國立臺灣大學電機資訊學院生醫電子與資訊學研究所

碩士論文

Graduate Institute of Biomedical Electronics and Bioinformatics
College of Electrical Engineering and Computer Science
National Taiwan University
Master Thesis

基於機器學習應用操弄後的腦波訊號來預測憂鬱症嚴重性頑固性及自殺風險

Manipulated EEG Signals Predict Severity of Depression, Refractoriness, and Suicide Risk in Patients with Major Depressive Disorder Based on Machine Learning Approaches

賴致安

Chih-An Lai

指導教授：陳中平博士

Advisor: Chung-Ping Chen, Ph.D.

共同指導教授：李正達博士

Co-Advisor: Cheng-Ta Li, M.D., Ph.D.

中華民國 112 年 7 月

July, 2023

doi:10.6342/NTU202301910

誌謝

能夠完成這篇論文，這兩年的時間使我從一個機器學習新手到能夠完成這篇研究論文，一路上受到許多人的幫助，在此我想表達衷心的感謝。首先，感謝我的指導教授陳中平教授與共同指導教授李正達醫師在這段期間的用心指導。感謝陳老師在這兩年來給予我穩定的學習環境及實驗室的資源，儘管在百忙之中還是會定期關心我的研究進度並且給予很大的自由空間來學習和研究，讓我能夠有良好的自主學習性。

感謝台北榮總精神科主任李正達醫師，李醫師關心我如同關心自己實驗室的學生一樣，分配我們研究內容以及每次開會時不停詢問以及修正我的研究進度與方向。就算平常醫院事務非常繁忙，還是會因為擔心我的進度而抽空與我個別討論。非常高興在碩士生涯有這個機會被醫師指導，一生受用無窮。另外，也非常感謝李醫師團隊中的學長姐及助理們，當我有問題時都會不厭其煩的教導我一些臨床上的專業知識，讓我對於跨領域之間的不熟悉能夠很快上手。

除此之外，也想要感謝我的好同學朱新源，能夠在這兩年的時間一起修課、一起玩樂、一起聊天互相討論研究，在各種大大小小的事上也幫助我許多，很高興這段時間有你在，期許未來我們都能飛黃騰達。也感謝各位口試委員能夠不辭千里來到博理館參加我的口試，並且給予許多寶貴的評論以及論文修改的意見。最後，感謝一路以來支持我的家人以及身邊的朋友們，有你們的各種支持、陪伴與良好的建議，才讓我能夠順利完成我的學業，謝謝你們。

賴致安 2023 年 7 月

doi:10.6342/NTU202301910

中文摘要

重度憂鬱症是一種反覆發作且嚴重影響生活質量的精神疾病，需要早期且有效的介入與治療。若未能及早緩解症狀，往往會變成慢性病。重度憂鬱症的症狀可能包括長時間的悲傷、極度疲勞、食慾和睡眠模式的改變、注意力困難、自我價值感的減低，甚至反覆出現有關死亡或自殺的念頭。部分患者患有輕度至中度嚴重的憂鬱症，但有些人可能有嚴重的憂鬱症狀（即多種症狀或在憂鬱評分量表中得分較高）。一些病患對抗憂鬱症藥物的治療反應不佳，可能需要個別化或其他進一步的治療方式，例如：重複經顱磁刺激（rTMS）。早期識別高度難治性（即對不同類型的抗憂鬱藥物反應不佳）的病患能夠避免浪費無效治療的時間並且及時決定下一步的抗憂鬱藥物治療（如 rTMS）。此外，重度憂鬱症中最極端的症狀為自殺意念，甚至是企圖自殺。及早識別病患症狀並及時介入對於應對自殺風險至關重要。因此，能夠在臨床治療之前預測重度憂鬱症患者的憂鬱嚴重程度、高度難治性和自殺風險，並提供個性化、精準和有效的治療，將是未來的重要技術發展。

本研究使用了 209 名重度憂鬱症患者的靜息和操弄的臨床腦電圖數據，訓練了多個機器學習算法，以預測憂鬱嚴重程度、高度難治性和自殺風險。我們的特徵集包括來自前額葉區域 7 個通道的電極（FP1、FP2、F7、F3、FZ、F4、F8），5 個子頻帶（Alpha、Beta、Delta、Theta、Gamma）和 6 個線性和非線性特徵（LLE、DFA、ApEn、KFD、HFD、Welch）。本研究所訓練出的機器學習模型能夠顯著預測憂鬱嚴重程度、高度難治性和自殺風險（校正後的 $p<0.05$），而最佳三個模型的準確率分別為 82.8%、86.7%和 97.6%。並依照模型重要特徵能夠發現，用於分類憂鬱嚴重程度的特徵位於 FP2 通道的 Alpha 頻帶。用於分類高度難治性的特徵主要分布在 FP1 通道的 Theta 頻帶和 F7 通道的 Beta 頻帶。用於分類自殺風險的特徵分佈在 FP1、F3 和 FZ 通道的 Beta 和 Gamma 頻帶。

本研究亦設計一個方便且易於操作的使用者介面，顯示了憂鬱嚴重程度、高度難治性和自殺風險的預測結果，以及每個模型在進行這些預測時所考慮的三個最重要特徵。該介面有助於患者更好地理解醫護人員所想要傳達的信息。

關鍵字：重度憂鬱症、腦電圖、憂鬱嚴重程度、高度難治性、自殺風險、機器學習

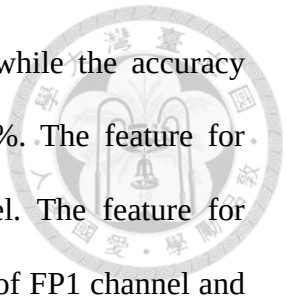
doi:10.6342/NTU202301910

ABSTRACT

Major depressive disorder (MDD) is a recurrent and highly disabled mental illness that warrants early and effective interventions. Without early symptomatic remission, MDD tends to be chronic. Symptoms of MDD may include prolonged periods of sadness, fatigue, changes in appetite and sleep patterns, difficulty concentrating, feelings of worthlessness, and even recurrent thoughts of death or suicide. Some patients have mild to moderate severity of depression, but some may have severe depression of severity (i.e., multiple symptoms or a higher rating in depression rating scale). A significant proportion of MDD individuals may not respond well to antidepressant treatments, requiring personalized or further treatments such as repetitive transcranial magnetic stimulation (rTMS). Early identifying patients with high levels of refractoriness (i.e., poor responses to different kinds of antidepressants) can avoid a waste of time on ineffective treatment and enable the timely decision for next-line antidepressant treatment (e.g., rTMS). Moreover, the most extreme symptoms of MDD are suicidal ideations and even attempts. Early recognition of symptoms and prompt intervention are vital in addressing this risk. Hence, predicting the severity of depression, high refractoriness of treatments, and suicide risk of MDD patients prior to clinical treatment, and delivering personalized, precise, and effective treatments, will be a crucial technological advancement in the future.

The present study used resting and modulated electroencephalography (EEG) data of 209 patients with MDD to train several machine learning algorithms to predict the severity of depression, high refractoriness, and suicide risk. Our feature set included 7 channels from the frontal region (FP1, FP2, F7, F3, FZ, F4, F8), 5 sub-bands (Alpha, Beta, Delta, Theta, Gamma), and 6 linear and nonlinear features (LLE, DFA, ApEn, KFD, HFD, Welch). The machine learning model in the study significantly predicted depression

doi:10.6342/NTU202301910



severity, high refractoriness, and suicide risk (corrected $p<0.05$), while the accuracy obtained on the best three models were 82.8%, 86.7%, and 97.6%. The feature for classifying depression severity lies in alpha band of FP2 channel. The feature for classifying high refractoriness is predominantly found in theta band of FP1 channel and beta band of F7 channel. The features for classifying suicide risk are distributed across beta and gamma band of FP1, F3, and FZ channel.

The study also developed a user-friendly interface for predicting the three dimensions of severity in MDD. The interface displays the predicting results for severity of depression, high refractoriness, and suicide risk, along with the top three most important features considered by each model in making these predictions. It also provides visualizations of EEG signals and explanations of the meaning behind the extracted features. As an aid for clinicians to identify levels of severity in MDD, the interface demonstrated objective data and meaningful information when explaining the results to patients. Such features facilitate patients' understanding of the information conveyed by their healthcare providers.

Keywords: Major Depressive Disorder, Electroencephalography, Depression Severity, High Refractoriness, Suicide Risk, Machine Learning;

doi:10.6342/NTU202301910

CONTENTS

doi:10.6342/NTU202301910

doi:10.6342/NTU202301910

doi:10.6342/NTU202301910

LIST OF FIGURES

doi:10.6342/NTU202301910

LIST OF TABLES

doi:10.6342/NTU202301910

doi:10.6342/NTU202301910

doi:10.6342/NTU202301910

Chapter 1 Introduction

1.1 Major Depressive Disorder

Major depressive disorder (MDD) is a serious mental health condition that can significantly impact a person's quality of life. It is characterized by sadness or irritability and accompanied by psychophysiological changes, such as crying, disturbances in sleep, appetite, or sexual desire, slowing of speech and action, constipation, loss of the ability to experience pleasure, and even suicidal thoughts [1[2]. These changes may last at least 2 weeks. According to the statistics, an estimated 3.8% of the population, which is approximately 280 million people worldwide, have MDD [3]. It is shown that MDD is a serious disorder and strongly impacts the patient's quality of life. As a consequence, how to diagnose whether a patient has depression and provide effective treatment to the patient is highly imperative [4].

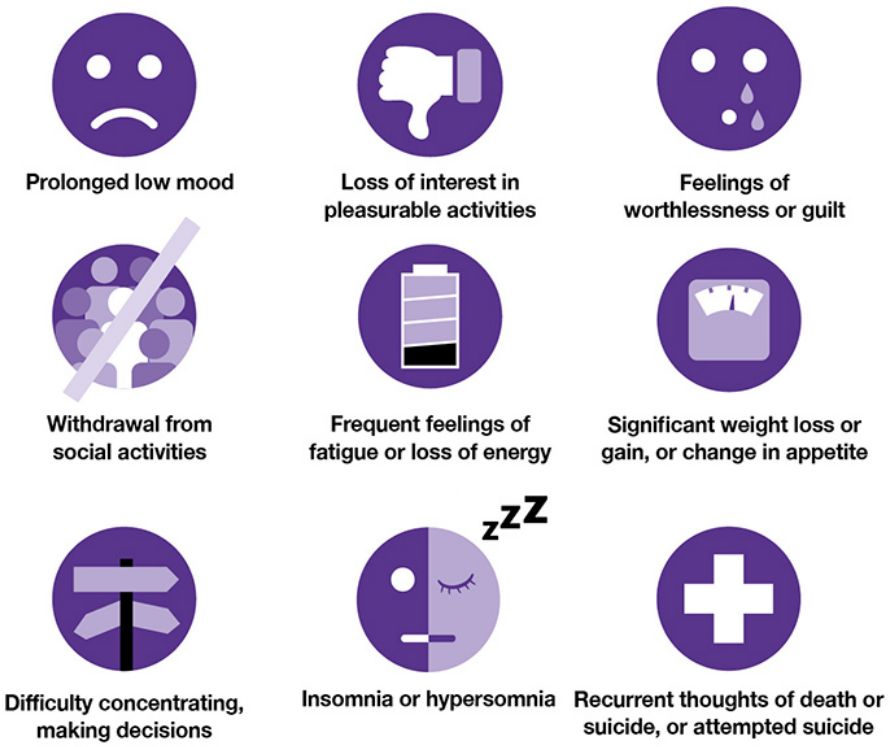


Figure 1-1. Symptoms of MDD [5]

doi:10.6342/NTU202301910

For clinical assessments, the Statistical Manual of Mental Disorders, fifth edition (DSM-5) [6] and the Hamilton Depression Rating Scale-17 (HAMD-17) [7] are standard depression scales used to diagnose depression. With the advancement of medical technology, electroencephalography (EEG), magnetic resonance imaging (MRI), and positron emission tomography (PET) have been found in recent studies try to find out reliable biomarkers to diagnose MDD [8[10]. Among them, EEG is a valuable tool to identify MDD due to its characteristic of low cost and easy recording.

