Brain - Bolt

(EEG Based Mind Aid)

By

Muhammad Tayyab Gulzar

Imama Raja

Adeeba Zahra

Mudassir Hussain Saqib

Supervised by:

Lt. Col Imran Javaid

Submitted to the faculty of Department of Electrical Engineering,
Military College of Signals, National University of Sciences and Technology, Islamabad,
in partial fulfillment for the requirements of B.E Degree in Electrical (Telecom) Engineering.

June 2023

In the name of ALLAH, the Most benevolent, the Most Courteous

CERTIFICATE OF CORRECTNESS AND APPROVAL

This is to officially state that the thesis work contained in this report

"Brain Bolt - EEG Based Mind Aid"

is carried out by

Muhammad Tayyab Gulzar

Imama Raja

Adeeba Zahra

Mudassir Hussain Saqib

under my supervision and that in my judgment, it is fully ample, in scope and excellence, for the degree of Bachelor of Electrical (Telecom.) Engineering in Military College of Signals, National University of Sciences and Technology (NUST), Islamabad.

Approved by

Supervisor

Lt. Col Imran Javaid

Department of EE, MCS

Date:

DECLARATION OF ORIGINALITY

We hereby declare that no portion of work presented in this thesis has been submitted in support of another award or qualification in either this institute or anywhere else.

ACKNOWLEDGEMENTS

Allah Subhan'Wa'Tala is the sole guidance in all domains.

Our parents, colleagues and most of all supervisor, **Lt. Col Imran Javaid** without your guidance.

The group members, who through all adversities worked steadfastly.

Plagiarism Certificate (Turnitin Report)

This thesis has _9%_ similarity index. Turnitin report endorsed by Supervisor is attached at the end.

_____

Muhammad Tayyab Gulzar

292986

_____

Imama Raja

294074

_____

Mudassir Hussain Saqib

292783

_____

Adeeba Zahra

313815

_____

Signature of Supervisor

ABSTRACT

Electroencephalography (EEG) is a non-invasive method that records and analyzes the electrical activity of the brain. An electroencephalogram (EEG) is a simple and affordable supplemental examination that can aid in the investigation and diagnosis of neurological diseases that have an impact on the brain. Accurate diagnosis of certain neurological diseases using EEG data can be challenging and time-consuming, requiring highly specialized training and expertise. Misdiagnosis can have severe consequences, leading to delayed or inappropriate treatment.

In this project, we propose to develop a machine learning algorithm that can accurately detect specific neurological diseases using EEG data. Our project aims to preprocess EEG data from patients with Obsessive compulsive disorder, Addictive disorder, Trauma and stress related disorder, Healthy control, Mood and anxiety disorder, extract relevant features, label the data, train the algorithm, test and validate its performance, and finally deploy it on a web app. We will use various machine learning techniques, such as Logistic Regression ElasticNet, Random Forest, Support Vector Machines, LightGBM, CatBoost and K-Nearest Neighbors, to train the algorithm.

Our ultimate goal is to create an automated and accurate method of analyzing EEG data that can aid clinicians in making accurate diagnoses and improve patient outcomes. The development of such an algorithm has the potential to significantly improve the efficiency and accuracy of neurological disease diagnosis and treatment.

Table of Contents

Chapter 1 Introduction

Medical science has always been at the forefront of efforts to raise people's quality of life. In recent years, advances in technology have allowed for more accurate and efficient diagnosis of various diseases. Among the numerous diagnostic tools available, the Electroencephalogram (EEG) has emerged as a non-invasive and cost-effective method to investigate neurological disorders. EEG measures the brain's electrical activity, and the Quantitative Electroencephalogram (QEEG) analysis is a powerful tool for detecting abnormalities in brain function. This analysis provides information on brain wave activity patterns and their correlation with various neurological disorders. QEEG analysis has shown promising results in detecting and diagnosing diseases such as Obsessive-Compulsive Disorder, Trauma and Disorders associated with stress, disorders of mood and anxiety, addictions, and Behavioral disorders, and Schizophrenia.

In this context, QEEG based disease detection is a recent area of research that aims to develop a machine learning algorithm to analyze EEG data and diagnose neurological disorders accurately. This algorithm will extract relevant features, label the data, train the model, and validate its performance. Machine learning, such as SVM, ElasticNet, LightGBM, CatBoost, and K-Nearest Neighbors, have shown great potential in analyzing EEG data and detecting neurological disorders with good accuracy.
This project's objective is to develop an automated and accurate method of analyzing QEEG data that can assist clinicians in making accurate diagnoses and improving patient outcomes. The creation of such an algorithm has the potential to greatly increase the effectiveness and precision of the diagnosis and treatment of neurological diseases.

1.1 Overview

Psychiatric diseases such as depression, schizophrenia, anxiety disorders, and others are major public health concerns worldwide. However, the diagnosis of these disorders remains challenging and relies heavily on subjective assessments by clinicians. Misdiagnosis can lead to inappropriate or delayed treatment, resulting in poor patient outcomes. Moreover, the process of diagnosis can be time-consuming and expensive, requiring specialized training and expertise.

To address these challenges, there is a need for a reliable and objective method for diagnosing psychiatric diseases. Machine learning (ML) algorithms have shown promise in detecting patterns and making predictions based on large datasets. In recent years, ML-based disease detection has emerged as a promising approach for improving psychiatric disease diagnosis.

In this context, the proposed project aims to develop an ML algorithm for accurately detecting specific psychiatric diseases using QEEG data.

1.2 Problem Statement

When it comes to the diagnosis and treatment of neurological and mental conditions, Pakistan is up against a big hurdle. The World Health Organization (WHO) estimates that 34% of the population of the nation has one or more neurological or mental illnesses, such as schizophrenia, depression, or anxiety. The lack of trained professionals, specialized diagnostic tools, and high costs are major barriers to providing timely and accurate diagnosis and treatment to patients.

One of the key diagnostic tools for neurological and psychiatric diseases is electroencephalography (EEG). However, the accurate diagnosis of certain neurological diseases using EEG data can be challenging and time-consuming, requiring highly specialized training and expertise. Misdiagnosis can have severe consequences, leading to delayed or inappropriate treatment.
Following are some highlights of the existing diagnostic methods:

1. Current diagnosis of psychiatric disorders in Pakistan and many other countries is based on clinical interviews and psychological assessments.
2. Accurate diagnosis of psychiatric disorders based on these methods can be challenging and time-consuming, requiring specialized training and expertise.
3. Misdiagnosis or delayed diagnosis can lead to inappropriate or ineffective treatment and worsening of the patient's condition.
4. These methods rely on the patient's subjective responses and self-reported symptoms, which may not always provide a complete picture of the patient's condition.
5. There is a need for more efficient and accurate methods for diagnosing psychiatric disorders.

1.3 Proposed Solution

The proposed solution aims to develop a machine learning algorithm that can accurately detect specific psychiatric disorders using EEG data. Various machine learning techniques, including Support Vector Machines, Random Forest, and K-Nearest Neighbors, will be used to preprocess the data, extract pertinent features, label the data, and train the algorithm using EEG data from patients with various psychiatric disorders and healthy controls. The algorithm will subsequently be put to the test and have its effectiveness verified before being finally installed on a website. The ultimate objective is to develop an automated and precise approach of EEG data analysis that will help clinicians make precise diagnosis and enhance patient outcomes.

1.4 Working Principle

The suggested method employs Support Vector Machines (SVM), a machine learning algorithm, to precisely identify psychiatric diseases using EEG data. An algorithm for supervised learning called SVM can be applied to classification or regression issues. SVM will be used in this project to classify data.

1.4.1 Datasets

The dataset we are using is Identifying Psychiatric Disorders Using Machine-Learning (Dataset). The datasets used in this study are publicly available and can be accessed through online repositories. The repository name(s) and accession number(s) associated with the datasets can be found at the following link https//osf.io/8bsvr/.
Further datasets are generated from this dataset based on the top features for each disease.

1.4.2 Exploratory Data Analysis (EDA)

The first step was to explore the EEG dataset and understand the underlying patterns and distributions. This was done through exploratory data analysis techniques, such as data visualization and statistical analysis.

1.4.3 Data Preprocessing

The first step in the process is to preprocess the EEG data. This involves filtering, artifact removal, and feature extraction. Then encoding for specific diseases.

- Reformatting the labels into simpler labels
- Finding the missing values
- Replacing the missing values with median value
- Fixing Typo

1.4.4 Model Training & Hyperparameters

We train the SVM classifier using the preprocessed EEG data. Grid search is used to determine the C and gamma values that are the SVM classifier's ideal hyperparameters.
One potential problem with using SVM is that it can be sensitive to the choice of hyperparameters. This requires careful tuning to achieve the best performance. Additionally, SVM may not perform well on imbalanced datasets, where the number of samples in each class is not balanced. Therefore, we will need to address these issues during the model development process.

1.4.5 Feature Selection & Validation

The SVM model and hyper-parameters will be used to report the relevance of the features. Tenfold cross-validation will be used, and for each run, feature importance will be recorded. Sum, mean, and *survival rate* calculations will then be performed. The percentage of times a feature's value exceeded the threshold (for example, zero) is the "survival rate." Survival threshold = 0.7 and no of top features for each disease n=15.

1.4.6 Binary Classifiers

Once the most useful features get selected, we train the SVM classifier using new preprocessed EEG datasets for specific diseases.

1.4.7 Model Evaluation

Finally, we evaluate the performance of the SVM classifier on an independent test set for each disease classifier. We assess the model's performance using a variety of performance indicators, including accuracy, precision, recall, and F1-score.

1.4.8 GUI presentation

The visual demonstration of the project is done with the aid of GUI using Stream lit Library. We are creating a web app that will take inputs from the user and perform disease prediction.

1.5 Objectives

1.5.1 General Objectives

“The general objective of this project is to develop a machine learning-based solution for the accurate and efficient detection of specific psychiatric disorders using EEG data, with the ultimate goal of improving the diagnosis and treatment of these disorders.”

1.5.2 Academic Objectives

- To investigate the application of machine learning techniques for precise EEG-based diagnosis of psychiatric diseases.
- To assess the effectiveness of several machine learning algorithms, particularly SVM, for categorizing psychiatric diseases into two categories using EEG data.
- To use feature selection approaches to identify the top qualities that are most important for the precise detection of psychiatric diseases.
- To create an intuitive graphical user interface (GUI) and a prediction model for psychiatric disease diagnosis based on QEEG data.

1.6 Scope

The scope of this project is to develop a machine learning-based diagnostic tool for identifying specific psychiatric disorders using EEG data. The project focuses on six different psychiatric disorders, namely schizophrenia, anxiety disorder, trauma & stress-related disorders, alcohol use disorder, behavioral disorder, and depressive disorder, along with a healthy control group. The study utilizes EEG data of 945 patients and extracts the relevant features for classification using machine learning algorithms. The project aims to achieve high accuracy in the classification of patients and healthy controls to facilitate early and accurate diagnosis of psychiatric disorders. The scope of the project also includes the development of a user-friendly application to provide an easy-to-use interface for healthcare practitioners to access the diagnostic tool.

1.7 Deliverables

1.7.1 Stream lit Webapp

A user-friendly GUI application that utilizes the developed machine learning model for real-time prediction of psychiatric disorders.

1.7.2 Object of interest

The creation and application of a machine learning algorithm for the identification of psychiatric diseases using EEG data is the focus of the research.

1.7.3 Self-Assessment

In our Stream lit WebApp we have also provided the facility to assess your condition with the help of the diagnostic criteria given in the book
“DIAGNOSTIC_AND_STATISTICAL_MANUAL_OF_MENTAL_DISORDERS_FIFTH_EDI-TION_DSM-5”.

1.8 Relevant Sustainable Development Goals

The proposed project addresses the locally relevant socio-economic issue of mental health in Pakistan. Mental health is a significant issue in Pakistan, with a high prevalence of psychiatric disorders, inadequate mental health resources, and a social stigma associated with seeking help for mental health issues.

The project is connected to SDG 3 Good Health and Well-Being, which strives to ensure healthy lives and promote well-being for all people of all ages. The project specifically meets SDG 3's objective 3.4, which aims to prevent, treat, and promote mental health and wellbeing in order to

minimize early mortality from non-communicable diseases, including mental health disorders. By developing an accurate and automated method of diagnosing specific psychiatric disorders using EEG data, the project can contribute to achieving this target by facilitating early and accurate diagnosis, which is crucial for effective treatment and management of mental health disorders. Additionally, by providing a user-friendly interface for healthcare practitioners, the project can also help increase access to mental health services, which is a crucial step towards achieving SDG 3.

