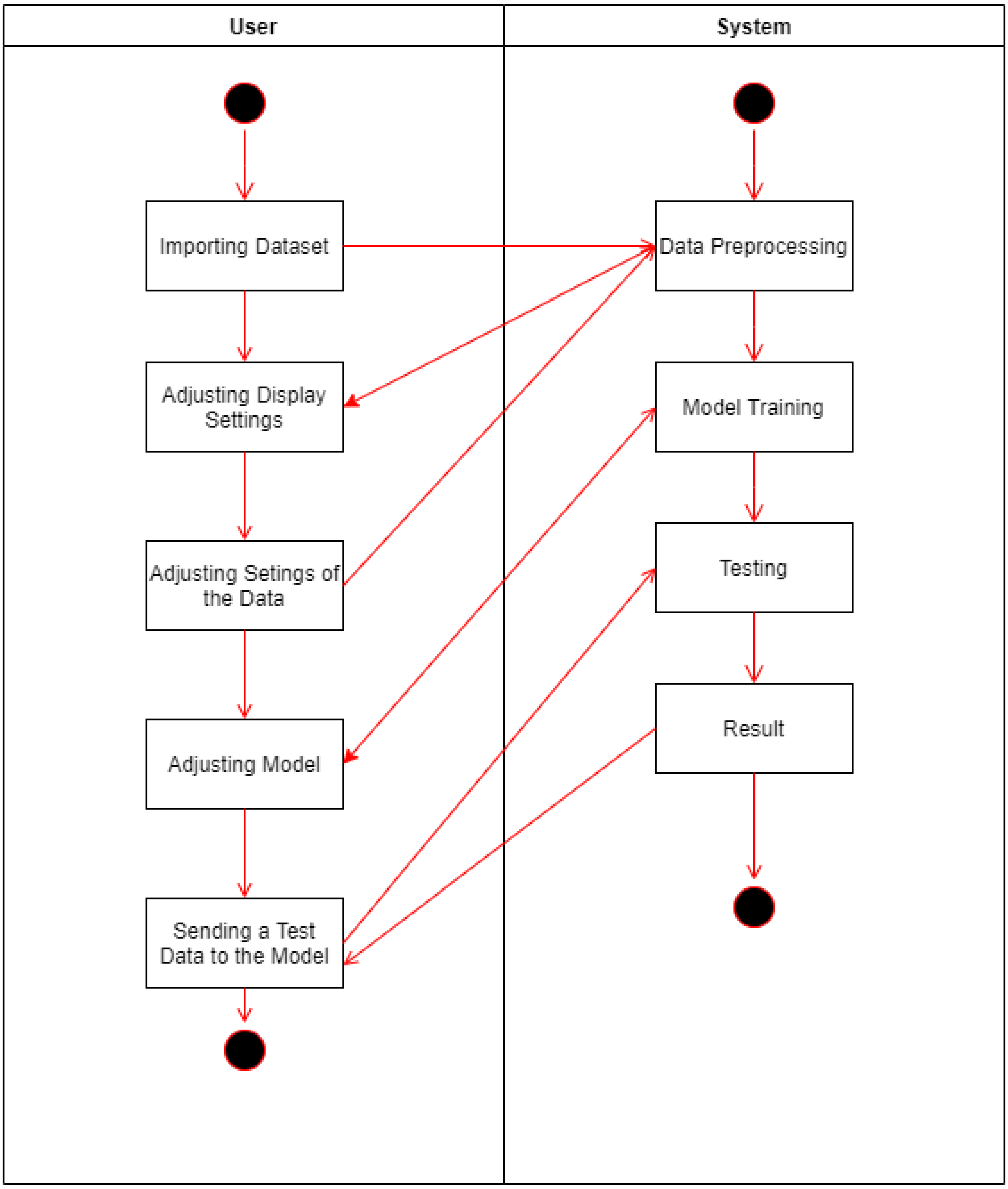


Figure 23 Activity Diagram for User and System

When the user starts the system, it defines a dataset by the user and then sets some changes to be used on the data in the display frame. The system then graphically displays an image of the data to the user. Then the user divides the data into two as test and train in the frame called settings and the system saves this information in itself. The user then chooses a model method in the Model frame section and creates the model, and the system saves the model and makes a return to the user about the expected time. Finally, the user has come to the stage of testing the model. In the detection section, the system sends a request to test it, and the system provides a detailed feedback to the user after testing the model.