1.2 Treatment Resistant Depression

Treatment Resistant Depression (TRD) is a subset of MDD which is characterized that depression symptoms persist even after adequate antidepressant therapy [11]. The European Union's Committee for Human Proprietary Medicinal Products (CHMP) [12] states: 'a patient is considered TRD when consecutive treatments with two different antidepressants products, used for a sufficient length of time at an adequate dose with a good affirmation of treatment adherence, fail to induce a clinically meaningful improvement'.

A meta-analysis by Fava and Davidson [13] shows that 29% to 46% of depressed patients treated with adequate doses and duration failed to respond fully. Specifically, 12% to 15% of patients attained only a partial response, whereas 19% to 34% of this population was nonresponsive. Another study [14] assessing treatment outcomes in MDD patients found that only 50.4% achieved full remission. The remaining patients experienced either a partial response or no response. According to the results above, it is obvious that quite large number percentage of clinical patients who do not remit after adequate antidepressant treatments. So, if there is a tool to determine whether a patient is

doi:10.6342/NTU202301910

TRD in advance, time spent on antidepressant treatment can be saved and more advanced treatment modalities, such as repetitive transcranial stimulation (rTMS) and intermittent theta burst stimulation (iTBS), can be used earlier to cure the patient.

Brain functional abnormalities play a significant role in the pathophysiology of MDD. Also, our team's previous research has already showed that depression is associated with many brain regions, such as the prefrontal cortex (PFC), anterior cingulate cortex (ACC), posterior cingulate cortex (PCC), and thalamus. Among them, PFC has been shown to have a strong correlation with refractoriness to antidepressant treatment [29].

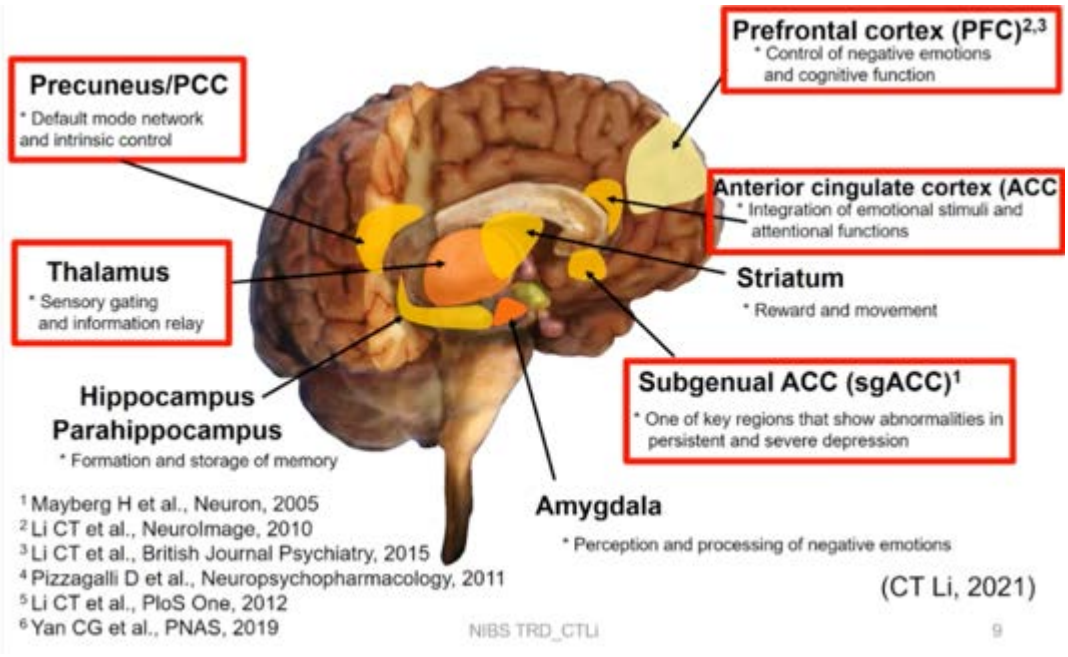


Figure 1-2. Brain regions associated with depression and treatment refractoriness. [24]

doi:10.6342/NTU202301910

1.3 Suicide Risk of Major Depressive Disorder

More than 700,000 people die due to suicide every year [15] and people with depression have a significantly higher suicide risk than non-depressed people [16]. There are various outcomes associated with suicide that are commonly referred as suicidal behaviors, including suicidal ideation (SI), suicide planning (SP), and suicide attempts (SA) [17]. The most serious symptom of MDD is SI and even SA. Unfortunately, patients with MDD are also at increased risk of engaging in suicidal behaviors. Research has consistently shown that a significant proportion of individuals with MDD experience SI, with estimates 37.7%.

Furthermore, approximately 15.1% of MDD patients have SP [18]. These statistics highlight the severity of SI and behavior in MDD patients and underscore the importance of early intervention and effective treatment to reduce the risk of suicide. It is critical to develop tools to predict suicide risk for identifying individuals who may be at high risk for suicidal behavior and providing timely intervention to prevent suicide.

From figure 1-3, listed brain regions are related to suicidal behaviors. The intersection brain region with both the target region of rTMS, intravenous injection of low-dose ketamine, and suicide behaviors is the dorsal prefrontal cortex (DPFC), colored in dark pink, which is responsible for executive control and is a key region within the executive network (EN). It has been shown to highly associated with depression and play a role in controlling suicidal ideations to behaviors.

doi:10.6342/NTU202301910

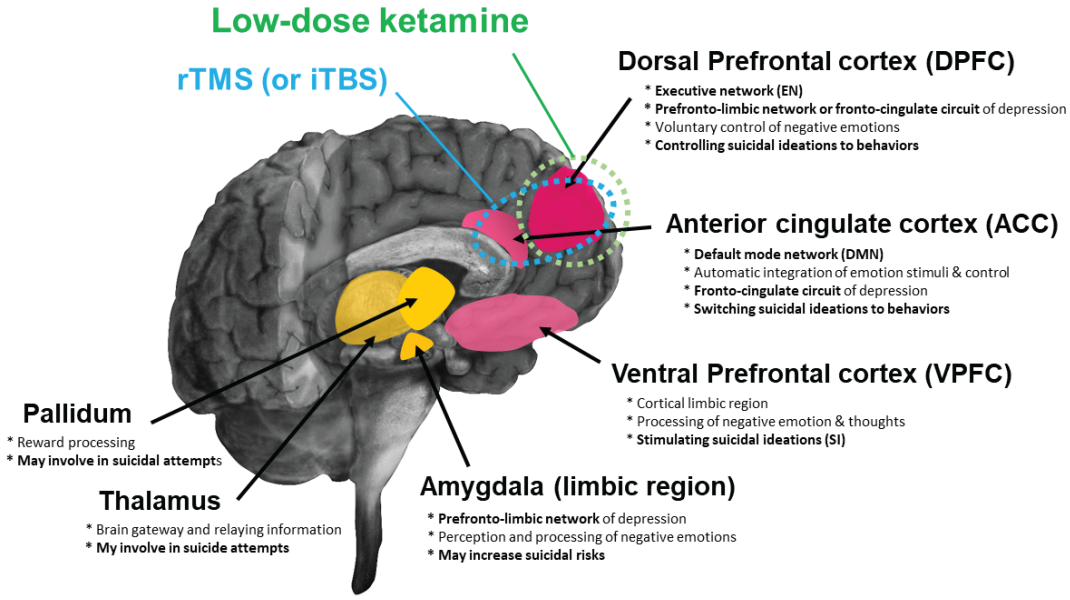


Figure 1-3. Key brain regions and circuits implicated in suicidal ideations and behaviors in patients with depression.[73]

1.4 Electroencephalogram

Electroencephalography (EEG) is a non-invasive neurophysiological technique used to record the brain's electrical activity. It measures the electrical potentials generated by large groups of neurons in the brain and provides insights into neural processing in real-time. EEG has been widely used in various fields of neuroscience research, including the study of depression and suicidal behavior [19]. By examining the patterns of electrical activity in the brain, EEG can offer valuable information on brain activity associated with mood regulation, decision-making, and cognitive processes, which are all relevant to understanding the pathophysiology of depression and suicide.

EEG can be divided into five frequency bands based on its physiological phenomenon: delta (1-4 Hz), theta (4-8 Hz), alpha (8-15 Hz), beta (15-30 Hz), and gamma (30-60 Hz).

doi:10.6342/NTU202301910

(1) Delta band:

Delta waves are often observed during deep sleep or unconsciousness and primarily observed in the frontal and central regions of the brain. The delta band is often associated with cognitive processes related to attention, memory consolidation, and emotional processing.

(2) Theta band:

Theta band is typically associated with the brain's "idling" state, such as during light sleep or drowsiness. It also involves working memory, attention, and other cognitive processes. The theta wave is often observed in the hippocampus and is essential for spatial navigation and memory consolidation.

(3) Alpha band:

Alpha band is often recorded in the occipital and parietal regions of the brain. It is observed when the brain is in a resting state or during closed-eye relaxation. Fast alpha waves can be seen when people are highly alert, whereas slow alpha waves are typically found when people are entering an unconscious state.

(4) Beta band:

Beta band is often related to alert and active mental states such as concentration, problem solving, and decision making. Beta band is involved in motor control and can be observed during muscle movement or motor planning. High levels of beta activity have been associated with anxiety and stress, while low levels have been linked to cognitive impairment and disorders such as Parkinson's disease.

(5) Gamma band:

Gamma band is associated with higher-order cognitive processes, such as attention, memory, and consciousness. Gamma activity has been observed in various brain regions, including the frontal, parietal, and temporal lobes. It is thought to reflect cortical network

doi:10.6342/NTU202301910

synchronization and is often seen during cognitive tasks requiring mental effort and concentration.