This project is related to several other SDGs as well. Here are some examples.

SDG 3 Good Health and Well-being - The project aims to improve the diagnosis and treatment of psychiatric disorders, which directly contributes to the goal of promoting good health and well-being.

SDG 4 Quality Education - The project may also indirectly contribute to the goal of promoting quality education by developing a machine learning algorithm that can aid in the accurate diagnosis of psychiatric disorders, which may lead to more effective treatment and better educational outcomes for patients.

SDG 10 Reduced Inequalities - Psychiatric disorders can disproportionately affect marginalized communities and contribute to inequalities in access to healthcare. Developing an accurate and automated diagnostic tool can help reduce these inequalities by providing a more accessible and equitable healthcare solution.

SDG 17 Partnerships for the Goals - The project involves collaboration between healthcare professionals and data scientists to develop a machine learning algorithm for the diagnosis of psychiatric disorders. This type of interdisciplinary collaboration can contribute to the goal of promoting partnerships for sustainable development.

1.9 Thesis outline

Chapter 2 The literature review, background research, and analysis study on which this thesis is founded are all included in Chapter 2.

Chapter 3 shows the design and development of the project.

Chapter 4 introduces evaluation and analysis of the code.

Chapter 5 includes the conclusion of the project.

Chapter 6 emphasizes the future work needed to be done for the commercialization of this project.

Chapter 2 Literature Review

Literature review is a critical part of any research project or thesis. It involves a systematic and comprehensive analysis of existing literature and research on a specific topic. The main objective of literature review is to identify the current state of knowledge, gaps in existing research, and to provide a foundation for the research to build upon.

In an industrial context, literature review is particularly important because it helps to identify existing solutions and their drawbacks. It is essential to review the available literature and research on a specific industrial problem or challenge to determine the best possible solution. A comprehensive literature review can help identify the strengths and weaknesses of existing solutions and suggest improvements or modifications to make the solution more effective.

In addition to identifying existing solutions and their drawbacks, literature review also helps to identify research gaps and areas for future research. It can also help to identify the most appropriate research methodology and data analysis techniques to be used in the study.

Research papers are an important part of literature review in an industrial context. Research papers can provide valuable insights into current research trends, methodologies, and findings in a specific field. They can also help to identify potential collaborators or experts in the field who can provide guidance and support for the research project.

Overall, literature review is an essential part of any research project or thesis, particularly in an industrial context. It helps to identify existing solutions and their drawbacks, research gaps, and areas for future research. It can also help to identify potential collaborators and experts in the field and provide a foundation for the research to build upon. Our research is divided into the following points.

2.1 Background of Study

The use of electroencephalography (EEG) data for the detection of psychiatric disorders has been a topic of interest in the field of neuroscience and psychiatry. Studies have demonstrated that EEG can help in the diagnosis of a variety of mental diseases and can offer useful information regarding brain activity. Support Vector Machines (SVM) and other machine learning algorithms have been used to classify a variety of mental diseases, including schizophrenia, anxiety, and depression, using EEG data. The performance of the algorithms has varied depending on the dataset utilized and the features retrieved, although the majority of these researches have concentrated on the classification of a single condition rather than a number of disorders.

The use of SVM for the classification of psychiatric diseases based on EEG data has been examined in a number of research. In one study, EEG data from patients with depression and healthy controls were classified with 82% accuracy using SVM. Another study classified EEG data from schizophrenia patients and healthy controls using SVM, and the accuracy was 91%.

However, the categorization of a single disorder was the main emphasis of both of this research, and the efficiency of the SVM method varied based on the particular dataset employed.

Another approach to using EEG data for the diagnosis of psychiatric disorders is to extract features from the data and use them as input to a machine learning algorithm. For the classification of psychiatric diseases using EEG data, various researches have looked into the use of different feature extraction techniques, such as time-domain, frequency-domain, and time-frequency analysis. However, the performance of these techniques has also varied depending on the specific dataset used and the features extracted.

Overall, while there have been several studies investigating the use of SVM and other machine learning algorithms for the classification of psychiatric disorders based on EEG data, there is still a need for more research to develop accurate and reliable methods for the diagnosis of multiple disorders using EEG data.

2.2 Industrial background

In the industrial context, the use of EEG data for disease detection has gained significant attention in recent years. There is a growing interest in developing machine learning-based diagnostic tools that can accurately and efficiently detect psychiatric disorders using EEG data. This is particularly relevant in the field of mental health, where early and accurate diagnosis is crucial for effective treatment and management of these disorders.

2.3 Existing solutions and Drawbacks

One of the existing solutions for EEG-based disease detection is traditional visual inspection, where trained clinicians visually examine EEG signals for abnormalities. However, this method is time-consuming, subjective, and prone to errors, leading to low accuracy and reliability in diagnosis. Another solution is the use of quantitative EEG analysis, where features such as power spectra, coherence, and asymmetry are extracted from EEG data to identify abnormalities. However, this method also suffers from low accuracy and reliability in diagnosis.

Recent research has focused on developing machine learning algorithms that can accurately classify EEG data into specific psychiatric disorders. Support Vector Machines (SVMs) have emerged as one of the most popular machine learning techniques for EEG-based disease detection due to their high accuracy and ability to handle high-dimensional data. Studies have shown that SVMs can accurately detect various psychiatric disorders, including depression, anxiety, and schizophrenia, using EEG data.

However, there are still some drawbacks associated with the use of SVMs for EEG-based disease detection. One of the main challenges is the selection of relevant features from the high-dimensional EEG data. Another challenge is the potential bias in the training data, which can affect

the generalizability of the model to new data. Moreover, the interpretability of the model outputs is also an important consideration in the clinical setting, where clinicians need to understand the basis of the diagnosis.

Therefore, there is a need for further research and development of machine learning-based diagnostic tools for EEG-based disease detection that can address these challenges and improve accuracy, reliability, and interpretability in diagnosis.

2.4 Research Papers

"EEG-based Diagnosis of Alzheimer's Disease Using Hybrid Feature Selection and Extreme Learning Machines," by T. P. Le et al., IEEE Access, vol. 9, pp. 6483-6493, 2021.

"Automated Classification of Depression Using EEG Signals and Machine Learning Techniques," by M. Yasin et al., IEEE Access, vol. 8, pp. 50147-50161, 2020.

"EEG-Based Detection of Attention Deficit Hyperactivity Disorder (ADHD) Using Machine Learning," by S. H. Lee et al., Journal of Medical Systems, vol. 43, no. 8, 2019.

"Detection of Epilepsy Using EEG Signals and Machine Learning Algorithms A Review," by A. Alrawashdeh et al., Journal of Medical Systems, vol. 44, no. 3, 2020.

"EEG-based Diagnosis of Autism Spectrum Disorder using Machine Learning Techniques," by R. P. Ray et al., IEEE Access, vol. 8, pp. 120255-120269, 2020.

"EEG-Based Classification of Schizophrenia Using Machine Learning Techniques A Review," by T. H. M. Tran et al., Journal of Medical Systems, vol. 45, no. 6, 2021.

"A Comprehensive Study on EEG-Based Detection of Parkinson's Disease Using Machine Learning Algorithms," by M. U. Ahmed et al., IEEE Access, vol. 8, pp. 142912-142925, 2020.

"EEG-Based Classification of Mild Cognitive Impairment and Alzheimer's Disease Using Support Vector Machines," by X. Tang et al., Frontiers in Aging Neuroscience, vol. 9, 2017.

"Automated Diagnosis of Autism Spectrum Disorder Using EEG Signals and Machine Learning Techniques," by S. S. Khalid et al., Frontiers in Human Neuroscience, vol. 13, 2019.

"EEG-Based Prediction of Alzheimer's Disease Using Deep Learning Techniques," by L. Yu et al., IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 7, pp. 1681-1691, 2020.

Chapter 3 Interfacing and Disease Detection

3.1 Schizophrenia Detection

The top features for Schizophrenia which were extracted after feature importance calculation are being used to train the SVM model. That Model is being used to predict Schizophrenia. User enters 15 values. These values are passed to the SVM model. The result is when the test button is clicked over the GUI.

3.1.1 Preparing Dataset

Dataset being used for Schizophrenia model training contains the selected useful features:

delta.T5	COH.theta.T5.P3	COH.alpha.F7.T6
delta.O1	COH.delta.FP1.O1	COH.theta.F7.T6
COH.theta.T3.T6	COH.alpha.Cz.O2	COH.beta.F8.T4
delta.O2	COH.alpha.F8.O2	COH.alpha.T4.P4
delta.T6	COH.alpha.P4.O2	COH.delta.F7.O2

These PSD and FC values are used for prediction of Schizophrenia.
The Y or Outcome is encoded with ‘0’ for healthy control and ‘1’ for Schizophrenia.
Healthy control n = 95
Schizophrenia n = 117