3.2.9 Data Analysis Process Diagram

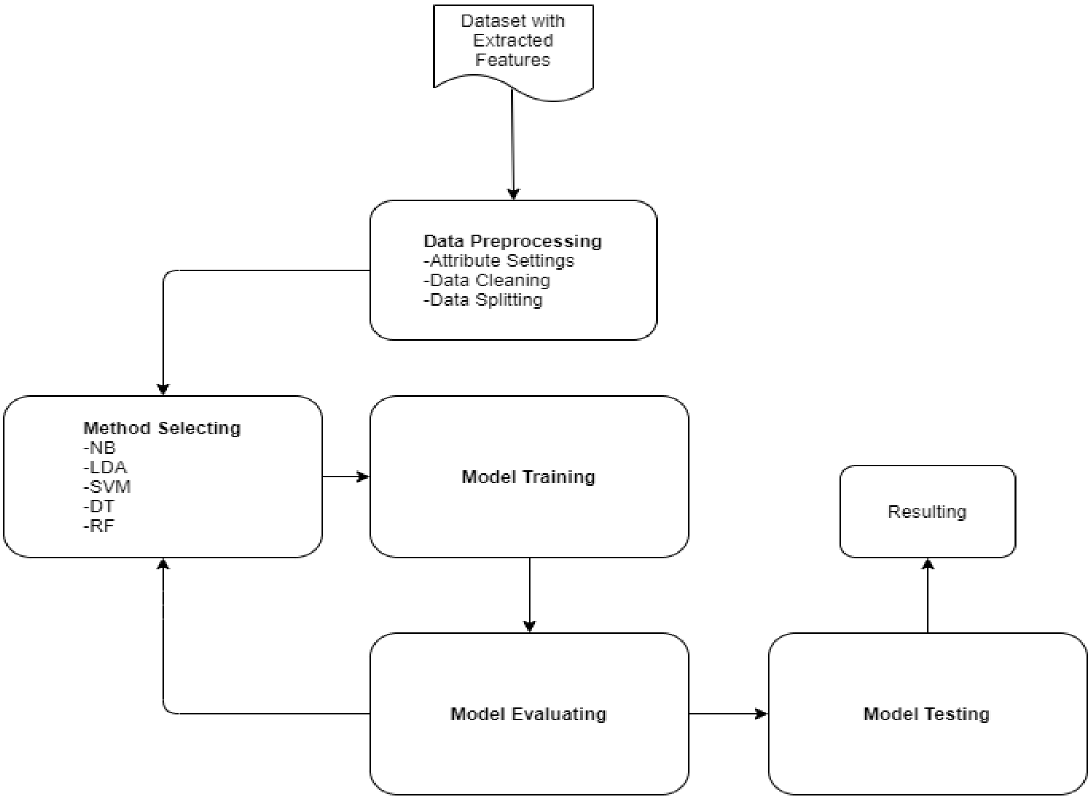


Figure 24 Data Analysis Process Diagram

Datas are taken as extracted with features. Then the data is taken to a preliminary process and cleaning and separation methods are applied here. Then, when the data is ready for use, it is decided which machine learning method to apply. After this process, the model is trained with the training data, tested with the test data and the results of the model are evaluated. If a high rate of success is achieved in the results, it means that the whole process was successful, if the success rate is not sufficient, the beginning is returned to make a change on the data.

4. CONCLUSION

This article explains what problems are encountered in imaging brain signals provided by EEG and how to solve these problems. How preprocessing is carried out is supported with step-by-step figures. The process we foresee for our future work will proceed as follows: Feature extraction, feature selection, and classification. Later, m-learning will be done using Python and MATLAB algorithms and it will be very easy to clean the data from artifacts.

5. REFERENCES

1. D. Vikrant, P. Matin, “A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals”, *J Big Data* 7, 18 (2020). https://doi.org/10.1186/s40537-020-00289-7
2. Albasri, Ahmed (2019), “EEG dataset of Fusion relaxation and concentration moods”, Mendeley Data, V1, doi: 10.17632/8c26dn6c7w.1. http://dx.doi.org/10.17632/8c26dn6c7w.1
3. MNEtools, URL: https://mne.tools/dev/auto_tutorials/intro/plot_10_overview.html
4. HackathonTutorial, URL: https://bigdatawg.nist.gov/HackathonTutorial.html
5. Youtube, URL: https://www.youtube.com/watch?v=sttf-Rgfl1Q
6. Z. H. Kilimci, A. Güven, M. Uysal, S. Akyokus, "Mood Detection from Physical and Neurophysical Data Using Deep Learning Models", Complexity, vol. 2019, Article ID 6434578, 15 pages, 2019. https://doi.org/10.1155/2019/6434578
7. Scott Makeig SCCN Director, Preprocessing Tools, 17 January 2020
8. Chuang C.-H., Ko L.-W., Lin Y.-P., Jung T.-P., Lin C.-T. (2014b). Independent component ensemble of EEG for brain-computer interface. IEEE Trans Neural Syst. Rehabil. Eng. pp. 22, 230–238. 10.1109/TNSRE.2013.2293139
9. Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. NeuroImage, 86, 446–460. Document id:10.1016/j.neuroimage.2013.10.027
10. International Journal of Computer Theory and Engineering, Vol. 7, No. 2, April 2015
11. Benjamin U. Cowley1,2, Jussi Korpela1, Jari Torniainen3, Computational testing for automated preprocessing: a Matlab toolbox to enable large scale electroencephalography data processing, March 6, 2017
12. Anderer, P., Roberts, S., Schlögl, A., Gruber, G., Klösch, G., Herrmann, W., … Saletu, B. (1999). Artifact Processing in Computerized Analysis of Sleep EEG – A Review. Neuropsychobiology, 40(3), 150–157. doi:10.1159/000026613
13. R. Ramirez, Z. Vamvakousis, “Detecting emotion of EEG signals using the emotive epoc device”, Proceedings of the 2012 international conference on Brain Informatics, (2012).
14. M.L.R. Menezes, A. Samara, L.Galway, A. Santanna, “Towards emotion recognition for virtual environments an evaluation of EEG features on brenchmark dataset”, Personal and Ubiguitous Computing 21, 6, (2017)