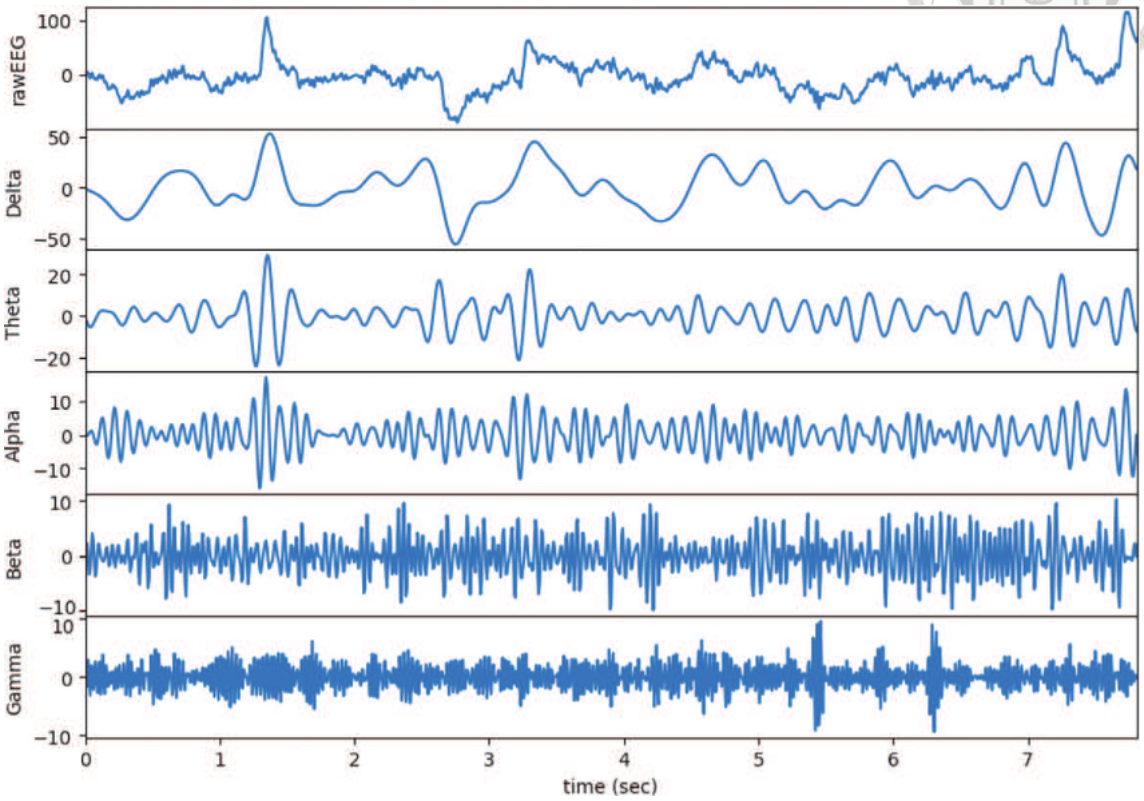


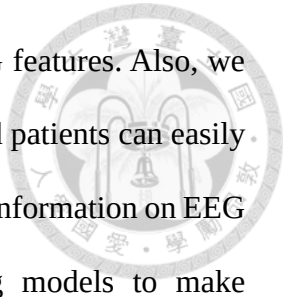
Figure 1-4. Illustration of different frequency bands in EEG signal [20].

1.5 Thesis Motivation

Previous studies in our team had used resting-state functional MRI (rs-fMRI) and positron emission tomography (PET) as raw data to predict whether the patient classifies as healthy control (HC), non-TRD or TRD and explored some potential biomarkers or features among three groups [21][22]. In addition, there were studies using EEG features to predict the response of repetitive transcranial stimulation (rTMS) and intermittent theta burst stimulation (iTBS) before the treatment [23].

In this study, we would like to use several machine learning approaches, such as Random Forest, XGBoost, and CatBoost, to make robust models to predict a patient's

doi:10.6342/NTU202301910



severity, high refractoriness, and suicide risk of MDD by using EEG features. Also, we use Python to create a user-friendly interface that doctors, nurses, and patients can easily use. This interface shows the prediction results and provides detailed information on EEG processing and important features by different machine learning models to make predictions. We hope this interface can enhance the transparency and interpretability of our work and promote effective communication between doctors and patients.

1.6 Thesis Organization

This thesis is structured into several chapters. Chapter 1 briefly introduces MDD, TRD, suicidal behaviors within depression symptoms, and EEG. Chapter 2 provides background information and previous research relevant to the topic. Chapter 3 describes the data acquisition process and clinical design used in the study. Chapter 4 outlines the methodology used to analyze the EEG signals and machine learning approaches. Chapter 5 presents the experimental results obtained from the research and the prediction interface designed for clinical use. Chapter 6 discusses the results, while Chapter 7 summarizes the main findings and draws conclusions. Finally, Chapter 8 outlines directions for future research.

doi:10.6342/NTU202301910

Chapter 2 Previous Research

2.1 DLPFC dysfunction in MDD

The pathological mechanisms of MDD involve multiple brain regions, including the prefrontal cortex (PFC), precuneus (PCC), anterior cingulate cortex (ACC), hippocampus, thalamus, and amygdala. Our team's research highlighted the importance of prefrontal cortex-anterior cingulate cortex-amygdala (PFC-ACC-Amygdala) in MDD patients [24]. Also, dorsolateral prefrontal cortex (DLPFC) plays the most essential role in executive dysfunctions in MDD patients. At the same time, DLPFC links to a range of cognitive and emotional processes, such as working memory, attention, decision-making, and emotion regulation [25].

In healthy individuals, DLPFC tends to regulate negative emotions [26]. It means that when individuals are experiencing negative emotions, DLPFC is often activated and helps to reduce the intensity of these emotions. However, studies have found that patients with MDD may show decreased DLPFC activity and interior lateral prefrontal cortex (PFC) during emotional processing [27[28]. This suggests that individuals with MDD may have difficulty regulating their negative emotions, contributing to depressive symptoms. Low functionality of DLPFC may also be associated with treatment refractoriness of depression [29]. Furthermore, DLPFC is evidenced to involve in mediating suicidal ideations to behaviors [74]. Therefore, we intend to use frontal lobe (DLPFC) EEG signals as data to predict the severity, treatment refractoriness, and suicide risk of individuals with MDD.

doi:10.6342/NTU202301910

2.2 Rostral Anterior Cingulate Cortex (rACC)-Engaging Cognitive Task (RECT)

However, the frontal theta wave can only partly reflect rostral anterior cingulate cortex (rACC) activity. A previous study of our team found that pre-treatment of modulated neural activity by applying a linear approach can induce rACC activity more than non-modulated one [30]. The modulating method is computerized RECT. The computerized RECT is a test designed according to the flexibility task of the Test for Attentional Performance [31]. During the RECT test, the computer screen will simultaneously and randomly show a range of sharp and circular patterns on the left and right sides. Patients are asked to click on the corresponding graphic immediately with their index finger with the rules explained before. Sharp and circular graphics are chosen to present because discriminating between sharp and circular patterns has been established to increase the prefrontal theta power [32]. Also, the increase of prefrontal theta power can increase the predictability of antidepressant response (Figure 2-2, [30]). Patients will do the RECT test for 10 minutes. After 5 minutes, the test will put more pressure to the patients (Figure 2-1, [32]). Resting-state five-minute EEG will be collected before and after performing RECT (i.e., baseline RECT and post-RECT).

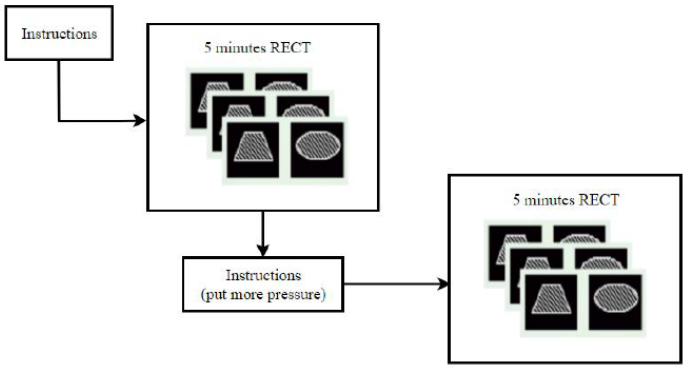


Figure 2-1. The procedure of RECT [32].

doi:10.6342/NTU202301910

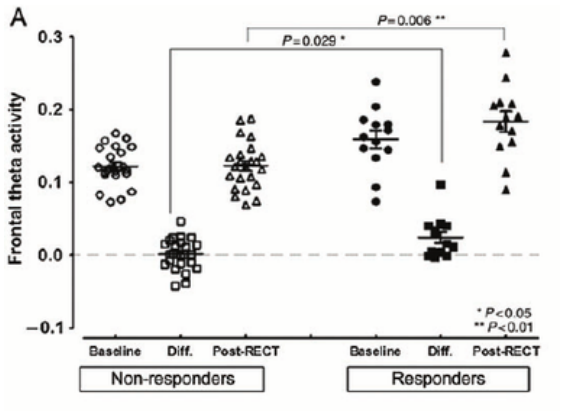


Figure 2-2. RECT-modulated frontal theta more reflects rACC activity than non-modulated one [30].

2.3 Machine Learning on MDD EEG Data

There are several types of research on using machine learning approaches for EEG [33-43]. However, the classification models they used are support vector machine (SVM) [33-41], logistic regression (LR) [33][40][42][43], and k-nearest neighbor (KNN) [35][37-42], which are relatively simple methods. Recently, few types of research have been using bagging or boosting methods to classify major depressive disorder and healthy control. Y. Mohammadi et al. used neural networks proposed by itself with manually extracted features and Beck depression inventory (BDI-II) scores to discriminate the levels of depression [44]. Many research tried to assess treatment resistant disorder, but not by using EEG as data or features [45]. D. K. Prasad et al. used selected electrodes related to attention, emotion, and memory to detect suicide ideation among all individuals' data they acquired [39]. There is no research on using machine learning methods with EEG features to make a comprehensive prediction including severity, treatment resistibility, and suicide risk. Therefore, this thesis employs a range of bagging and boosting machine learning techniques to train robust models capable of predicting various

doi:10.6342/NTU202301910

aspects of major depressive disorder. The predictions include the severity of major depressive disorder, the possibility of failed antidepressant treatments, and the risk of suicide.