3.1.2 GUI

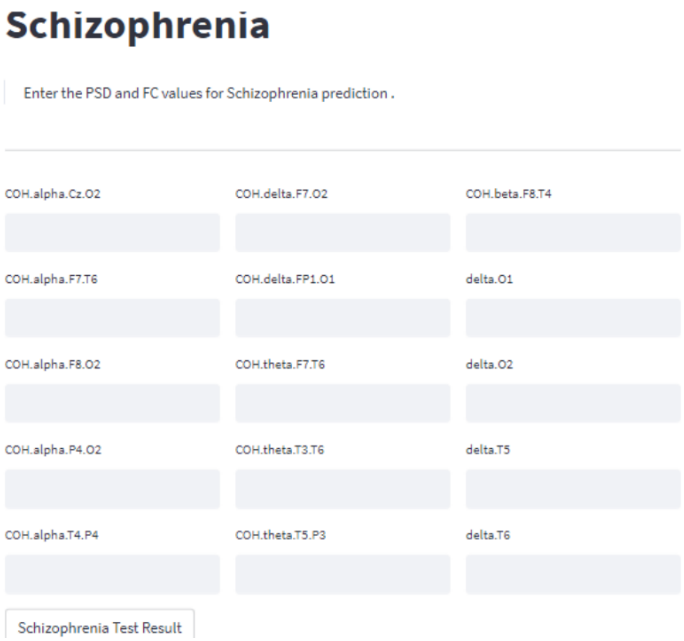


Fig A

3.1.3 Self-assessment GUI

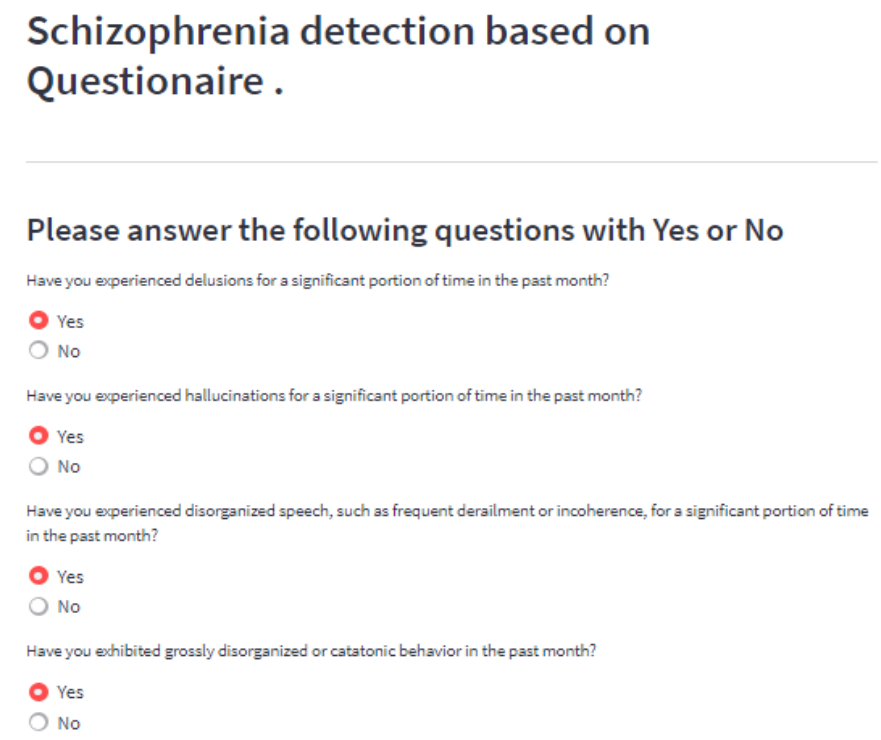


Fig A2

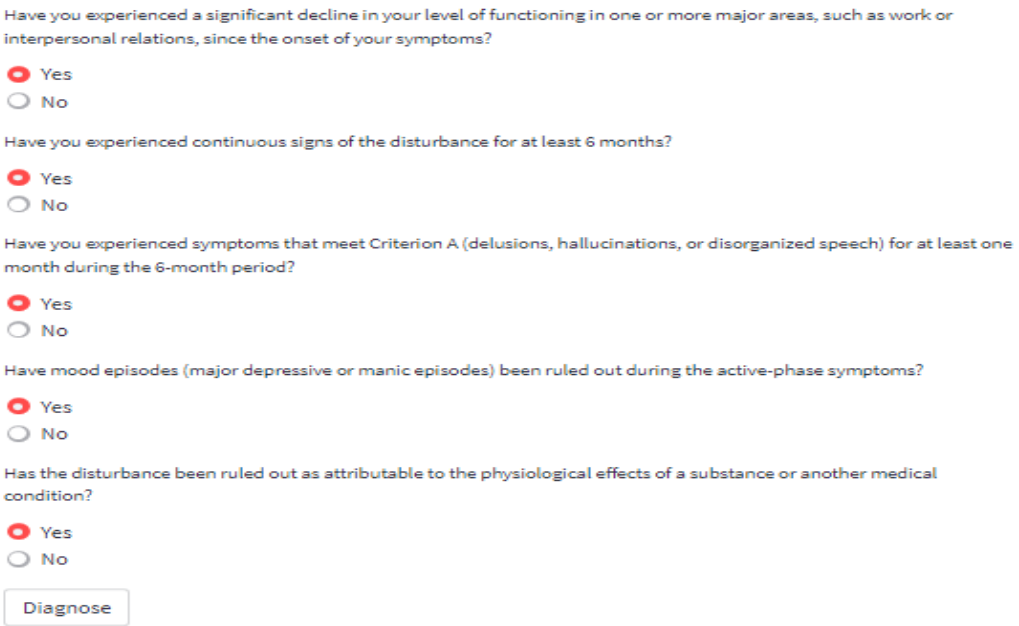


Fig A3

3.1.4 Decision Making

The choice of selecting the upper 15 top features was made after the feature extraction.

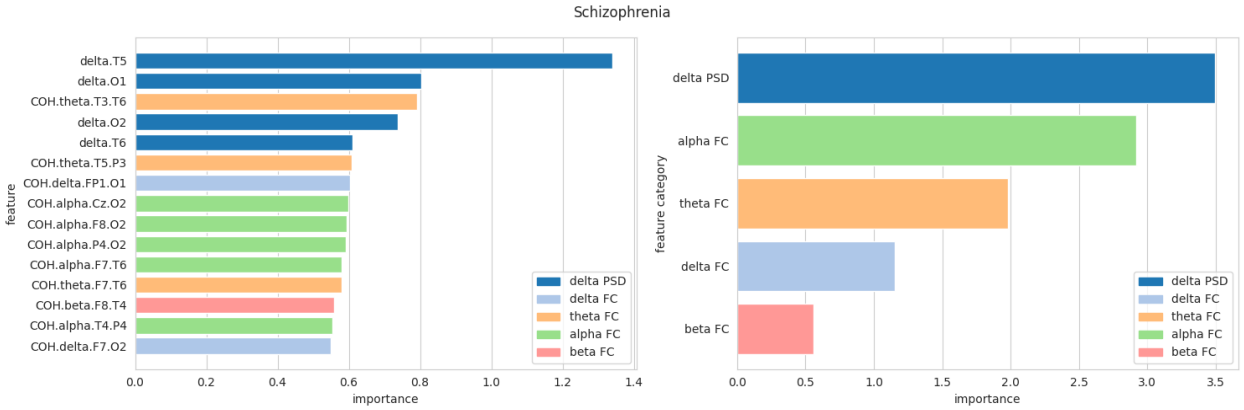


Fig A4

The model Schizophrenia.sav trained on the above dataset is used by the prediction function. The Prediction Function passes the PSD and FC input Values to the model and shows a result dialogue based on model Prediction.

3.1.5 Questionnaire

For the texted based self-assessment, the decision making is done on the basis of the following questions and conditions obtained from DSM-5.

Have you experienced delusions for a significant portion of time in the past month?
Have you experienced hallucinations for a significant portion of time in the past month?
Have you experienced disorganized speech, such as frequent derailment or incoherence, for a significant portion of time in the past month?
Have you exhibited grossly disorganized or catatonic behavior in the past month?
Have you experienced negative symptoms, such as diminished emotional expression or avolition, for a significant portion of time in the past month?
Have you experienced a significant decline in your level of functioning in one or more major areas, such as work or interpersonal relations, since the onset of your symptoms?
Have you experienced continuous signs of disturbance for at least 6 months?
Have you experienced symptoms that meet Criterion A (delusions, hallucinations, or disorganized speech) for at least one month during the 6-month period?
Have mood episodes (major depressive or manic episodes) been ruled out during the active-phase symptoms?
Has the disturbance been ruled out as attributable to the physiological effects of a substance or another medical condition?

If the answer to at least two of questions 1-5 is yes and the answer to question 6 is yes, it could be indicative of schizophrenia. Additionally, if the answer to questions 7 and 8 is yes and the disturbance cannot be attributed to the effects of a substance or medical condition, the diagnosis of schizophrenia could be confirmed. Finally, if mood episodes have been ruled out or have been present for only a minority of the total duration of the active and residual periods of the illness, the diagnosis of schizophrenia may be more likely. It is important to note that a trained mental health professional should be consulted for an accurate diagnosis.

3.2 Anxiety Detection

The top features for Anxiety which were extracted after feature importance calculation are being used to train the SVM model. That Model is being used to predict Anxiety. User enters 15 values. These values are passed to the SVM model. The result is when the test button is clicked over the GUI.

3.2.1 Preparing Dataset

Dataset used for Anxiety disorder model training contains these features:

delta.O2	COH.beta.F4.O2	theta.C3
gamma.T4	COH.delta.T5.P3	delta.O1
COH.beta.F8.O2	gamma.F8	theta.Cz
delta.T5	COH.alpha.P4.O2	COH.theta.T5.P3
COH.alpha.F8.T4	beta.FP2	COH.gamma.T5.P3

These PSD and FC values are used for prediction of Anxiety.
The Y or Outcome is encoded with ‘0’ for healthy control and ‘1’ for Anxiety Disorder.
Healthy control n = 95
Anxiety disorder n = 107

3.2.2 GUI

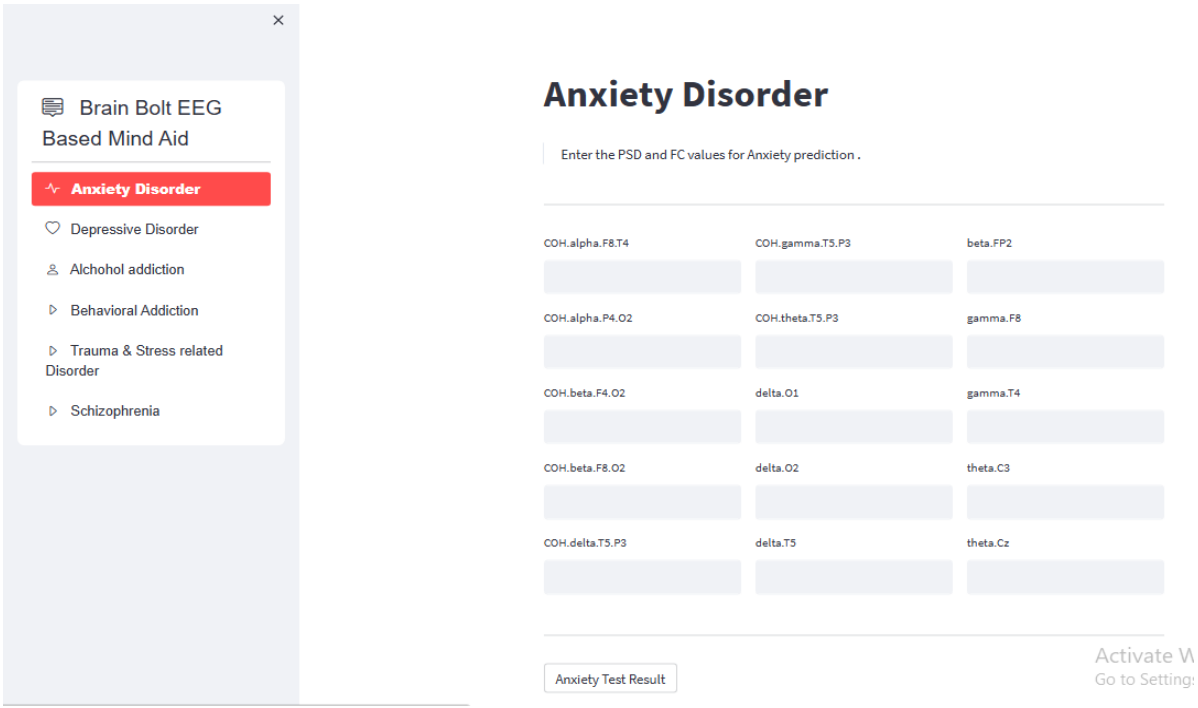


Fig B1

3.2.3 Self-assessment GUI

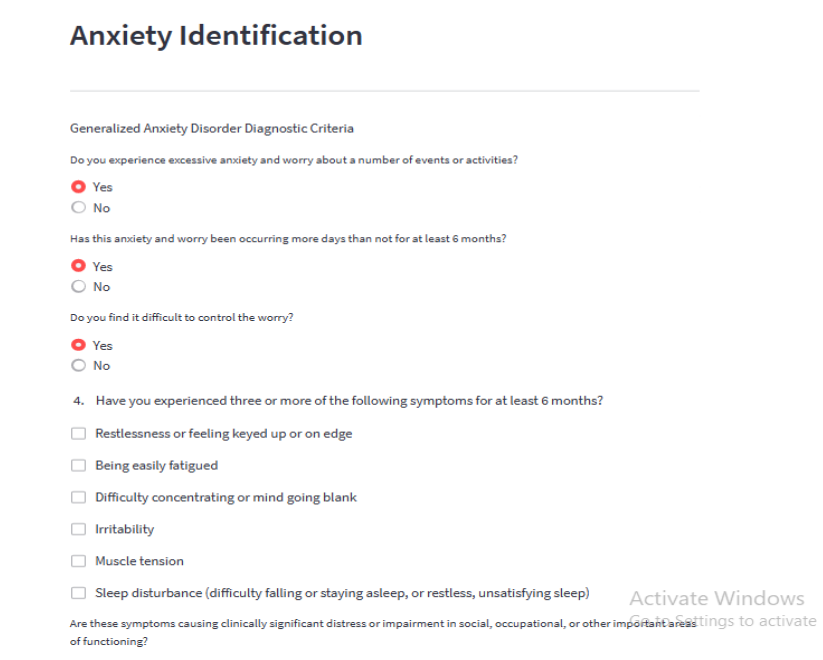


Fig B2

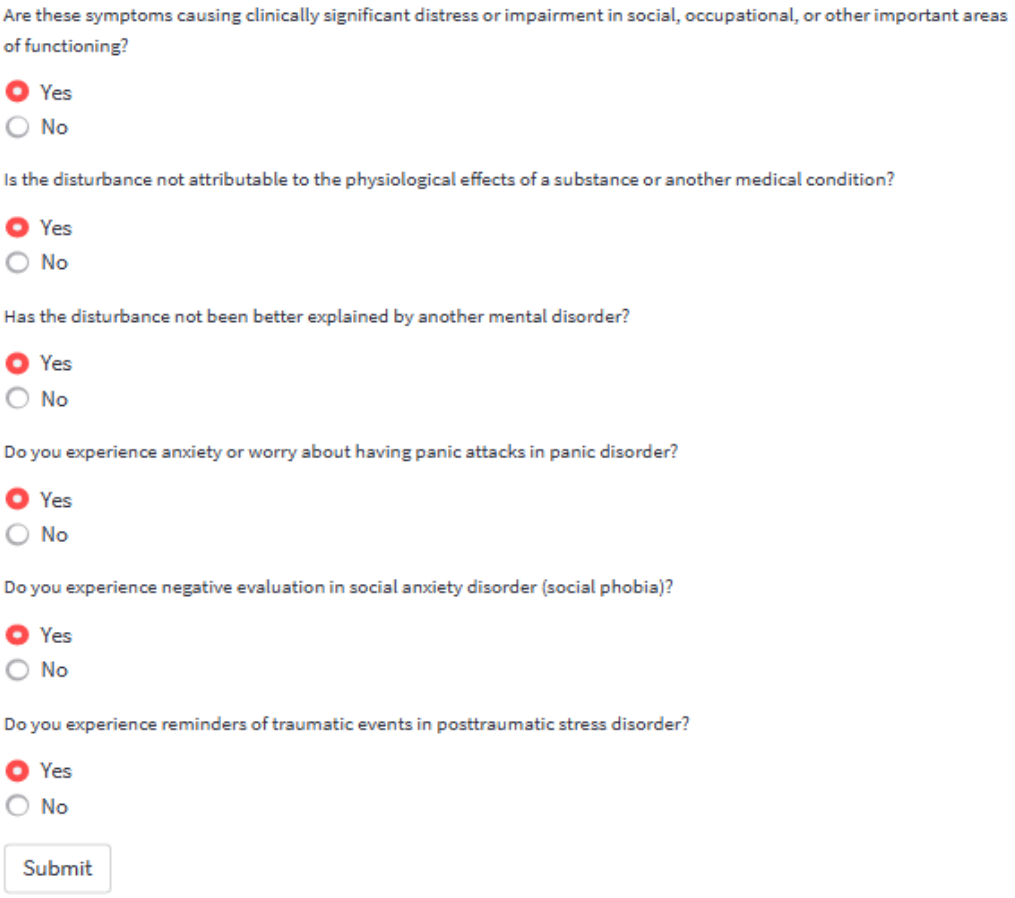


Fig B3

3.2.4 Decision Making

The choice of selecting the upper 15 top features was made after the feature extraction.

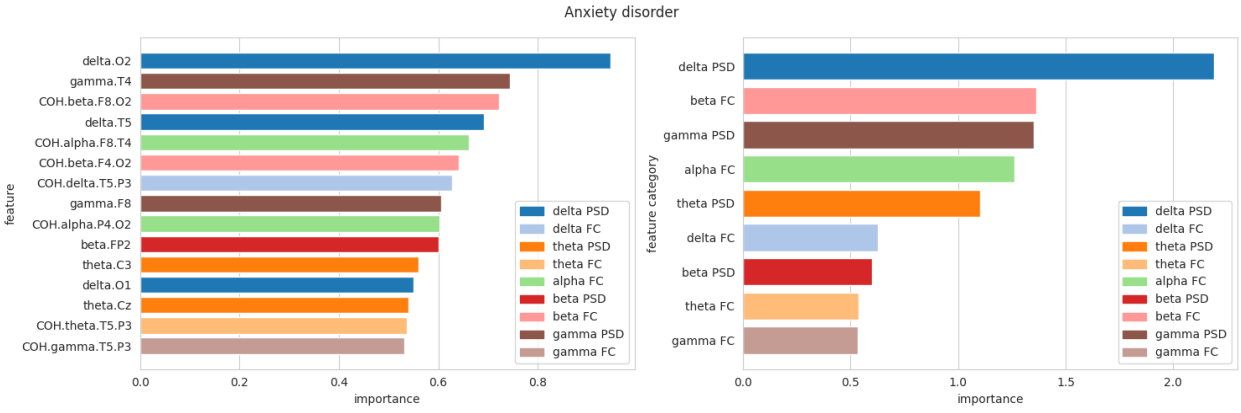


Fig B4

The model Anxiety.sav trained on the above dataset is used by the prediction function. The Prediction Function pass the PSD and FC input Values to the model and shows a result dialogue, Based on the model prediction.

3.2.5 Questionnaire

For the texted based self-assessment, the decision making is done based on the following questions and conditions obtained from DSM-5.

1. Do numerous events or activities cause you undue concern or worry?
2. Has it been at least six months since this concern and worry have occurred more days than not?
3. Is it tough for you to restrain your worry?
4. Have you had three or more of the following symptoms for at least six months impatience, muscle tension, restlessness, feeling tense or on edge, being quickly exhausted, having trouble focusing or going blank?
5. Are these symptoms impairing social, vocational, or other critical areas of functioning or producing clinically significant distress?
6. Is the ailment not caused by another medical condition or the physiological effects of a substance?
7. Could a different mental illness not explain the problem more effectively?
8. Do you suffer from anxiety or worry about having a panic attack?
9. Do you experience negative evaluation in social anxiety disorder (social phobia)?
10. Do you suffer from posttraumatic stress disorder and remember horrific events?

When symptoms have been present for at least six months and the responses to questions 1 through 5 are yes and the responses to questions 6 through 10 are no, generalized anxiety disorder may be suspected. However, a correct diagnosis may only be made following a complete evaluation by a trained mental health practitioner.

3.3 Trauma and stress related disorder detection

The top features for Trauma and stress related disorder which were extracted after feature importance calculation are being used to train the SVM model. That Model is being used to predict Trauma and stress related disorders. User enters 15 values. These values are passed to the SVM model. The result is when the test button is clicked over the GUI.

3.3.1 Preparing Dataset

Dataset being used for Trauma and stress related disorder model training contains the selected useful features:

COH.alpha.FP2.F8	COH.delta.P3.O2	highbeta.O1
COH.delta.FP2.F7	COH.alpha.F8.T4	COH.delta.FP1.F8
COH.alpha.F7.T6	beta.FP2	COH.delta.F7.T3
delta.T5	COH.beta.FP2.O1	beta.FP1
COH.beta.FP2.F8	COH.alpha.O1.O2	COH.alpha.Fz.F4

These PSD and FC values are used for prediction of Trauma and stress related disorder.
The Y or Outcome is encoded with '0' for healthy control and '1' for Trauma and Stress related disorder.
Healthy control n = 95
Trauma and Stress related disorder n = 128

3.3.2 GUI

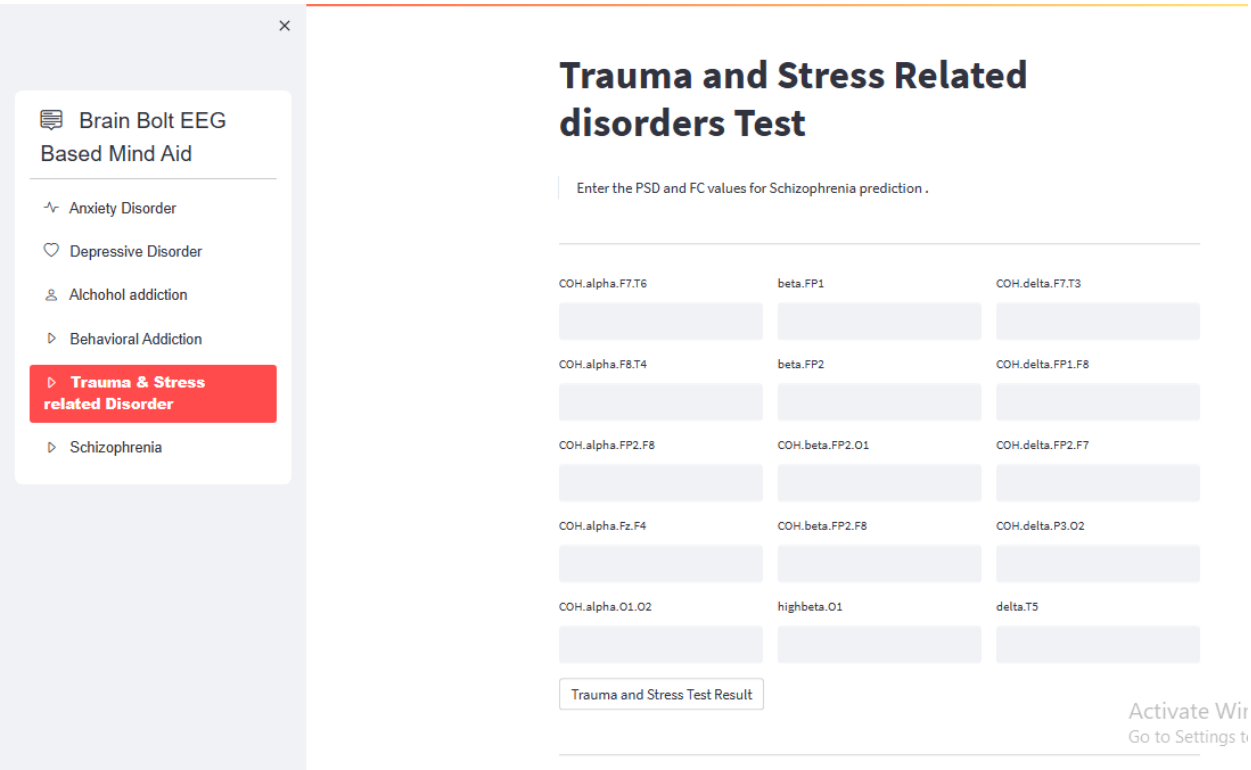


Fig C1

3.3.3 Self-assessment GUI

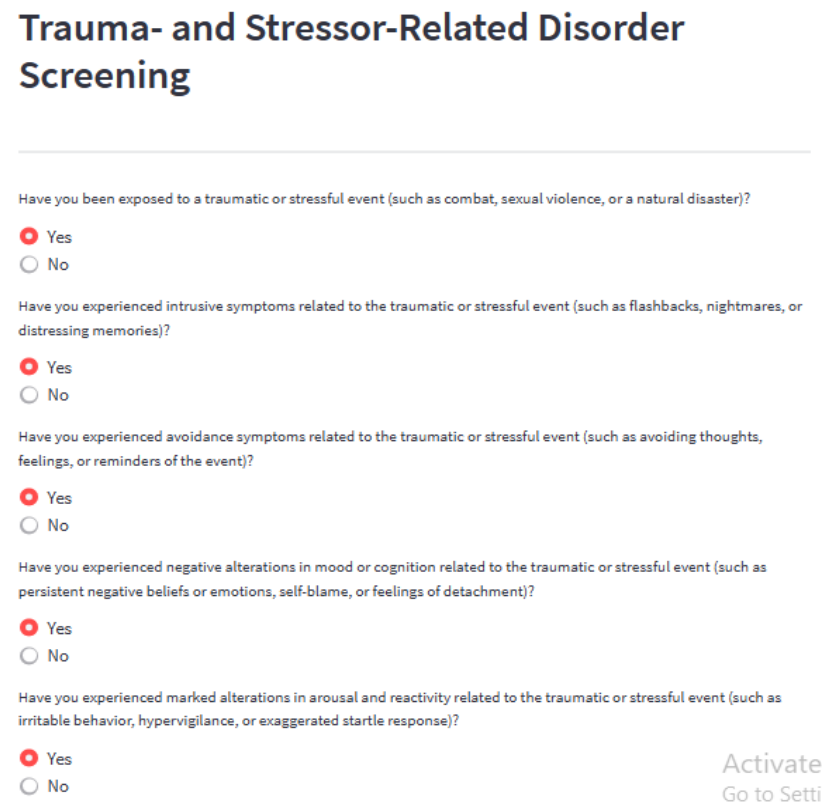


Fig C2

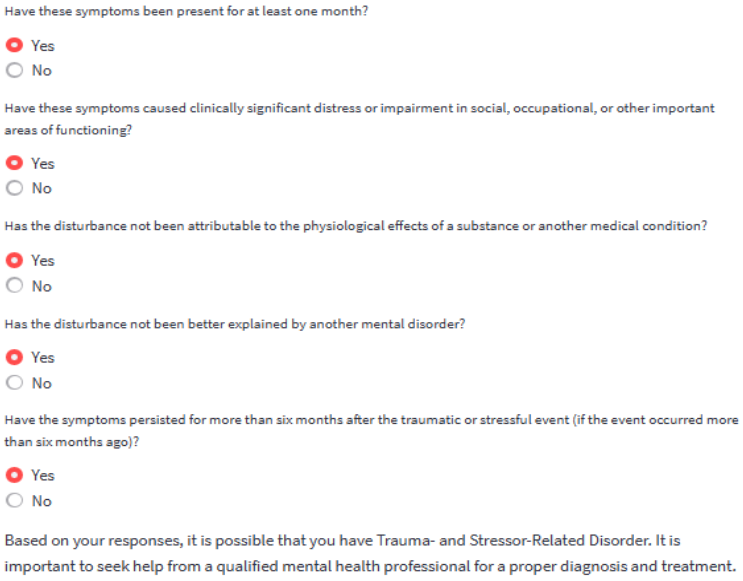


Fig C3

3.3.4 Decision Making

The choice of selecting the upper 15 top features was made after the feature extraction.