2.4 Aim and Hypothesis

In this thesis, we aim to use the bagging and boosting machine learning models to train prediction models for severity, treatment resistibility, and suicide risk of MDD patients by using EEG features extracted from frontal signals. We hypothesize that the bagging and boosting methods, which respectively based on fusing decision trees and residual training, can perform superior to SVM, improve the accuracy and reliability of our predictions, and contribute to the development of more effective clinical tools for MDD diagnosis and treatment.

doi:10.6342/NTU202301910

Chapter 3 Data Acquisition

3.1 Psychiatric Evaluations

Psychiatric evaluations are performed to assess the severity of mood and somatic symptoms in patients. The evaluation tools used includes the 17-item Hamilton Depression Rating Scale (HAMD-17), Young Mania Rating Scale (YMRS), Clinical Global Index-Severity (CGI-S), Maudsley staging method (MSM), Depression and Somatic Symptoms Scale (DSSS), and Life Stress Scale. These assessments were conducted before initiating any treatment (W0, baseline). The clinician evaluated HAMD-17, YMRS, CGI-S, and MSM, while the patients self-filled the DSSS and Life stress scale. After 2 weeks of treatment (W2), all patients were re-evaluated using HAMD-17, YMRS, CGI-S, and DSSS. HAMD-17 is the most widely used scale for assessing depression, which measures various symptoms, such as depressed mood, feelings of guilt, suicide, insomnia, and anxiety. The severity of each symptom is scored, and the total score reflects the overall level of depression. A higher score indicates a more severe level of depression. In this study, we use total scores of HAMD-17 to determine the severity of depression and the scores of suicides (Question 3 of HAMD-17) to determine the suicide risk. Another scale, MSM, can evaluate duration, symptom severity, and treatment failures. In this study, we use the scores of antidepressant treatment failures to determine whether the patient is high refractoriness of MDD.

doi:10.6342/NTU202301910

3.2 Clinical Subjects

Professional physicians diagnosed all MDD patients. They aged from 13 to 88 years old. The inclusion and exclusion criteria of MDD patients are described below.

(1) Inclusion criteria:

A. Meets the criteria for major depressive disorder in DSM-5.
B. Current severity: CGI-S >= 4.
C. Currently antidepressant-free for at least one week.
D. With the ability to comprehend the contents of consent forms.

(2) Exclusion criteria:

A. Patients diagnosed with other types of mental disorder, including schizophrenia, psychotic disorder, bipolar disorder are excluded.
B. Patients diagnosed with neurological disorder, including stroke, seizure, neurocognitive diseases, and brain tumors are excluded.
C. Patients with brain surgical history or brain implants (neurostimulators) or cardiac pacemakers are excluded.
D. Pregnant patients are excluded.
E. Patients with inability to comprehend the contents of consent forms are excluded.
F. Patients who cannot cooperate (ex., fail to get informed consent) are excluded.

This study classified MDD patients into different groups based on various standards. Details are described below (Ch. 3.2.1-3.2.3).

doi:10.6342/NTU202301910

3.2.1 Classifying Severity

For classifying severity, 144 patients are classified into high-severity and low-severity groups based on HAMD-17 scores.

A. Low-severity patients (72 patients)

HAMD-17 scores: 16 or less

B. High-severity patients (72 patients)

HAMD-17 scores: 17 or more

Table 3-1. Demographic data of classifying severity. (Note. Student's t-test were applied to compare the continuous variables among different groups; chi-square tests was applied to compare the categorical variables among different groups.)

Groups / Variables	Low severity	High severity	P-value
Sex, male/female	33/39	31/41	0.73947
Age	40.11 (17.0)	39.04 (19.9)	0.72937
HAMD-17	16.01 (7.2)	22.97 (5.9)	2.80E-09
MSM	7.78 (2.4)	8.67 (1.8)	0.01287

doi:10.6342/NTU202301910

3.2.2 Classifying High Refractoriness

For classifying high refractoriness, 150 patients are classified into Non-TRD and TRD groups based on the scores of antidepressant treatment failures.

A. Non-TRD patients (75 patients)

MSM (Antidepressant treatment failures) score: 0-1

B. TRD patients (75 patients)

MSM (Antidepressant treatment failures) score: 2-5

Table 3-2. Demographic data of classifying high refractoriness. (Note. Student's t-test was applied to compare the continuous variables among different groups; chi-square tests were applied to compare the categorical variables among different groups.)

Variables \ Groups	Low refractoriness	High refractoriness	P-value
Sex, male/female	32/43	31/44	0.86969
Age	42.53 (19.6)	40.52 (17.2)	0.50488
HAMD-17	19.63 (7.4)	21.15 (7.1)	0.19963
MSM	6.64 (1.5)	8.88 (1.6)	1.13E-14

doi:10.6342/NTU202301910

3.2.3 Classifying Suicide Risk

For classifying suicide risk, only patients from 18 to 65 years old are included and 209 patients are classified into five groups based on the suicidality scores.

A. None suicide risk patients (37 patients)

   HAMD-17 (Suicidality) score: 0

B. Mild suicide risk patients (48 patients)

   HAMD-17 (Suicidality) score: 1

C. Moderate suicide risk patients (90 patients)

   HAMD-17 (Suicidality) score: 2

D. Severe suicide risk patients (28 patients)

   HAMD-17 (Suicidality) score: 3

E. Profound suicide risk patients (6 patients)

   HAMD-17 (Suicidality) score: 4

Table 3-3. Demographic data of classifying suicide risk.

Groups / Variables	None (37)	Mild (48)	Moderate (90)	Severe (28)	Profound (6)
Sex, male/female	14/23	22/26	32/58	13/15	4/2
Age	43.73 (13.9)	40.54 (14.7)	35.40 (14.1)	35.25 (14.0)	38.33 (15.2)
HAMD-17	14.00 (6.8)	19.38 (5.2)	22.97 (5.9)	26.75 (6.2)	27.83 (3.6)
MSM	6.70 (1.8)	8.15 (1.8)	8.76 (1.9)	9.39 (2.1)	10.67 (1.9)

From Table 3.1, Table 3.2, and Table 3.3, we can see that all ages and genders are not significant different between or among all groups (**P>0.05**), respectively, so they will not affect the subsequent analysis.

doi:10.6342/NTU202301910

3.3 EEG Data Acquisition

MDD patients' EEG data were acquired in a dimly lit, electrically shielded quiet space. A standard 32-channel digital EEG cap (Quik-Cap) with Ag/AgCl electrodes was positioned on the scalp following the international 10/20 system. The 10/20 system is a widely used method for electrode placement in EEG recordings, ensuring consistent and standardized electrode positions across different subjects and research settings (Figure 3-1, [46]), and the impedances of all electrodes were kept below 5 kΩ. The "10" and "20" refer to the fact that the actual distances between adjacent electrodes are either 10% or 20% of the total front-back or right-left distance of the skull [47]. Neuroscan amplifiers (Nuamps) and Neuroscan 4.3 software were used for EEG recording (each EEG signal for 5 minutes) while patients were seated in a comfortable arm-chair with eyes closed [48].

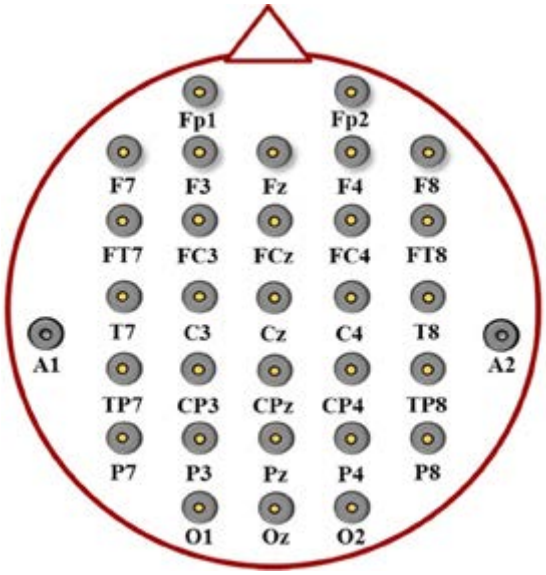


Figure 3-1. International 10/20 system of 32 channels placement [46].

doi:10.6342/NTU202301910

Chapter 4 Methodology

In this thesis, we aim to use EEG features to predict the severity, high refractoriness, and suicide risk of MDD patients. First, we need to do some EEG preprocessing, including EEG signal resampling, band pass filtering, and Independent Component Analysis (ICA) (detailed in Ch. 4.1). After EEG preprocessing, EEG signals were decomposed into 5 bands: alpha, bets, delta, gamma, and theta band. Then, EEG features were extracted by linear and nonlinear methods from EEG signals (detailed in Ch. 4.2). Techniques for dealing with imbalanced data were described in Ch.4.3. Finally, machine learning approaches were applied for classifying the severity, high refractoriness, and suicide risk (detailed in Ch. 4.4).

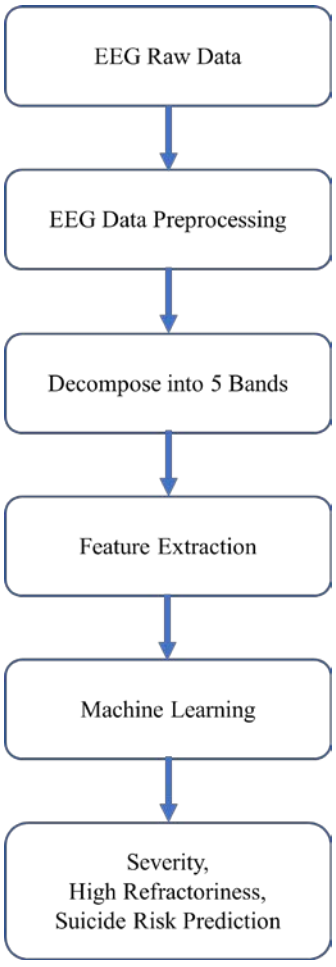


Figure 4-1. The overview of the methodology of this thesis.

doi:10.6342/NTU202301910

4.1 EEG Data Preprocessing

Because the EEG raw data from the hospital contains lots of biological artifacts such as eye blink, eye movement, heart beats, muscle movement, and so on, which may interfere the analysis, we need to first eliminate these noises. This section will introduce details of some preprocessing steps of EEG signals, including resampling, filtering, and Independent Component Analysis (ICA) will be introduced.

4.1.1 EEG Signal Resampling

The original raw data is about 300 seconds and its sampling frequency is 1,000 Hz. However, according to the Nyquist sampling theorem [49], the maximum frequency in this work is about 60 Hz. Therefore, the sampling frequency must be greater than 120 Hz in order to prevent aliasing effect. Due to that, we resample the EEG data from 1,000 Hz to 250 Hz.

4.1.2 Band Pass Filter

The EEG data is processed through a band-pass finite-duration impulse response (FIR) filter with the application of a Hamming window. The Hamming window helps achieve a smooth and accurate filtering process by reducing unwanted frequency components outside the desired range. The low cut-off frequency is set to 1 Hz, and high cut-off frequency is set to 60 Hz because our interesting region of EEG is from delta band (1-4 Hz) to gamma band (30-60 Hz) and we would like to filter out the direct current (DC) noise.

doi:10.6342/NTU202301910

4.1.3 Independent Component Analysis

Independent Component Analysis (ICA) is a technique used to discover a linear representation of data that is not normally distributed. The goal of ICA is to identify components that are as statistically independent from each other as possible. By doing so, ICA can separate mixed signals into their original sources, which is particularly useful in scenarios where the data consists of a mixture of different sources or signals. The resulting independent components provide valuable insights into the underlying structures and patterns within the data [50]. With the linear representation of EEG signals, we can easily remove artifacts such as eye blink, eye movement, etc. The algorithm of ICA is described below [50].