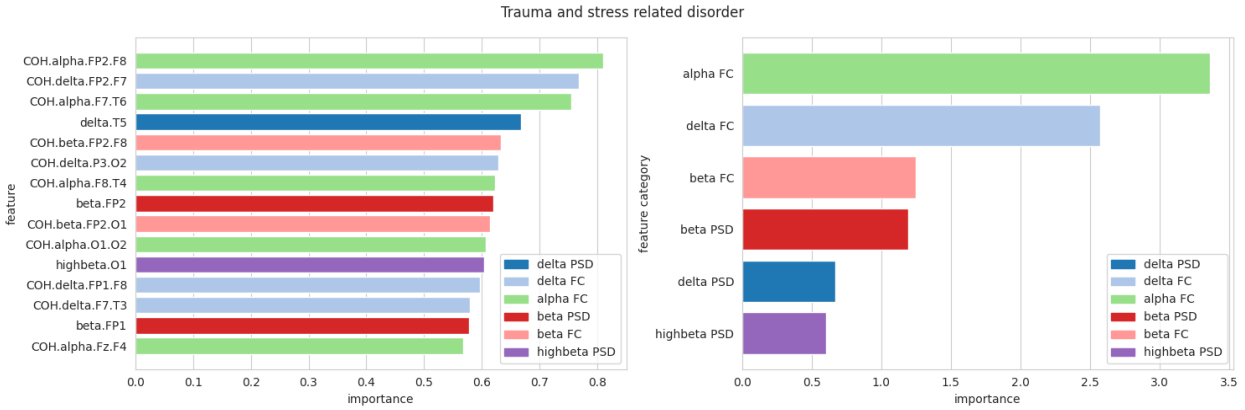


Fig C4

The model Trauma_and_stress_disorder.sav trained on the above dataset is used by the prediction function. The Prediction Function passes the PSD and FC input Values to the model and shows a result based on conditions given in DSM-5.

3.3.5 Questionnaire

For the texted based self-assessment, the decision making is done on the basis of the following questions and conditions obtained from DSM-5.

1. Have you ever been exposed to a traumatic or stressful incident (such a natural disaster, sexual assault, or combat)?
2. Have you had invasive symptoms (such flashbacks, nightmares, or upsetting memories) from the traumatic or stressful event?
3. Has the traumatic or stressful event caused you to suffer avoidance symptoms (such as avoiding thoughts, feelings, or reminders of the event)?
4. Have you seen any unfavorable changes in your mood or cognition as a result of the traumatic or stressful occurrence (such as enduring negative thoughts or sentiments, self-blame, or detachment feelings)?
5. Have traumatic or stressful incidents caused noticeable changes in your arousal and reactivity (such as irritable behavior, hypervigilance, or an excessive startle response)?
6. Has it been at least a month since these symptoms first appeared?
7. Have these symptoms significantly impacted your ability to operate in your social, professional, or other essential areas?
8. Has the disruption not been caused by a medical condition or a drug's physiological effects?
9. Hasn't another mental illness been able to explain the disruption more effectively?

10. If the traumatic or stressful incident happened more than six months ago, have the symptoms lasted for longer than that time?

A diagnosis of Trauma- and Stressor-Related Disorder may be made if the responses to questions 1 through 5 are positive and questions 6 through 9 are negative, and if the symptoms have been present for at least one month and have produced clinically substantial distress or impairment. However, a correct diagnosis may only be made following a complete evaluation by a trained mental health practitioner.

3.4 Depressive Disorder Detection

The top features for Depressive disorder which were extracted after feature importance calculation are being used to train the SVM model. That Model is being used to predict Depressive Disorder. User enters 15 values. These values are passed to the SVM model. The result is when the test button is clicked over the GUI.

3.4.1 Preparing Dataset

Dataset being used for Depressive Disorder model training contains the selected useful Features:

delta.O1	COH.alpha.F8.T4	alpha.T4
COH.alpha.Pz.P4	COH.beta.F8.O1	COH.alpha.P4.O2
COH.theta.F7.T6	COH.theta.FP1.O2	delta.T5
delta.O2	COH.theta.T5.P3	beta.T3
COH.alpha.T5.T6	COH.theta.T3.P4	COH.beta.P3.O2

These PSD and FC values are used for prediction of Depressive disorder.
The Y or Outcome is encoded with ‘0’ for healthy control and ‘1’ for Depressive Disorder.
Healthy control n = 95
Depressive disorder n = 199

3.4.2 GUI

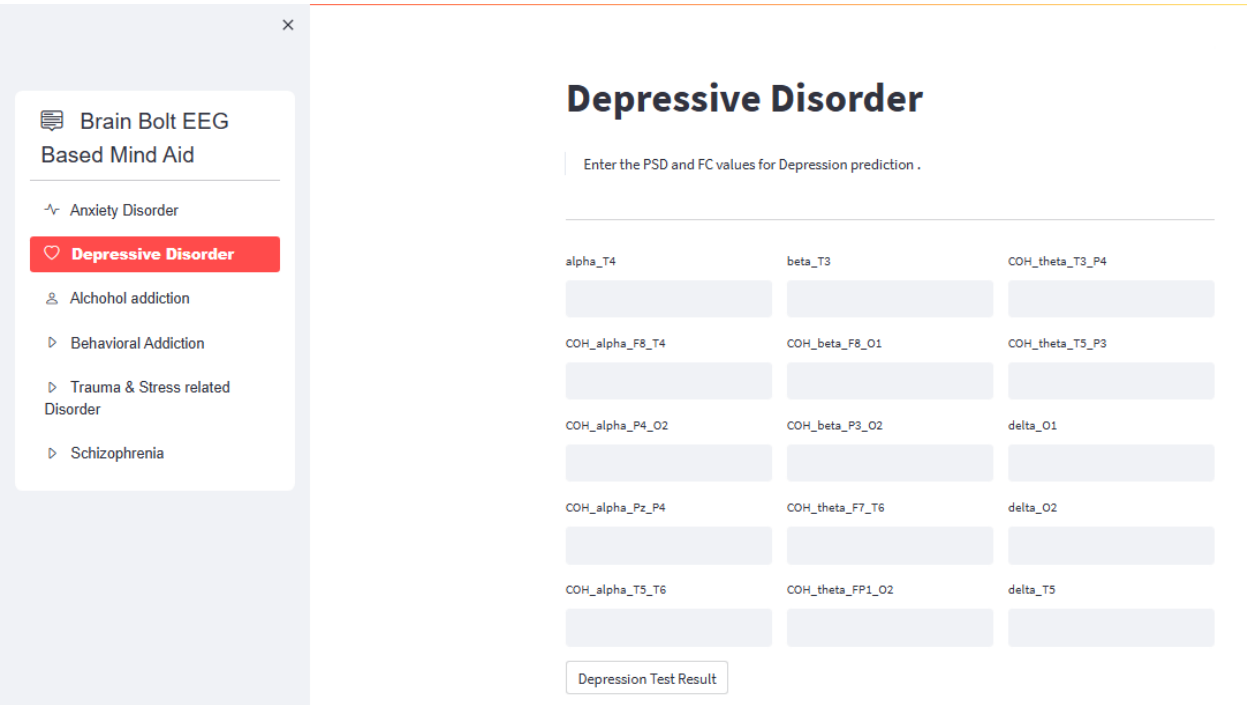


Fig D1

3.4.3 Questionnaire

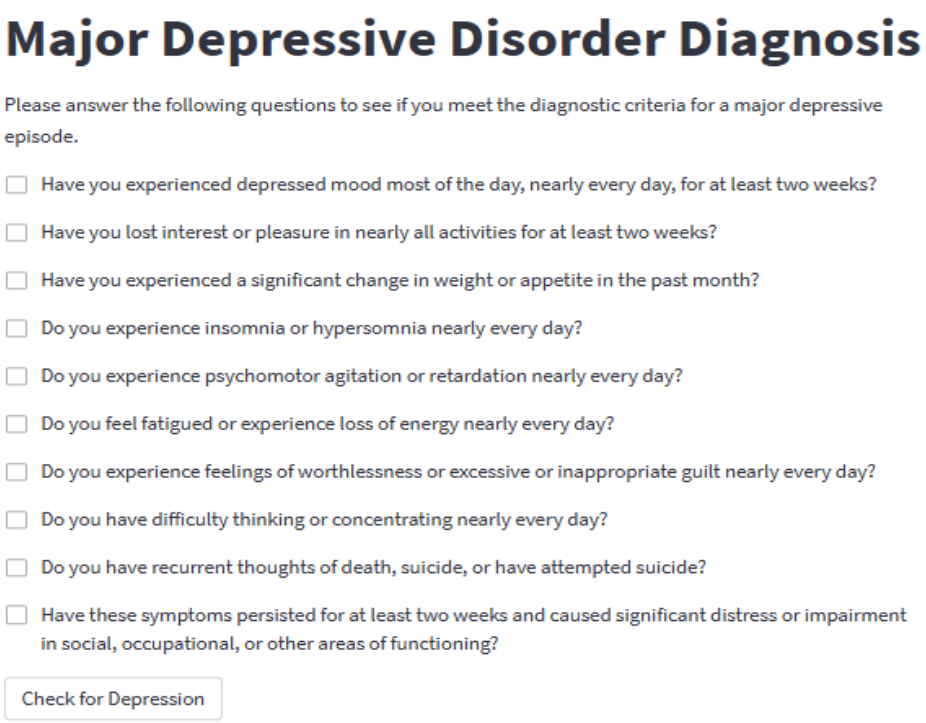


Fig D2

3.4.4 Decision Making

The choice of selecting the upper 15 top features was made after the feature extraction.

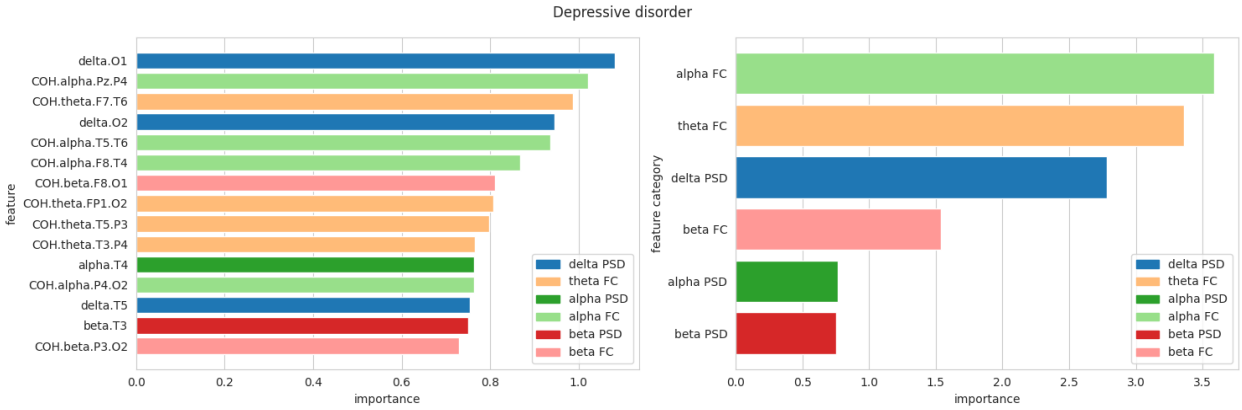


Fig D3

The model Depressive_disorder.sav trained on the above dataset is used by the prediction function. The Prediction Function passes the PSD and FC input Values to the model and shows a result dialogue based on model Prediction.

3.4.5 Questionnaire

For the texted based self-assessment, the decision making is done on the basis of the questions derived from criteria given in DSM-5.

Have you been feeling sad or empty most of the day, almost every day, for at least 2 weeks?
Criterion A1 for Major Depressive Disorder Depressed mood most of the day, nearly every day, for at least 2 weeks.
Have you lost interest or pleasure in almost all the activities that you usually enjoy, for at least 2 weeks?
Criterion A2 for Major Depressive Disorder Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day, for at least 2 weeks.
Have you experienced significant weight loss or weight gain, or changes in appetite almost every day, for at least 2 weeks?
Criterion A3 for Major Depressive Disorder Significant weight loss when not dieting or weight gain, or decrease or increase in appetite nearly every day, for at least 2 weeks.
Have you been having difficulty sleeping or sleeping too much almost every day, for at least 2 weeks?
Criterion A4 for Major Depressive Disorder Insomnia or hypersomnia nearly every day, for at least 2 weeks.
Have you been feeling restless or physically slowed down almost every day, for at least 2 weeks?
Criterion A5 for Major Depressive Disorder Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings), for at least 2 weeks.