Assume that we observe n linear mixtures $x_1, x_2, \dots, x_n$ of n independent components like

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \cdots + a_{jn}s_n, \forall j$$

We consider that each mixture x_j and independent component s_k is a random variable. We assume that both the mixtures variables and the independent components have a mean of zero for convenience. We represent all the mixtures as a random vector x with elements $x_1, x_2, \dots, x_n$. Similarly, we represent all the independent components as another random vector s with elements $s_1, s_2, \dots, s_n$. We use a matrix A with elements a_{ij} to describe the mixing process. This matrix represents the linear combination of the independent components to form the observed mixtures. The mixing model, in vector-matrix notation, can be written as

$$x = As$$

doi:10.6342/NTU202301910

Denoting the columns of matrix A and the equation can also be written as

$$x = \sum_{i=1}^{n} a_i s_i$$

The equation described above represents the Independent Component Analysis (ICA) model, which is a generative model. It explains how the observed data x_i is generated by mixing the independent components s_i through the unknown mixing matrix A. Since we only have access to the observed random vector x, our task in ICA is to estimate both the unknown mixing matrix A and the independent components s_i from this observed data. To achieve this, ICA relies on certain assumptions and statistical methods to separate the mixed signals and uncover the original independent components. Once we estimate the mixing matrix A, we can compute its inverse matrix A^{-1}. By multiplying the observed data x with A^{-1}, we can obtain the estimated independent components $s_1, s_2, \dots, s_n$ by

$$s = A^{-1}x$$

Then, we can remove the independent components, artifacts more specifically and be able to use EEG signals without noise for further analysis.

doi:10.6342/NTU202301910

4.2 Feature Extraction

For feature extraction, we integrate both the linear and nonlinear feature extraction methods and all the linear and nonlinear features extracted in our previous work [23][74](Table 4.1). The nonlinear methods can be classified into three categories: trend (Largest Lyapunov Exponent (LLE) [51] and Detrended Fluctuation Analysis (DFA) [52]), complexity (Approximate Entropy (ApEn) [53]), and fractal dimension (Katz Fractal Dimension (KFD) [54] and Higuchi Fractal Dimension (HFD) [55]). The linear method is to compute the band power of EEG signals (Welch periodogram [56]).

Table 4-1. Feature extraction for EEG signals.

Categories		**Methods**	**Features**
Nonlinear	Trend	Largest Lyapunov Exponent (LLE) [51]	Instability or unpredictability
		Detrended Fluctuation Analysis (DFA) [52]	Represent the long-range temporal correlation
	Complexity	Approximate Entropy (ApEn) [53]	Regularity and complexity
	Fractal Dimension	Katz Fractal Dimension (KFD) [54]	Obtains fractal dimension based on morphology, which measures the roughness of time series
		Higuchi Fractal Dimension (HFD) [55]	Quantifies the self-similarity of an objection in the time domain
Linear	Band Power	Welch periodogram (Welch) [56]	Compute band power by integration

doi:10.6342/NTU202301910

4.2.1 Largest Lyapunov Exponent

Lyapunov Exponents (LE), which are the average exponential rates of divergence or convergence of nearby orbits in phase space, provide a quantitative characterization of dynamic behavior. LE is related to the degree of divergence or convergence of nearby trajectories in phase space over time. It measures the separation rate of initially close trajectories in a chaotic system, where each LE corresponds to a specific direction in phase space. Positive LE indicates that trajectories diverge exponentially, meaning the system is chaotic and sensitive to initial conditions, while negative LE indicates that nearby trajectories converge over time, meaning the system is stable and predictable. In most applications, the Largest Lyapunov Exponent (LLE) [51] is commonly used because it determines predictability for a dynamical system. The operation steps of LLE are described below [51].

Assume $S = [x(1), x(2), \dots, x(N)]$ is the sequence of the data. Reconstructed phase space X is defined as

$$X(i) = [x(i), \dots, x(i + m - 1)]$$

where $i = 1,2, \dots, (N - m + 1)$ and m is the embedding dimension. The maximum Lyapunov exponent λ_1 is defined as

$$d_j(i) = d_j(0)e^{\lambda_1 i\Delta t}$$

where $d_j(i)$ is the average Euclidian distance between two nearby paths at t_i, and $d_j(0)$ is the Euclidian distance between $j\,th$ pair of nearest neighbors. We can have

$$\ln\left(d_j(i)\right) = \ln\left(d_j(0)\right) + \lambda_1 i\Delta t$$

Then we can calculate the LLE by fitting the slope of the curve of the best least square to the mean log curve, which is defined as

$$y(i) = \frac{1}{\Delta t}\left(\ln\left(d_j(t)\right)\right)$$

doi:10.6342/NTU202301910

4.2.2 Detrended Fluctuation Analysis

Detrended fluctuation analysis (DFA) [52] is mainly used to distinguish the influence of intrinsic and extrinsic stimuli on complex systems, as changes in the internal dynamics of the system can cause changes in long-range correlations, while external stimuli can cause local changes in the system. DFA is used to measure the patterns' similarity and evaluate trends of various scales. The operation steps of DFA are described below [52].

Assume $S = [x(1), x(2), \dots, x(N)]$ is the sequence of the data. N new sequences are constructed for $k = 1,2,3, \dots, N$,

$$y(k) = \sum_{i=1}^{k} [x(i) - \langle S \rangle]$$

where $\langle S \rangle$ is average value of S, which is defined as

$$\langle S \rangle = \frac{1}{N} \sum_{i=1}^{N} x(i)$$

y(k) is divided into equal box sizes n, each box size has a least square line, denoted by $y_n(k)$. Then we can compute the root mean square fluctuation $F(n)$ by

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [y(k) - y_n(k)]^2}$$

Then the scaling exponent α of DFA can be computed as:

$$\alpha = \frac{\log (F(n))}{\log (n)}$$

doi:10.6342/NTU202301910

4.2.3 Approximate Entropy

Approximate entropy (ApEn) [53] is a nonlinear statistic technique used to quantify the regularity and the unpredictability of fluctuations over time-series data. The operation steps of ApEn are described below [53].

Assume $S = [x(1), x(2), \dots, x(N)]$ is the sequence of the data, $x^*(i)$ is the subsequence of the original data S.

$$x^*(i) = [x(i), x(i+1), \dots, x(i-m+1)]$$

where m is the length of sampling window. The threshold $r = SD * k$, where SD means the standard deviation of S and k represents a constant between 0.1 and 0.9. For each $1 \le i, j \le N - m + 1, i \ne j, C_i^m$ can be expressed as

$$C_i^m(r) = \frac{\sum_{j=1}^{N-m+1} \theta(r - |x^*(i) - x^*(j)|)}{N - m + 1}$$

where θ is the Heaviside function. The quantity $\emptyset^m(r)$ can be calculated as

$$\emptyset^m(r) = \frac{\sum_{i=1}^{N-m+1} \ln C_i^m(r)}{N - m + 1}$$

Then ApEn is defined as

$$ApEn = \emptyset^m(r) - \emptyset^{m+1}(r)$$

doi:10.6342/NTU202301910

4.2.4 Fractal Dimension

Fractal dimension (FD) is a measure of complexity that counts the effective number of degrees of freedom in a dynamical system. The Katz FD (KFD) [54] and Higuchi FD (HFD) [55] are two different methods used to compute self-similarity in FD. These methods are used to evaluate trends and fractal tendencies of various scales. KFD measures the degree of curve tortuosity in a signal, while HFD evaluates the scale-invariant properties of the signal.

(1) Katz Fractal Dimension

The KFD obtains fractal dimension based on morphology, which measures the distance between two successive points, considered as a measure of the irregularity of a time series. The operation steps of KFD are described below [54].

Assume $S = [x(1), x(2), \dots, x(N)]$ is the sequence of the data. For $j = 1,2, \dots, N$. Find the maximum Euclidian distance between $x(1)$ and $x(j)$ and call it d. Compute L as the total length of the time series

$$L = \sum_{i=2}^{N} d(x(i), x(i-1))$$

Compute a as the average distance between continuous points of S

$$a = \frac{L}{N-1}$$

Then we can compute KFD as

$$KFD = \frac{ln\frac{L}{a}}{ln\frac{d}{a}}$$

doi:10.6342/NTU202301910

(2) Higuchi Fractal Dimension

The HFD is a metric used to quantify the complexity and self-similarity of a time series in the time domain. It is a valuable tool for obtaining stable indices and identifying the time scale corresponding to a given dataset's characteristic frequency. HFD measures how much the time series deviates from a straight line, with a higher HFD indicating greater complexity and less predictability. The operation steps of HFD are described below [55].

Assume $S = [x(1), x(2), \dots, x(N)]$ is the sequence of the data. New sequences of length k are constructed for $m = 1, 2, \dots, k$.

$$X_m^k = \left\{x(m), x(m+k), \dots, x\left(m + \left\lfloor\frac{N-m}{k}\right\rfloor k\right)\right\}$$

where k determines the delay in continuous points. The normalized average length L_m can be calculated as

$$L_m(k) = \frac{1}{k}\frac{\sum_{i=1}^{\left\lfloor\frac{N-m}{k}\right\rfloor}|x(m+ik) - x[m+(i-1)]k| \times (N-1)}{\left\lfloor\frac{N-m}{k}\right\rfloor}$$

Compute the total average length for scale k

$$L(k) = \frac{1}{k}\sum_{m=1}^{k} L_m(k)$$

Then we can calculate the HFD as the slope of the best least squares line for the curve $logL(k)$ versus $log(1/k)$.

doi:10.6342/NTU202301910

4.2.5 Welch Periodogram

Welch periodogram method is commonly used to estimate a signal's power spectral density (PSD). It converts the signal from the time domain to the frequency domain using periodogram spectrum estimates. This method is known for its ability to reduce noise, making it a popular choice in many applications. The frequency spectrum is given by the operation step below [56].

$$P_w^p(f) = \frac{1}{P}\sum_{p=0}^{D-1}\frac{1}{UDT}\left|T\sum_{n=0}^{D-1}w[n]x^p[n]e^{-j2\pi fnT}\right|^2$$

where U is the discrete-time window energy of $w[n]$

$$U = T\sum_{n=0}^{D-1}w^2[n]$$

doi:10.6342/NTU202301910

4.3 Imbalanced Data

Imbalanced data refers to a situation where the distribution of classes in a dataset is not equal. In other words, one class may have a much larger number of samples than the other classes, resulting in an imbalanced dataset. This may cause the classification model fails to give accurate predictions about the minority class. To address this problem, there are several approaches, such as resampling (over-sampling and under-sampling), adjusting class weights, using the right evaluation metrics, etc. Under-sampling of the majority class has been proposed as a suitable means of increasing the sensitivity of a classifier to the minority class [57]. In this work, we apply Synthetic Minority Oversampling Technique (SMOTE) [58], a combination of under-sampling and over-sampling techniques, to balance our dataset and obtain better classification results with our finite data.

4.3.1 Synthetic Minority Oversampling Technique

Synthetic Minority Oversampling Technique (SMOTE) [58] is a powerful tool for addressing imbalanced data in classification problems. One of the critical advantages of SMOTE is that it can effectively increase the size of the minority class by creating synthetic examples using existing minority class examples. This can significantly improve the performance of machine learning models trained on imbalanced datasets. SMOTE can also help reduce bias towards the majority class, often dominating the learning process in imbalanced datasets. Another advantage of SMOTE is that it can be used with a wide range of machine learning algorithms, making it a versatile tool for addressing imbalanced data in many different contexts. Overall, SMOTE is a valuable technique for improving the accuracy and reliability of classification models trained on

doi:10.6342/NTU202301910

imbalanced data. The main concept of SMOTE is to generate synthetic samples for the minority classes. The operation steps of SMOTE are described below [58].

Step 1. Choose a minority class instance as a starting point

Step 2. Find k nearest neighbors for the chosen instance

Step 3. Choose one of the k neighbors and create a synthetic new sample by

$$x_{new} = x_{chosen} + (x_{nearest} - x_{chosen}) * \delta$$

Where δ is a random number between 0 and 1. The new synthetic instance x_{new} is then added to the minority class data to increase its representation in the dataset

Step 4. Repeat steps 1-3 until the desired level of minority class over-sampling is achieved.

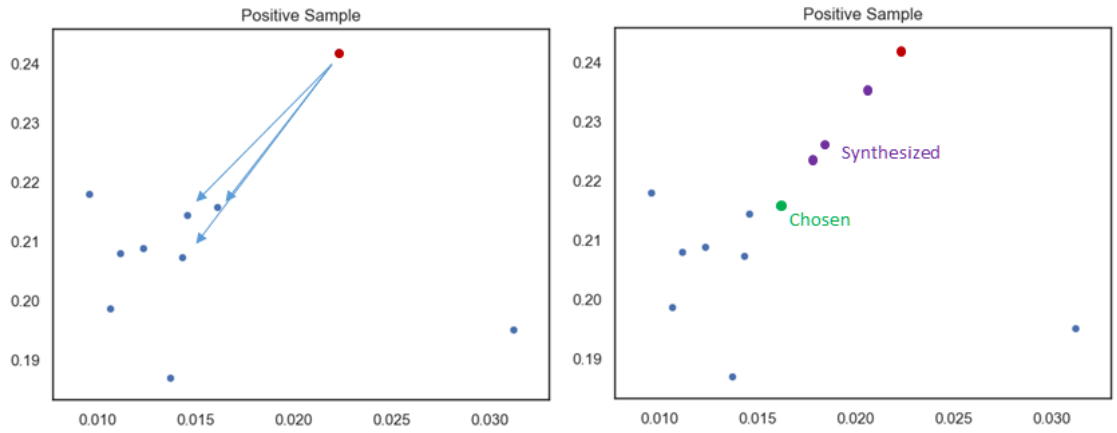


Figure 4-2. Illustration of the SMOTE over-sampling approach.

doi:10.6342/NTU202301910

4.4 Machine Learning Approaches

This section uses three machine learning approaches to train classifiers: Random Forest (RF), XGBoost, and CatBoost. We combine these three classifiers using an ensemble voting technique to obtain the final classification result. In addition, we include Support Vector Machine (SVM) as a reference model to further compare the classification performance.

4.4.1 Support Vector Machine

Support Vector Machine (SVM) [59] is a supervised machine learning algorithm used for binary classification tasks, both linear and non-linear. SVM works by finding the hyperplane that can separate data points of different classes with the most significant margin distance between them (Figure 4-3, [60]). This margin is optimal because it minimizes the chances of misclassifying future data points. The Gaussian SVM [61], which is used in the present study, uses a radial basis function kernel to transform the data into a higher-dimensional space, where the hyperplane can be found to separate the data. This approach has been shown to be effective in various classification tasks.

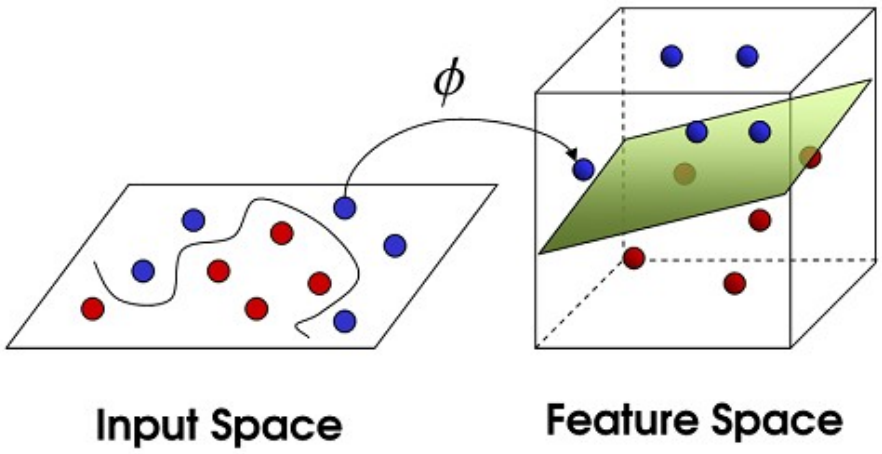


Figure 4-3. Illustration of SVM computation [60]

doi:10.6342/NTU202301910

4.4.2 Random Forest

Random Forest (RF) [62] is the advanced version of decision tree, which combines multiple decision trees, the reason for the name ‘forest’, and applies the ‘bagging’ (bootstrap aggregating) algorithm with random feature sampling in order to prevent overfitting and improve the prediction ability. RF randomly select n samples from total N samples, where $n < N$, and then also randomly extract k features as the training features to train a decision tree. Repeatedly do the abovementioned step M times to get M decision trees and integrate the prediction results by a majority vote (Figure 4-4, [63]). Due to samples and features used in each decision tree both randomly determined and applying voting mechanism for the final prediction results, RF can reduce risk of overfitting and provide flexibilities. Also, every decision tree is independent, so it can run in parallel during the training and predicting phase to speed up the process.

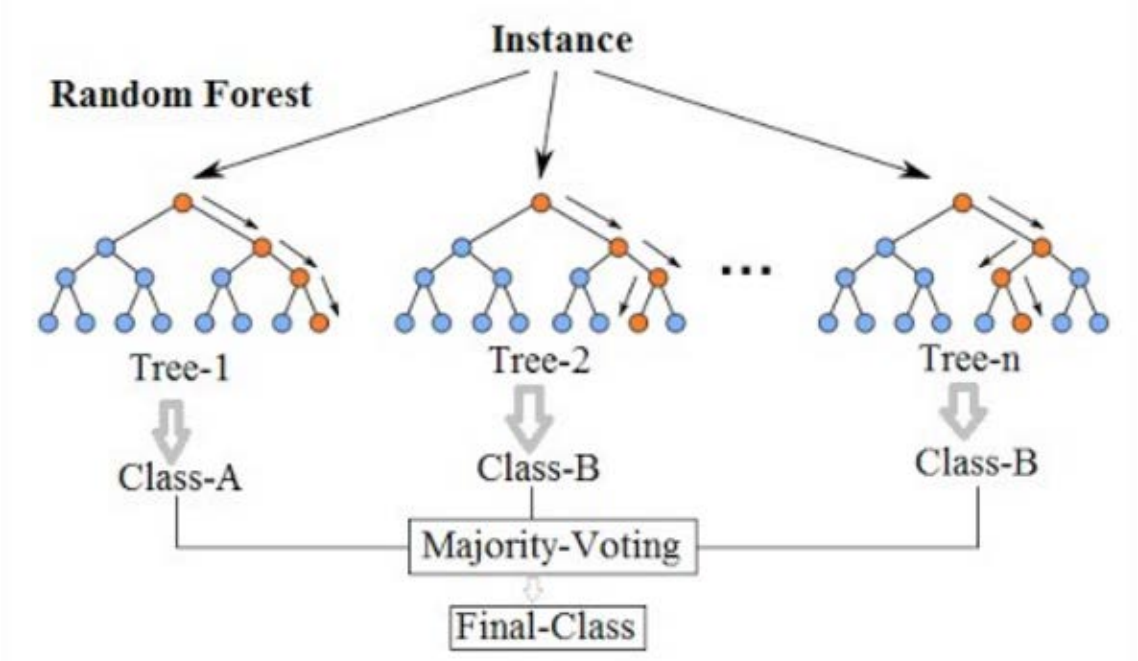


Figure 4-4. Illustration of Random Forest algorithm [63].

doi:10.6342/NTU202301910

4.4.3 XGBoost

XGBoost (eXtreme Gradient Boosting) [64], an algorithm based on gradient boosting decision tree (GBDT) with some new techniques, combines the advantages of bagging and boosting. Like bagging, boosting methods also build trees with random sampling, which means not all the features participate in making the tree. The essential boosting technique is that every tree is related to the former tree, with the hope that the latter tree can revise some mistakes made by the former (Figure 4-5, [65]). When every latter tree is a little bit better than the former one continuously for a while, we can expect that the tree will be better enough. The model may generate many high-order functions to fit the training data during training, but it is easy to become overfitted by a little noise interference. Therefore, XGBoost adds normalization to the objective function to prevent the model from becoming over-complicated and uses L1/L2 Regularization to smooth the loss function and become more resistant to noise interference.

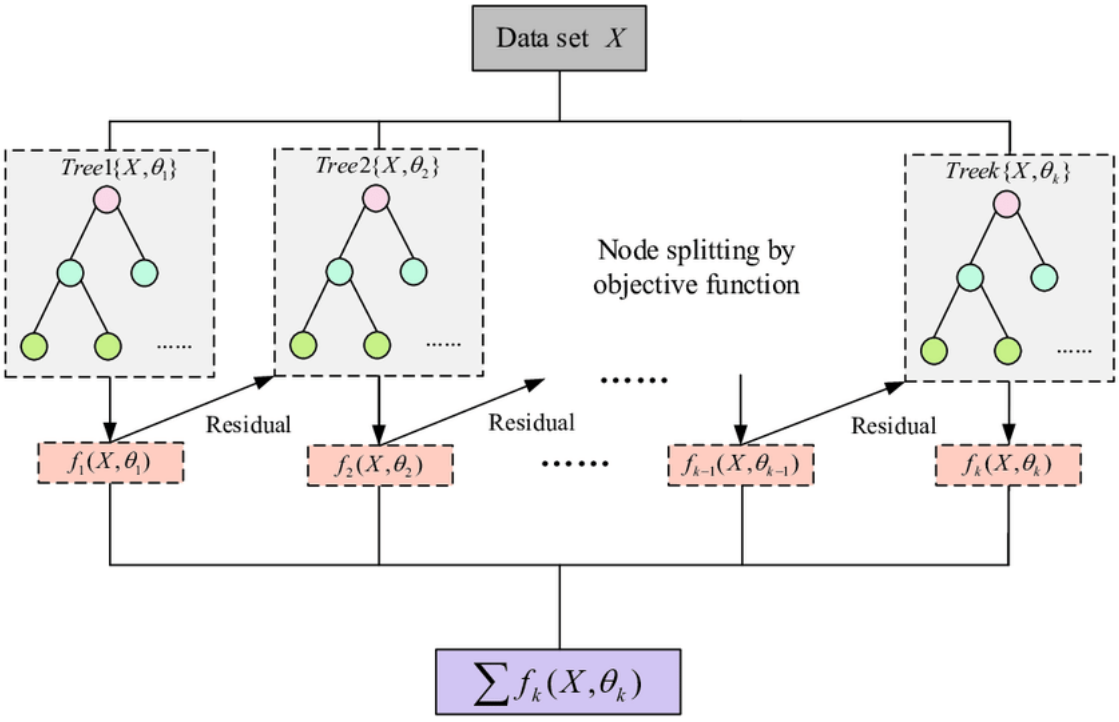


Figure 4-5. Illustration of XGBoost algorithm [65].