Have you been feeling tired or lacking in energy almost every day, for at least 2 weeks?
Criterion A6 for Major Depressive Disorder Fatigue or loss of energy nearly every day, for at least 2 weeks.
Have you been feeling worthless or guilty almost every day, for at least 2 weeks?
Criterion A7 for Major Depressive Disorder Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day, for at least 2 weeks.
Have you been having trouble concentrating or making decisions almost every day, for at least 2 weeks?
Criterion A8 for Major Depressive Disorder Diminished ability to think or concentrate, or indecisiveness, nearly every day, for at least 2 weeks.
Have you had thoughts of death or suicide, or have you attempted suicide or made a plan to commit suicide?
Criterion A9 for Major Depressive Disorder Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.
Have these symptoms caused significant distress or impairment in your social, occupational, or other important areas of functioning?
Criterion B for Major Depressive Disorder The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

3.5 Alcohol use disorder Detection

The top features for Alcohol use disorder which were extracted after feature importance calculation are being used to train the SVM model. That Model is being used to predict Alcohol use disorder. User enters 15 values. These values are passed to the SVM model. The result is when the test button is clicked over the GUI.

3.5.1 Preparing Dataset

Dataset being used for Alcohol use disorder model training contains the selected useful features:

beta.FP1	beta.T3	COH.delta.C3.Cz
gamma.T4	COH.alpha.P4.T6	COH.alpha.FP2.O2
gamma.F8	COH.alpha.Pz.T6	gamma.T6
COH.delta.O1.O2	gamma.P3	COH.delta.T6.O1
COH.highbeta.O1.O2	COH.alpha.T3.T4	gamma.C3

These PSD and FC values are used for prediction of Alchohol use disorder .
The Y or Outcome is encoded with ‘0’ for healthy control and ‘1’ for Alchohol use disorder.
Healthy control n = 95
Alchohol use disorder n = 93

3.5.2 GUI

Figure E1

3.5.3 Self-Assessment GUI

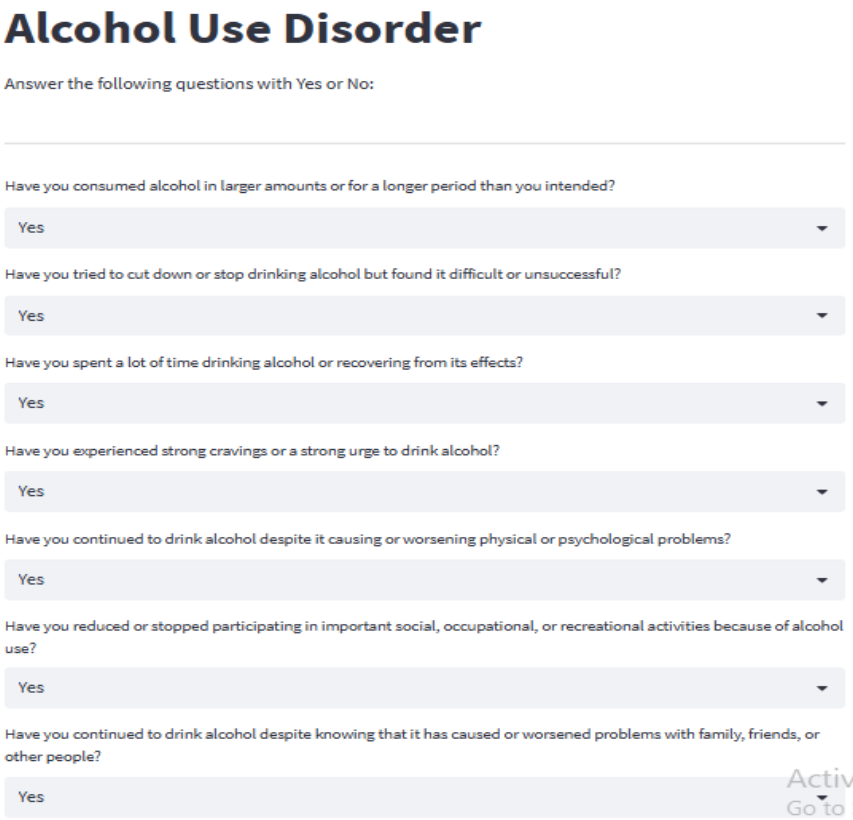


Fig E2

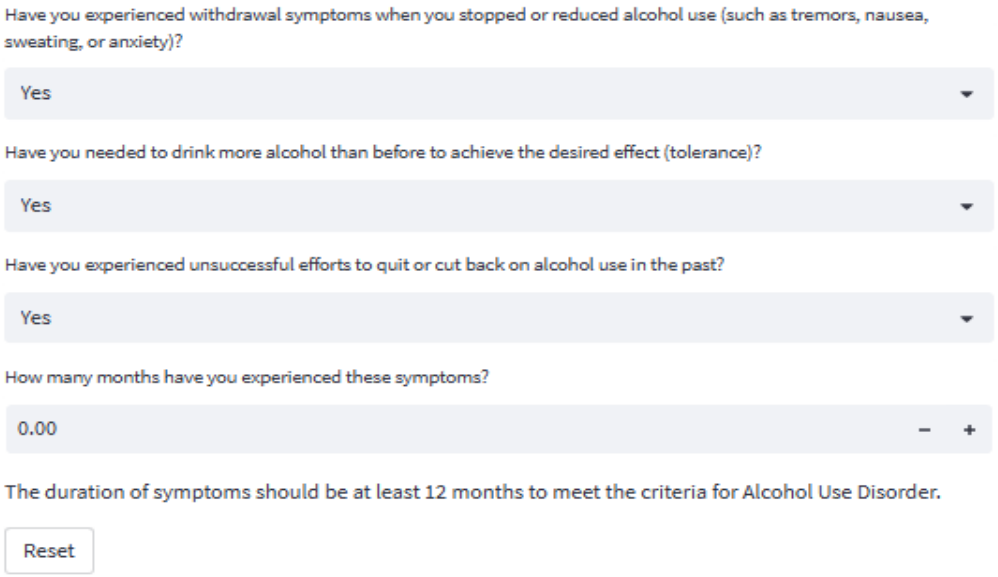


Fig E3

3.5.4 Decision Making

The choice of selecting the upper 15 top features was made after the feature extraction.

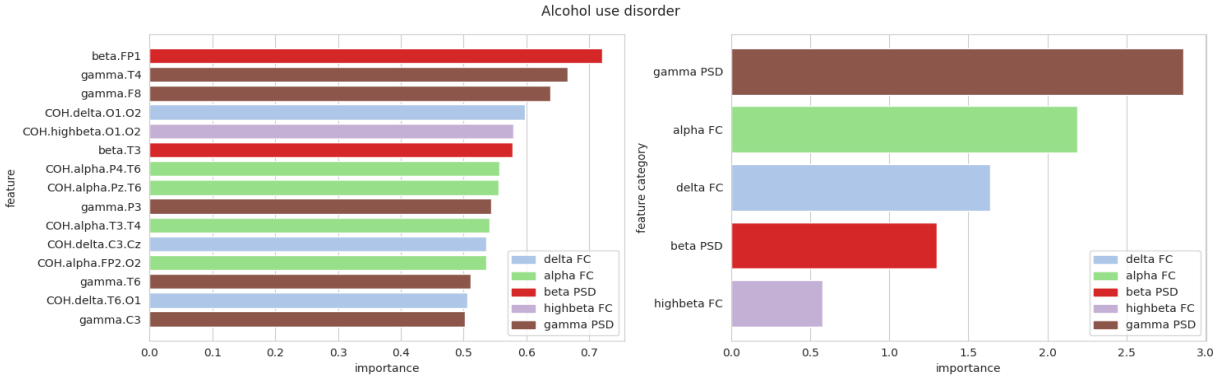


Fig E4

The model Alchohol_use_disorder.sav trained on the above dataset is used by the prediction function. The Prediction Function passes the PSD and FC input Values to the model and shows a result dialogue based on model Prediction.

3.5.5 Questionnaire

For the texted based self-assessment, the decision making is done based on the following questions and conditions obtained from DSM-5.

1. Have you drank more alcohol than you meant or for a longer period of time?
2. Have you tried to reduce your alcohol consumption or stop drinking altogether but failed?

3. Have you spent a lot of time drinking alcohol or recovering from its effects?
4. Have you experienced strong cravings or a strong urge to drink alcohol?
5. Have you continued to drink alcohol despite it causing or worsening physical or psychological problems?
6. As a result of your alcohol usage, have you cut back on or stopped engaging in significant social, professional, or recreational activities?
7. Have you continued to consume alcohol despite being aware that it has exacerbated or created conflicts with loved ones, friends, or others?
8. Have you experienced withdrawal symptoms when you stopped or reduced alcohol use (such as tremors, nausea, sweating, or anxiety)?
9. Have you needed to drink more alcohol than before to achieve the desired effect (tolerance)?
10. Have you experienced unsuccessful efforts to quit or cut back on alcohol use in the past?

A diagnosis of alcohol consumption disorder may be made if the responses to questions 1 through 10 are positive, the symptoms have been present for at least 12 months, and they have produced clinically significant distress or impairment. But only a skilled mental health practitioner with extensive evaluation can provide an accurate diagnosis.

3.6 Behavioral Addiction and gambling and Internet related addiction Detection

The top features for Behavioral Addiction which were extracted after feature importance calculation are being used to train the SVM model. That Model is being used to predict Behavioral Addiction. User enters 15 values. These values are passed to the SVM model. The result is when the test button is clicked over the GUI.

3.6.1 Preparing Dataset

Dataset being used for Behavioral Addiction model training contains the selected useful features:

dCOH.alpha.T3.T5	COH.delta.T3.C3	COH.alpha.T3.T6
COH.theta.T5.P3	COH.delta.P3.O2	COH.beta.F8.O1
COH.alpha.O1.O2	COH.alpha.Fz.T5	COH.alpha.T5.P3
COH.alpha.F8.O2	COH.beta.F8.O2	COH.alpha.T4.T5
delta.F8	COH.delta.F3.T3	COH.delta.F8.T4

These PSD and FC values are used for prediction of Behavioral Addiction.
The Y or Outcome is encoded with '0' for healthy control and '1' for Behavioral Addiction
Healthy control n = 95
Behavioral Addiction n = 93