doi:10.6342/NTU202301910

4.4.4 CatBoost

CatBoost [66], derived from ‘category’ and ‘boosting’, is also an algorithm based on GBDT. Different from XGBoost, there are some innovations and highlights in CatBoost. First, CatBoost propose a new algorithm that can automatically transfer categorical features into numerical features, not just treat them as numerical features. It will use statistic methods to calculate the frequency of the categorical feature and add hyperparameters to become a new numerical feature, preserving the categories’ information and enriching the feature dimension. Second, CatBoost come up with ‘order boosting’ to prevent the prediction shift, caused by using the same datasets to get the model’s gradient during every iteration in GDBT. Order boosting will train a separate model M for each sample x and the model M is trained by the dataset except the x. Then, we can get the gradient estimation with the model M, use the gradient to train base classifier, and get the final model. Last, CatBoost use oblivious trees, balanced and symmetric decision trees, as the base classifiers. The same split criterion is used on the entire tree level, which can prevent the possibility of overfitting problem. Due to the techniques and advantages above, CatBoost is known for its robustness and excellent accuracy compared with any advanced machine learning algorithms.

doi:10.6342/NTU202301910

Chapter 5 Experimental Results

5.1 Evaluation Method

In this section, we will introduce how we evaluate the performance of our models or results by using different functions or calculations.

A. Confusion Matrix

Confusion matrix is a visualization tool specifically used in supervised learning. Each row of the matrix represents the instances of a class predicted, while each column represents the instances of an actual class. The reason for this naming is that through this matrix, it is easy to see whether the model has confused two different classes (such as misclassifying one class as another). Figure 5-1 shows the four different combinations of predicted and actual values, named as True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN).

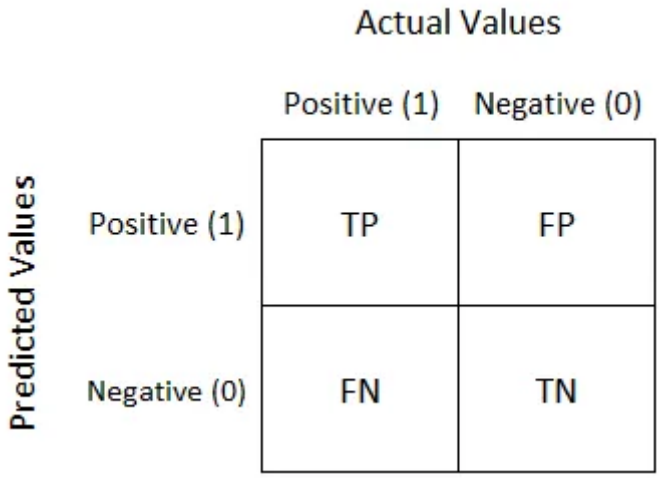


Figure 5-1. Illustration of Confusion Matrix [67]

doi:10.6342/NTU202301910

B. Area Under Curve

Area under curve (AUC) is the area under the receiver operating characteristic (ROC) curve, which plots the true positive rate (TPR) against the false positive rate (FPR). The AUC ranges from 0 to 1, with higher score means the better discrimination ability between the positive and negative classes. A perfect classifier has an AUC of 1. An AUC of 0.5 indicates that the model has no discrimination ability and an AUC of less than 0.5 indicates that the model's predictions are worse than random guessing.

C. Precision, Recall, Specificity, F1-Score, Accuracy

Precision measures the proportion of true positive predictions among all positive predictions made by the model. It tells how often the model is correct when it predicts a positive label.

$$Precision = \frac{TP}{TP + FP}$$

Recall (also known as sensitivity) measures the proportion of true positive predictions among all actual positive instances. It tells us how well the model can identify positive instances.

$$Recall = \frac{TP}{TP + FN}$$

Specificity is the opposite of recall, which measures the proportion of true negative predictions among all actual negative instances. It tells us how well the model can identify negative instances.

$$Specificity = \frac{TN}{TN + FP}$$

F1-score is the harmonic mean of precision and recall. It is a balanced metric that considers both precision and recall. It is particularly useful when the classes are

doi:10.6342/NTU202301910

imbalanced. When the weights of the classes are equal, the calculation formula is described below

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

Accuracy measures the proportion of correct predictions made by the model among all predictions. It tells us how often the model is correct in its predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

D. P-value (ACC>NIR)

No information rate (NIR) is the accuracy achieved by a model that always predicts the most frequent class in the training data without using any input variables. In the one-tailed binomial test, a p-value less than certain significance level (e.g., 0.05) would indicate that the ACC is significantly better than NIR. Thus, the model has some predictive power. On the other hand, if the p-value is greater than the significance level, it suggests that the difference between ACC and NIR could be due to chance, and the model may not be useful for prediction.

doi:10.6342/NTU202301910

5.2 Classification Machine Learning Performance

In this thesis, we utilize SVM, a widely used classification model, for time series prediction and classification tasks as a comparative model. We have extracted five nonlinear features, including LLE, DFA, KFD, HFD, and ApEn, which capture complex patterns in the EEG signals, and a linear feature Welch that represents the linear relationship between variables. By combining these features, we aim to improve the predictive accuracy compared with SVM by using a bagging and boosting model including Random Forest, XGBoost, and CatBoost for prediction.

This section shows the machine learning results of each classification, including the model's Accuracy, AUC, Precision, Recall/Sensitivity, Specificity, F1-score, and P-value (ACC>NIR), as shown from Table 5.1 to Table 5.4.

doi:10.6342/NTU202301910

5.2.1 Model Performance Result for Classifying Severity

Table 5.1 shows the machine learning methods used for classifying severity and compares the evaluating results with each other. From Table 5.1, we can see that nearly all the evaluation functions of the Random Forest model perform the best. Also, using the bagging and boosting methods to train classifiers performs better than SVM. Furthermore, the P-value (ACC>NIR) of the best SVM classifier is above 0.05, which indicates that it performs worse than random guessing.

Table 5-1. The machine learning results for classifying severity.

	Classifying Severity			
Number	72/72			
Feature	All linear and nonlinear with all bands			
Classification	SVM	Random Forest	XGBoost	CatBoost
Accuracy	0.65517	**0.82759**	0.72414	0.75862
AUC	0.60476	**0.82857**	0.75	0.77404
Precision	0.625	0.8	**1.0**	0.66667
Recall/Sensitivity	0.71429	0.85714	0.5	**0.92308**
Specificity	0.6	0.8	**1.0**	0.625
F1-score	0.6667	**0.82759**	0.66667	0.77419
P-value (ACC>NIR)	0.60675	0.00040	0.00274	0.00249

doi:10.6342/NTU202301910

5.2.2 Model Performance Result for Classifying High Refractoriness

Table 5.2 shows the machine learning methods used for classifying high refractoriness and compares the evaluating results with each other. From Table 5.2, we can see that nearly all the evaluation functions of the Random Forest model perform the best. Also, using the bagging and boosting methods to train classifiers performs slightly better than SVM. However, the best SVM classifier's P-value (ACC>NIR) is above 0.05, while the others are below 0.05.

Table 5-2. The machine learning results for classifying high refractoriness.

	Classifying High Refractoriness			
Number	75/75			
Feature	All linear and nonlinear with all bands			
Classification	SVM	Random Forest	XGBoost	CatBoost
Accuracy	0.73333	**0.86667**	0.73333	0.8
AUC	0.81333	**0.86667**	0.74107	0.81944
Precision	0.68421	0.92308	0.83333	**0.92857**
Recall/Sensitivity	**0.86667**	0.8	0.625	0.72222
Specificity	0.6	**0.93333**	0.85714	0.91667
F1-score	0.76471	**0.85714**	0.71429	0.81250
P-value (ACC>NIR)	0.30544	5.06E-05	0.00716	0.00059

doi:10.6342/NTU202301910

5.2.3 Model Performance Result for Classifying Suicide Risk

Table 5.3 and Table 5.4 show the machine learning methods used for classifying suicide risk with and without SMOTE and compare the evaluating results with each other. Table 5.3 shows that all classifiers perform poorly due to multi-class classification with the imbalanced dataset. Only the CatBoost model's P-value (ACC>NIR) is below 0.05; however, the accuracy is about 30%. To address the issue of imbalanced data, we employ the SMOTE technique to balance the quantities among each label, and results are shown in Table 5.4. To mitigate the impact of synthetic data generated by SMOTE on the model's accuracy, we first split the data into training and testing datasets using the same proportion as each label in the entire dataset. Subsequently, we only apply SMOTE to the training data. By doing so, we ensure that the synthetic data generated by SMOTE was used exclusively for training the model, while the testing data remained untouched. This approach helps us maintain the integrity and generalizability of the model's performance during evaluation. From Table 5.4, we can observe that all the evaluation functions of the models are better than those in Table 5.3. Random Forest outperforms the other models significantly with only one testing data predicting wrong. The results of the different models are better than random guessing, but the models' accuracy still needs improvement to be more convincing.

doi:10.6342/NTU202301910

Table 5-3. The machine learning results based on the original imbalanced dataset for classifying suicide risk (without SMOTE).

	Classifying Suicide Risk			
Number (0/1/2/3/4)	37 / 48 / 90 / 28 / 6			
Feature	All linear and nonlinear with all bands			
Classification	SVM	Random Forest	XGBoost	CatBoost
Accuracy	0.19048	0.28571	0.26190	0.30952
AUC	0.57782	0.48045	0.47267	0.47149
Precision	0.18447	0.17898	0.16758	0.17956
Recall/Sensitivity	0.19048	0.28571	0.26190	0.30952
Specificity	0.8	0.82353	0.78571	0.83333
F1-score	0.17561	0.21693	0.20413	0.20413
P-value (ACC>NIR)	0.46908	0.06211	0.11817	0.02976

Table 5-4. The machine learning results based on the balanced dataset after applying SMOTE technique for classifying suicide risk (with SMOTE).

	Classifying Suicide Risk			
Number (0/1/2/3/4)	72 / 72 / 72 / 72 / 72			
Feature	All linear and nonlinear with all bands			
Classification	SVM	Random Forest	XGBoost	CatBoost
Accuracy	0.47619	**0.97619**	0.59524	0.52381
AUC	0.53985	**0.99891**	0.63671	0.71901
Precision	0.45805	**0.97744**	0.62648	0.47792
Recall/Sensitivity	0.47619	**0.97619**	0.59524	0.52381
Specificity	0.75758	**0.96970**	0.82857	0.78788
F1-score	0.46172	**0.97589**	0.56574	0.49427
P-value (ACC>NIR)	1.36E-05	1.11E-16	3.68E-09	6.70E-07

doi:10.6342/NTU202301910

5.3 Features of Importance

To enhance the interpretability of the machine learning model, bagging and boosting algorithms provide feature importance as an output. This section will analyze the feature importance using these algorithms with all features included. Additionally, we will use one-way ANOVA to calculate the p-value to determine the significance of each feature across different groups.