3.6.2 GUI

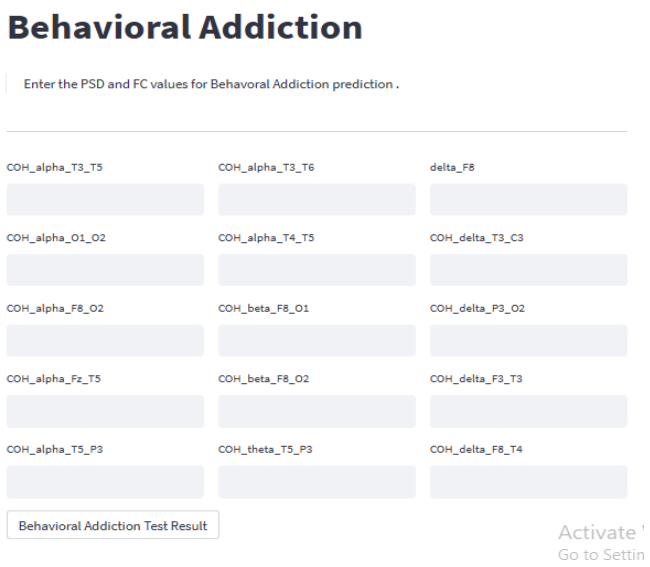


Fig F1

3.6.3 Self-assessment GUI

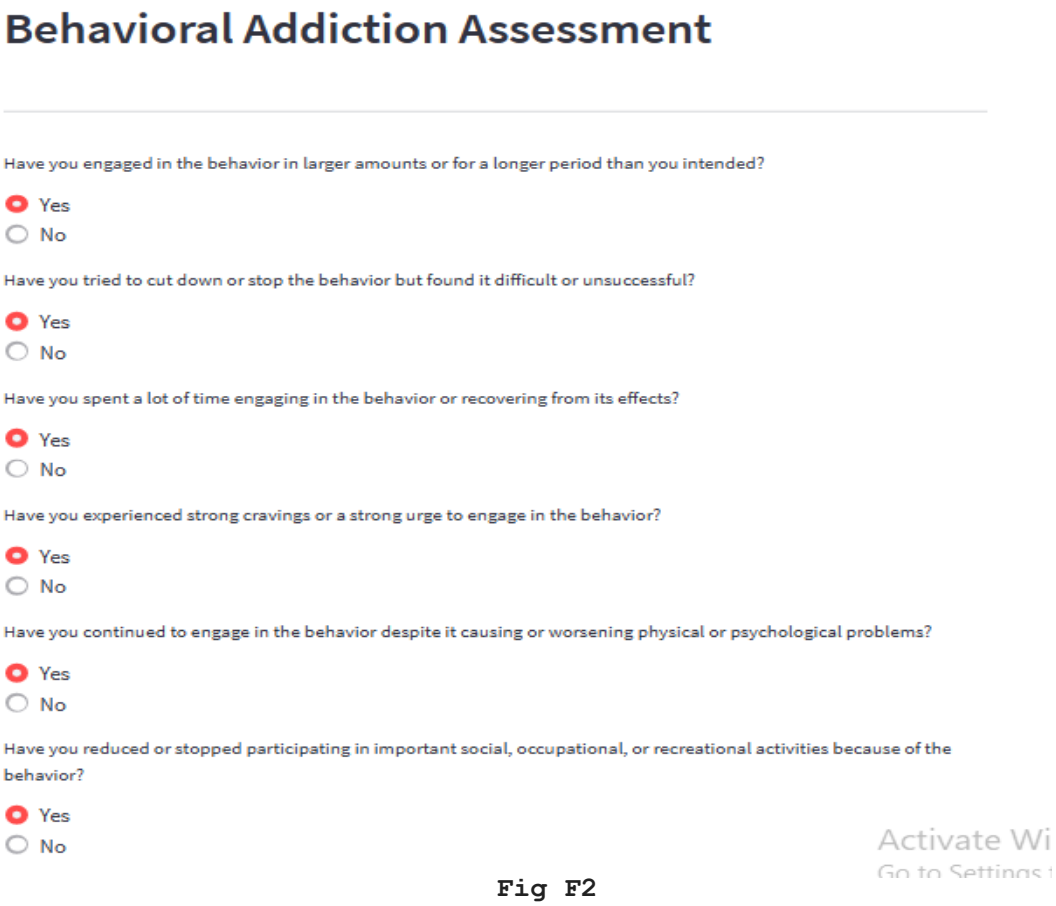


Fig F2

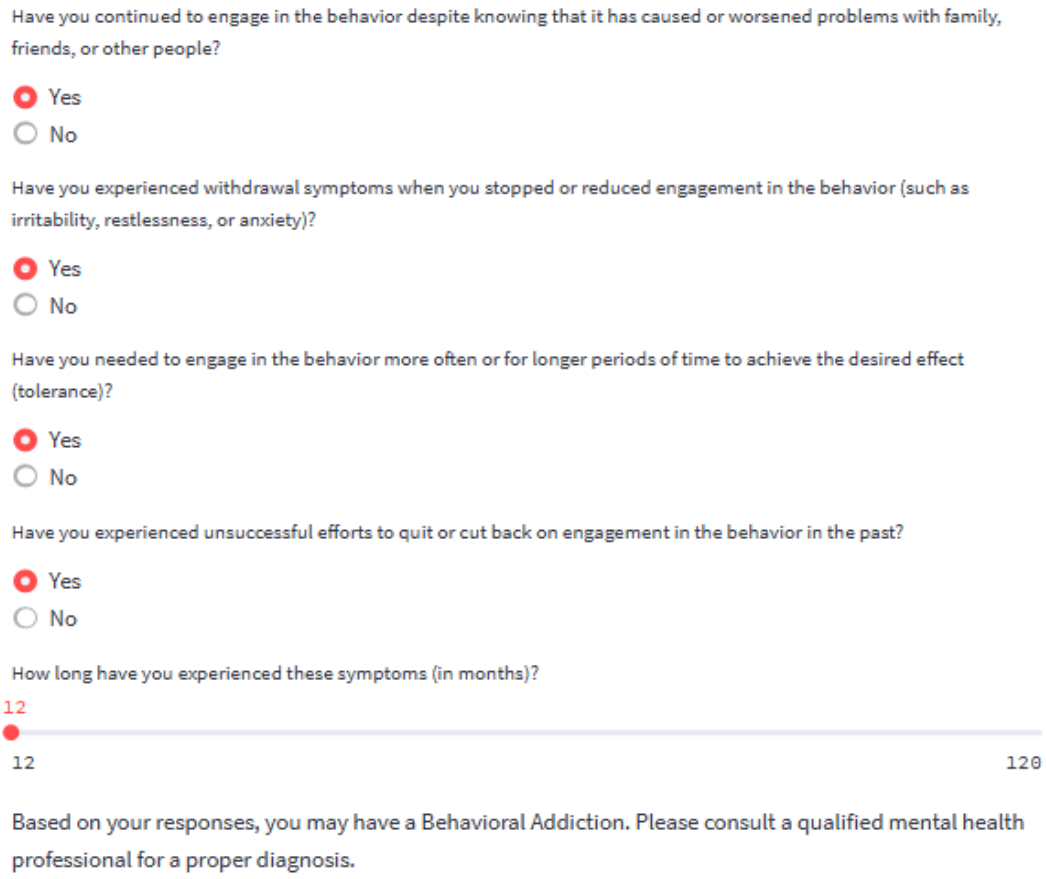


Fig F3

3.6.4 Decision Making

The choice of selecting the upper 15 top features was made after the feature extraction.

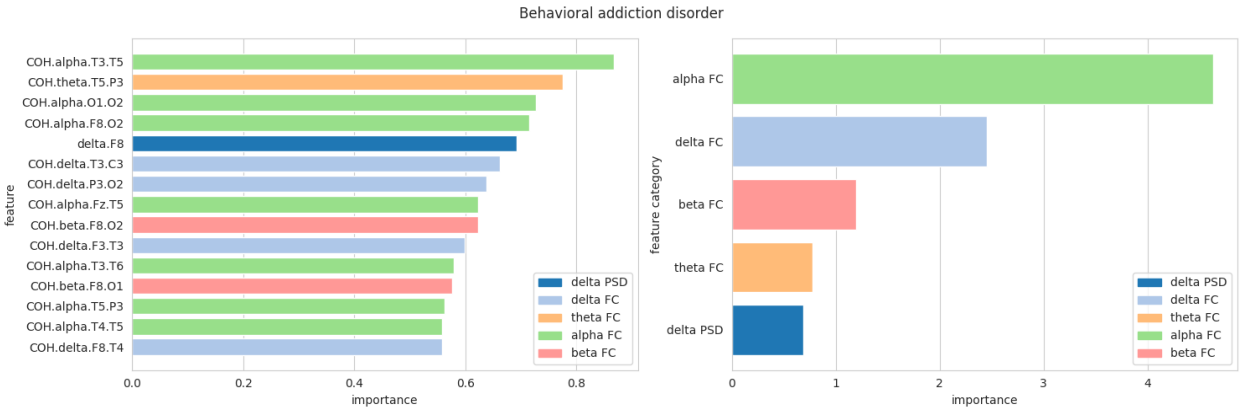


Fig F4

The model Behavioral_addiction.sav trained on the above dataset is used by the prediction function. The Prediction Function passes the PSD and FC input Values to the model and shows a result dialogue based on model Prediction.

3.6.5 Questionnaire

For the texted based self-assessment, the decision making is done based on the following questions and conditions obtained from DSM-5.

Note The DSM-5 does not currently recognize "behavioral addiction" as a separate diagnostic category. However, certain behaviors such as gambling disorder and internet gaming disorder are recognized as conditions in need of further research.

1. Have you continued to act in a certain way while knowing that it interferes with your relationships, career, or other vital aspects of your functioning?
2. Have you tried to stop or lessen the behavior in the past without success?
3. Have you felt a strong desire or craving to involve yourself in the behavior?
4. Have you needed to engage in the behavior more frequently or for longer periods of time to achieve the desired effect (tolerance)?
5. Have you experienced negative emotional states such as anxiety or irritability when attempting to stop or reduce the behavior?
6. Have you continued to engage in the behavior despite it causing physical or psychological harm?
7. Have you stopped doing or cut back on important social, professional, or recreational activities to engage in the behavior?
8. Have you continued to engage in the behavior despite knowing that it is causing problems?
9. Have you covered up how much you participated in the behavior by lying to loved ones, therapists, or others?
10. Have you experienced negative consequences such as legal problems, financial difficulties, or health problems because of the behavior?

If the answers to questions 1-10 are yes, and the symptoms have been present for at least 12 months and have caused clinically significant distress or impairment, then this may indicate a behavioral addiction. However, as mentioned earlier, the DSM-5 does not currently recognize a separate diagnostic category for behavioral addiction. A qualified mental health professional can assess the behavior and determine an appropriate diagnosis based on the individual's symptoms and circumstances.

Chapter 4 Evaluation of Models and Code Analysis

4.1 Schizophrenia

The classification report for the SVM model for Schizophrenia is:

	precision	recall	f1-score	support
0	0.78	0.74	0.76	19
1	0.80	0.83	0.82	24
accuracy			**0.79**	43
macro avg	0.79	0.79	0.79	43
weighted avg	0.79	0.79	0.79	43

4.2 Anxiety disorder

The classification report for the SVM model for Anxiety disorder is:

	precision	**recall**	**f1-score**	**support**
0	0.86	0.95	0.90	19
1	0.95	0.86	0.90	22
accuracy			**0.90**	41
macro avg	0.90	0.91	0.90	41
weighted avg	0.91	0.90	0.90	41

4.3 Trauma and Stress related Disorder

The classification report for the SVM model for Trauma and Stress related Disorder is:

	precision	recall	f1-score	support
0	0.82	0.74	0.78	19
1	0.82	0.88	0.85	26
accuracy			**0.82**	45
macro avg	0.82	0.81	0.81	45
weighted avg	0.82	0.82	0.82	45

4.4 Depressive Disorder

The classification report for the SVM model for Depressive Disorder is:

	precision	recall	f1-score	support
0	0.67	0.74	0.70	19
1	0.87	0.82	0.85	40
accuracy			**0.80**	**59**
macro avg	0.77	0.78	0.77	59
weighted avg	0.80	0.80	0.80	59

4.5 Behavioral Addiction Disorder

The classification report for the SVM model for Behavioral Addiction Disorder is:

	precision	recall	f1-score	support
0	0.71	0.79	0.75	19
1	0.76	0.68	0.72	19
accuracy			**0.74**	**38**
macro avg	0.74	0.74	0.74	38
weighted avg	0.74	0.74	0.74	38

4.6 Alcohol Use Disorder

The classification report for the SVM model for Alcohol Use Disorder is

	precision	recall	f1-score	support
0	0.76	0.84	0.80	19
1	0.82	0.74	0.78	19
accuracy			**0.79**	38
macro avg	0.79	0.79	0.79	38
weighted avg	0.79	0.79	0.79	38

4.7 Best hyperparameters

The values of the hyperparameters C, gamma, and kernel are included in the parameter grid that this pseudocode first defines for the search. The optimum hyperparameters are then found by searching the parameter grid with GridSearchCV and building an SVM model. A new SVM model is built using those hyperparameters once the best ones have been identified, and the model is then fitted to the training set of data. Finally, predictions are made based on the test data using the model, and the classification report is printed.

4.8 SVM

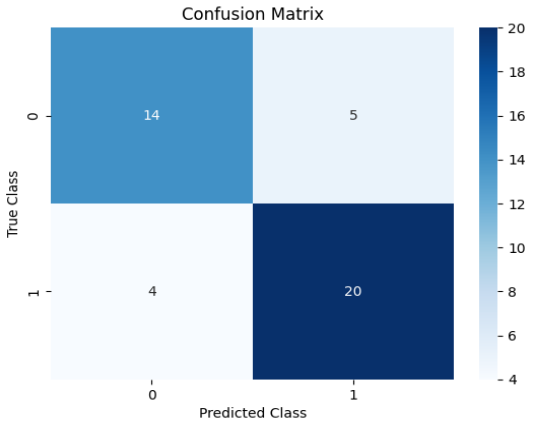
This is the general pseudo code for implementing SVM.

- Load the dataset and preprocess it as required (e.g., feature scaling, data normalization, etc.).
- Split the dataset into training and testing sets.
- Initialize the SVM model with the desired kernel and hyperparameters (e.g., regularization parameter, kernel coefficient, etc.).
- Train the SVM model using the training data.
- Use the trained SVM model to predict the classes of the testing data.
- Evaluate the performance of the SVM model using various metrics, such as accuracy, precision, recall, and F1-score.
- If the performance of the SVM model is not satisfactory, adjust the hyperparameters and repeat steps 4-6 until the desired performance is achieved.
- Once the optimal hyperparameters have been identified, train the SVM model on the entire dataset (including training and testing data) using these hyperparameters.
- Use the trained SVM model to make predictions on new, unseen data.

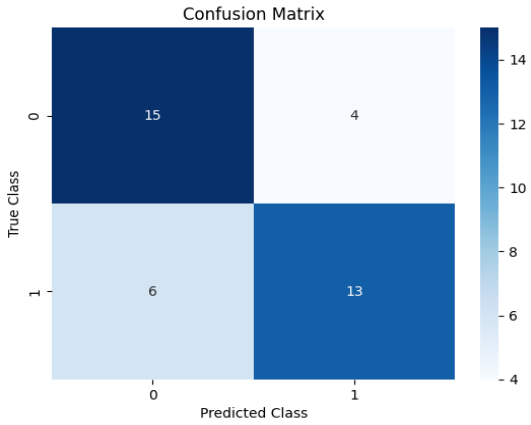
4.9 Confusion Matrices

The confusion matrices for each model are given below:

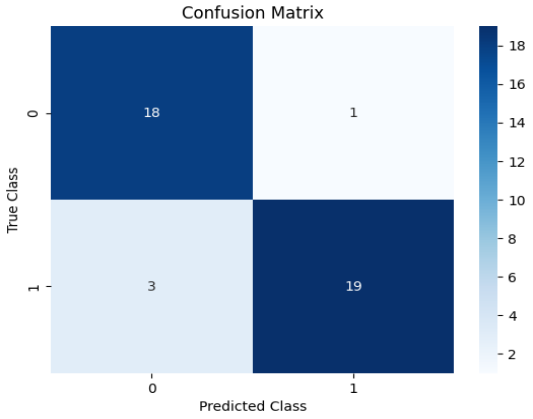
Schizophrenia



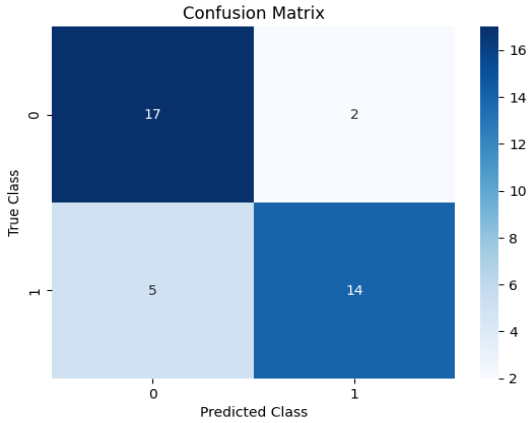
Behavioral addiction and gambling addiction and internet gaming addiction



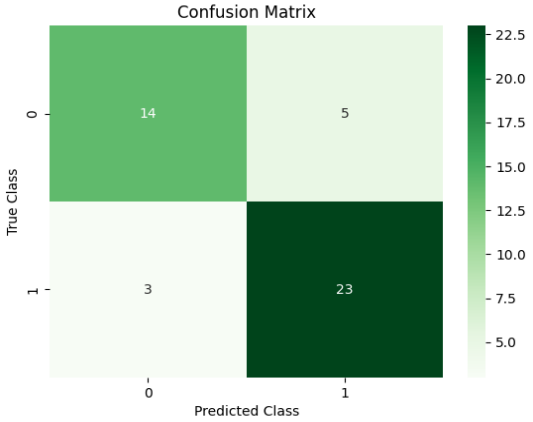
Anxiety disorder



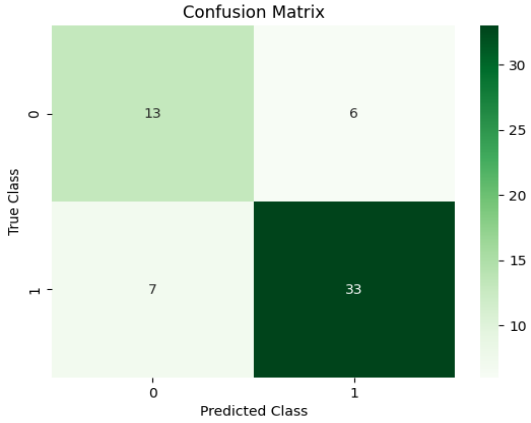
Alcohol Use Disorder



Trauma and Stress related disorders

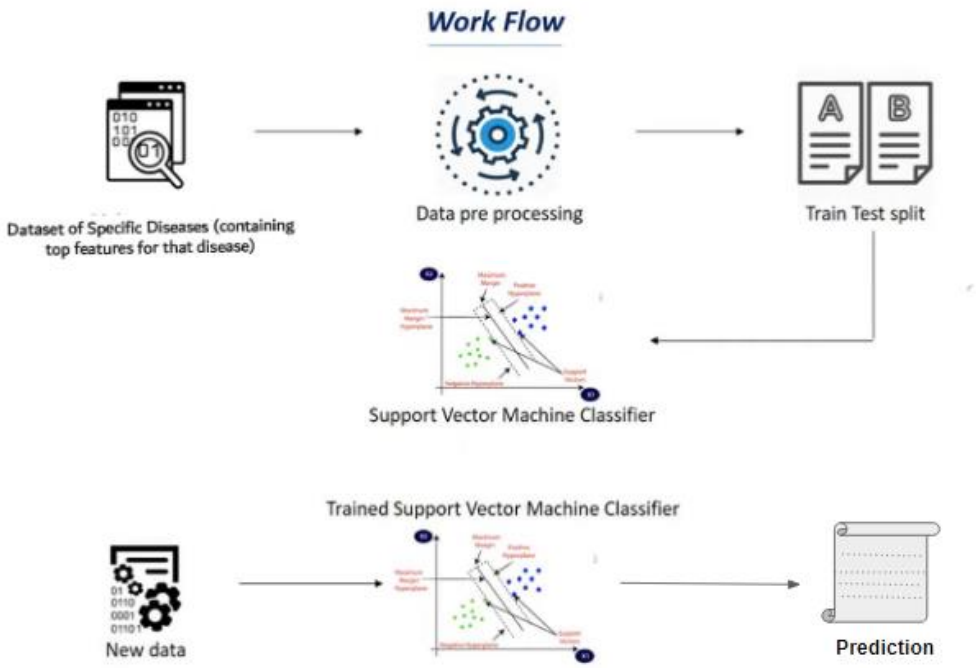


Depressive disorder

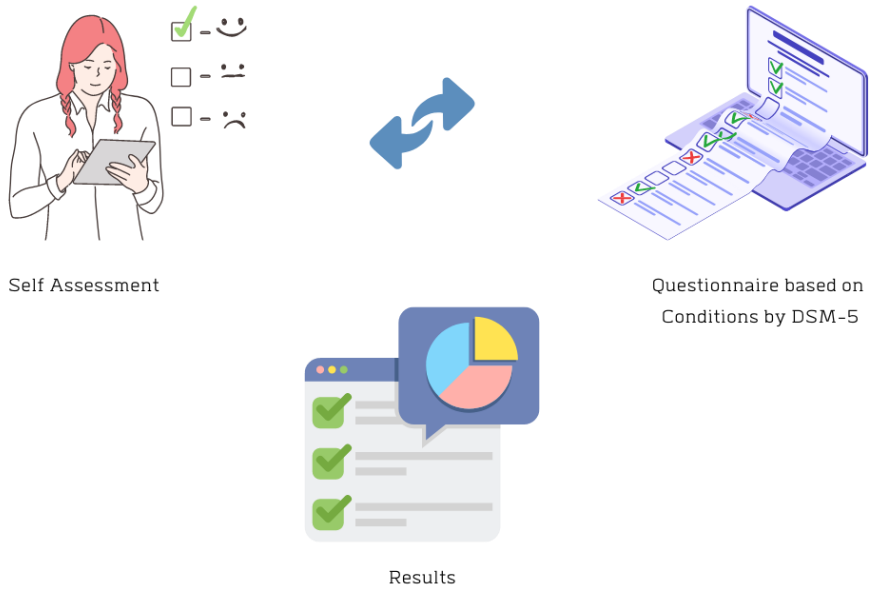


4.10 Workflow Diagrams

Diagram of using SVM model for prediction of psychiatric diseases.



Workflow diagram the questionnaire based on DSM-5 diagnostic criteria.



Chapter 5 Conclusion

In summary, our project focused on utilizing rest state QEEG data to detect common mental health disorders, such as schizophrenia, alcohol use disorder, anxiety disorders, depressive disorder, trauma and stress-related disorders, behavioral addiction, internet addiction, and gambling addiction. By using SVM, we achieved an accuracy of over 70% with good precision, recall, and f1 values. We also included a self-assessment option based on DSM 5 diagnostic criteria, allowing users to assess their mental health by answering 10 simple questions for each disease.

To improve the accuracy of our model, we utilized top features for binary classification for each disease with healthy control data. Our findings highlight the potential of using EEG-based analysis for mental health screening and diagnosis. However, future research and development are needed to increase the sample size and balance the data for further accuracy improvement.

Our GUI application built with Stream lit allows users to easily access the self-assessment feature and obtain assessments based on simple questions based on dsm-5 criteria and prediction scores for each disease by entering the QEEG values. Our project showcases the potential of pre-processed QEEG data for disease prediction using various ML algorithms, with SVM being a preferred model due to its ability to avoid overfitting. Overall, our proposed solution provides an effective and accessible method for mental health screening and diagnosis.

Chapter 6 Future Work

Future Milestones that need to be achieved to commercialize this project are given below:

6.1 Enhancing Classification Accuracy:

One important area of future work for this project is to further enhance the classification accuracy of the EEG-based disease detection model. This can be achieved by exploring advanced machine learning algorithms, optimizing feature selection techniques, and incorporating additional relevant data sources or biomarkers.

6.2 Extending the Dataset:

Expanding the dataset used for training and testing the model is crucial for improving its generalization and robustness. Collecting more diverse and representative EEG data from larger populations would help in capturing the full spectrum of neurological diseases and improving the overall performance of the classification model.

6.3 Real-Time Disease Detection:

A significant advancement would be to develop a real-time disease detection system using EEG data. This would involve integrating the trained model into a live data processing pipeline, enabling clinicians and healthcare professionals to obtain instant disease predictions and make prompt decisions regarding patient care and treatment.

6.4 Validation and Clinical Trials:

Conducting rigorous validation and clinical trials to evaluate the performance and effectiveness of the EEG-based disease detection model is an essential step towards its adoption in clinical practice. Collaborating with medical professionals and institutions to validate the model's accuracy, sensitivity, specificity, and reliability would provide valuable insights for its real-world application.

6.5 Integration with Clinical Decision Support Systems:

Integrating the EEG-based disease detection model with existing clinical decision support systems would enhance its utility and usability. This integration would enable seamless integration of EEG data analysis into the diagnostic workflow, supporting clinicians in making more informed and accurate diagnoses.

Plagiarism Report

Brain_Bolt_-_EEG_Based_Mind_Aid_F_-_Google_docs

ORIGINALITY REPORT

9%	8%	4%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.docdroid.net Internet Source	1%
2	dokumen.pub Internet Source	1%
3	www.coursehero.com Internet Source	1%
4	academic.oup.com Internet Source	<1%
5	clickfit.com Internet Source	<1%
6	centauri.erin.utoronto.ca Internet Source	<1%
7	chel-center.ru Internet Source	<1%
8	www.library.umaine.edu Internet Source	<1%
9	www.mdpi.com Internet Source	<1%

10	Submitted to University of Westminster Student Paper	<1 %
11	etd.aau.edu.et Internet Source	<1 %
12	Submitted to University of Hull Student Paper	<1 %
13	nda.nih.gov Internet Source	<1 %
14	Submitted to University of Western Sydney Student Paper	<1 %
15	docshare.tips Internet Source	<1 %
16	apps.dtic.mil Internet Source	<1 %
17	utpedia.utp.edu.my Internet Source	<1 %
18	mdpi-res.com Internet Source	<1 %
19	www.hcisj.com Internet Source	<1 %
20	brage.bibsys.no Internet Source	<1 %
21	Submitted to Monash University Student Paper	<1 %

22	link.springer.com Internet Source	<1 %
23	rayinc.ca Internet Source	<1 %
24	Submitted to Brigham Young University Student Paper	<1 %
25	eprints.qut.edu.au Internet Source	<1 %
26	Panagiota Anna Chousou, Rahul Chattopadhyay, Vasiliki Tsampasian, Vassilios S. Vassiliou, Peter John Pugh. "Electrocardiographic Predictors of Atrial Fibrillation", Medical Sciences, 2023 Publication	<1 %
27	dspace.bracu.ac.bd Internet Source	<1 %
28	healthcenter.uoregon.edu Internet Source	<1 %
29	www.nature.com Internet Source	<1 %
30	www.ncbi.nlm.nih.gov Internet Source	<1 %
31	www.hivguidelines.org Internet Source	<1 %