5.3.1 Feature Importance for Classifying Severity

Tables 5.3 to 5.5 are the feature importance list of the bagging and boosting machine learning methods for classifying severity. The larger the value of the feature importance in the table, the more influential the feature is in that decision tree-based machine learning model.

Table 5-5. The feature importance of the Random Forest model for classifying severity.

Random Forest Feature Importance for Classifying Severity			
Feature Importance	**Sub-band**	**Electrode-Feature**	**P-value**
0.0332	Alpha	FP2-DFA	0.0492*
0.0244	Delta	F7-HFD	0.7195
0.0242	Alpha	F7-Welch	0.8786
0.0231	Alpha	FP1-DFA	0.2371
0.0190	Alpha	F3-Welch	0.6970

doi:10.6342/NTU202301910

Table 5-6. The feature importance of the XGBoost model for classifying severity.

XGBoost Feature Importance for Classifying Severity			
Feature Importance	**Sub-band**	**Electrode-Feature**	**P-value**
0.1016	Gamma	F7-ApEn	0.1231
0.0733	Alpha	FP2-ApEn	0.0850
0.0615	Gamma	F4-HFD	0.0768
0.0584	Delta	F7-LLE	0.7151
0.0575	Theta	F8-Welch	0.0680

Table 5-7. The feature importance in the CatBoost model for classifying severity.

CatBoost Feature Importance for Classifying Severity			
Feature Importance	**Sub-band**	**Electrode-Feature**	**P-value**
0.3345	Delta	F8-LLE	0.3901
0.3300	Gamma	F7-HFD	0.4139
0.1716	Beta	FP2-KFD	0.6137
0.1639	Theta	FP1-KFD	0.7230

We analyze the feature importance by electrode channel-wise, feature-wise, and sub-band-wise to get detail findings. Table 5.8 to Table 5.10 shows that the best models among the bagging and boosting models prefer to use FP1, FP2, F7, and F8 channels, especially F7 channel. The number of linear and nonlinear features in the models is roughly equal. And then, the models tend to choose the elements from alpha, delta, theta, and gamma bands, especially alpha band, for making predictions. Among them, Random Forest model demonstrates the best performance, and based on the p-values, a unique feature, $p<0.05$, was identified in Random Forest model, specifically in alpha band of FP2 channel.

doi:10.6342/NTU202301910

Table 5-8. The statistics of the channel-wise biomarkers of bagging and boosting model for classifying severity.

	FP1	FP2	F3	F4	F7	F8	FZ
Number	2	3	1	1	5	2	0

Table 5-9. The statistics of the feature-wise biomarkers of bagging and boosting model for classifying severity.

	LLE	ApEn	DFA	KFD	HFD	Welch
Number	2	2	2	2	3	3

Table 5-10. The statistics of the sub-band-wise biomarkers of bagging and boosting model for classifying severity.

	Alpha	Beta	Delta	Theta	Gamma
Number	5	1	3	2	3

doi:10.6342/NTU202301910

5.3.2 Feature Importance for Classifying High Refractoriness

Tables 5.11 to 5.13 are the feature importance list of the bagging and boosting machine learning methods for classifying high refractoriness. The larger the value of the feature importance in the table, the more influential the feature is in that decision tree-based machine learning model.

Table 5-11. The feature importance of the Random Forest model for classifying high refractoriness.

Random Forest Feature Importance for Classifying High Refractoriness			
Feature Importance	**Sub-band**	**Electrode-Feature**	**P-value**
0.0363	Theta	FP1-Welch	0.0238*
0.0320	Theta	F3-KFD	0.0691
0.0309	Theta	F3-DFA	0.5472
0.0288	Gamma	FP2-HFD	0.5602
0.0271	Theta	FZ-HFD	0.5877

Table 5-12. The feature importance of the XGBoost model for classifying high refractoriness.

XGBoost Feature Importance for Classifying High Refractoriness			
Feature Importance	**Sub-band**	**Electrode-Feature**	**P-value**
0.0566	Alpha	FP1-DFA	0.0809
0.0546	Gamma	FP2-HFD	0.5602
0.0546	Delta	F4-ApEn	0.2512
0.0480	Theta	FP2-DFA	0.2589
0.0437	Beta	F7-Welch	0.0045*

doi:10.6342/NTU202301910

Table 5-13. The feature importance of the CatBoost model for classifying high refractoriness.

CatBoost Feature Importance for Classifying High Refractoriness			
Feature Importance	**Sub-band**	**Electrode-Feature**	**P-value**
0.4438	Delta	F8-LLE	0.5869
0.3023	Delta	F4-ApEn	0.2512
0.1980	Theta	F8-KFD	0.6207
0.0559	Beta	FP2-KFD	0.8684

We analyze the feature importance by electrode channel-wise, feature-wise, and sub-band-wise to get detail findings. According to Table 5.14 to Table 5.16, we know that the best models among the bagging and boosting models prefer to use FP1, FP2, F3, F4, and F8 channels, especially FP2 channel. The models' linear and nonlinear features are roughly equal except for LLE. And then, the models tend to choose the elements from the beta, delta, theta, and gamma bands, especially the theta band, for making predictions. Among them, Random Forest model demonstrate the best performance. However, based on the p-values, $p<0.05$, features were identified in Random Forest model and XGBoost model, specifically in theta band of FP1 channel and beta band of F7 channel, respectively.

doi:10.6342/NTU202301910

Table 5-14. The statistics of the channel-wise biomarkers of bagging and boosting model for classifying high refractoriness.

	FP1	FP2	F3	F4	F7	F8	FZ
Number	2	4	2	2	1	2	1

Table 5-15. The statistics of the feature-wise biomarkers of bagging and boosting model for classifying high refractoriness.

	LLE	ApEn	DFA	KFD	HFD	Welch
Number	1	2	3	3	3	2

Table 5-16. The statistics of the sub-band-wise biomarkers of bagging and boosting model for classifying high refractoriness.

	Alpha	Beta	Delta	Theta	Gamma
Number	1	2	3	6	2

doi:10.6342/NTU202301910

5.3.3 Feature Importance for Classifying Suicide Risk

Tables 5.17 to 5.19 are the feature importance list of the bagging and boosting machine learning methods for classifying suicide risk. The larger the value of the feature importance in the table, the more influential the feature is in that decision tree-based machine learning model.

Table 5-17. The feature importance of the Random Forest model for classifying suicide risk.

Random Forest Feature Importance for Classifying Suicide Risk			
Feature Importance	**Sub-band**	**Electrode-Feature**	**P-value**
0.0125	Delta	F7-HFD	0.1315
0.0094	Alpha	FP2-KFD	0.0681
0.0092	Gamma	FZ-HFD	0.0435*
0.0088	Gamma	F7-LLE	0.1124
0.0085	Gamma	FP1-KFD	0.0491*

Table 5-18. The feature importance of the XGBoost model for classifying suicide risk.

XGBoost Feature Importance for Classifying Suicide Risk			
Feature Importance	**Sub-band**	**Electrode-Feature**	**P-value**
0.0165	Delta	F3-Welch	0.9930
0.0161	Alpha	F7-DFA	0.6847
0.0158	Gamma	FP1-HFD	0.0046*
0.0154	Beta	F3-HFD	0.0249*
0.0137	Delta	F4-HFD	0.5535

doi:10.6342/NTU202301910

Table 5-19. The feature importance of the CatBoost model for classifying suicide risk.

CatBoost Feature Importance for Classifying Suicide Risk			
Feature Importance	**Sub-band**	**Electrode-Feature**	**P-value**
0.0207	Gamma	FZ-HFD	0.0435*
0.0180	Alpha	FP1-HFD	0.0856
0.0156	Delta	F7-HFD	0.1315
0.0151	Gamma	FP1-KFD	0.0491*
0.0127	Beta	FZ-HFD	0.6449

We analyze the feature importance by electrode channel-wise, feature-wise, and sub-band-wise to get some detail findings. According to Table 5.20 to Table 5.22, we can know that best models among the bagging and boosting models prefer to use FP1, F3, F7, and FZ channels. Among linear and nonlinear features, the models prefer to select KFD and HFD features, which comprise most of the top important features. And then, the models tend to choose the features from the alpha, beta, delta, and gamma bands, especially the gamma band for making predictions. Among them, Random Forest model demonstrate the best performance. However, based on the p-values, $p<0.05$, features were identified in all three models, specifically in beta and gamma band of FP1, F3, and FZ channel.

doi:10.6342/NTU202301910

Table 5-20. The statistics of the channel-wise biomarkers of bagging and boosting model for classifying suicide risk.

	FP1	FP2	F3	F4	F7	F8	FZ
Number	4	1	2	1	4	0	3

Table 5-21. The statistics of the feature-wise biomarkers of bagging and boosting model for classifying suicide risk.

	LLE	ApEn	DFA	KFD	HFD	Welch
Number	1	0	1	3	9	1

Table 5-22. The statistics of the sub-band-wise biomarkers of bagging and boosting model for classifying suicide risk.

	Alpha	Beta	Delta	Theta	Gamma
Number	3	2	4	0	6

doi:10.6342/NTU202301910

5.4 The Interface Displaying Details and Results

The interface is designed to enhance the transparency and interpretability of our work and make it easy for physicians to use it. The execution process is described below.

1. Click on the Python file to open the interface.
2. The default file extension for input EEG files is “edf”. If the file names are not “edf”, just simply modify in the entry.
3. Click on the “Load EEG Data” button to select the folder that contains the EEG data. Ensure that the folder is in the same directory as the Python file, then click on the folder.
4. Wait until the entry below displays “Loading Finished!!” and then click on the “Preprocessing” button to perform preprocessing of EEG data.
5. The preprocessing steps will be displayed sequentially in the entry below. Once it shows “Preprocessing Finish!!”, it indicates the process has been completed.
6. Click on the “Predict and Report” button. A table representing the average spectrum between different frequency bands will appear in the lower section. In the middle-upper section, each model will have buttons and a map of electrode positions. The upper-right section will display signal graphs for each EEG channel, while the lower-right section will show the prediction results for each task. Also, a pdf file containing the table and prediction results will be create in the same directory with the data.
7. Clicking on the buttons for each model will display the top three important features that contribute to the positive or negative prediction. The button color turns red when the model predicts positive and turns green when the model predicts negative. The textbox below in the middle also provides explanations in Chinese for the meaning of each feature, making it easier for patients to understand.
8. In the EEG signal graph, pressing the “z” key on the keyboard will hide the scrollbar

doi:10.6342/NTU202301910

at the bottom and right side. Clicking on the “Help” button in the bottom-left corner will provide instructions on how to navigate the window, allowing for a clearer view of the desired section of EEG signals.

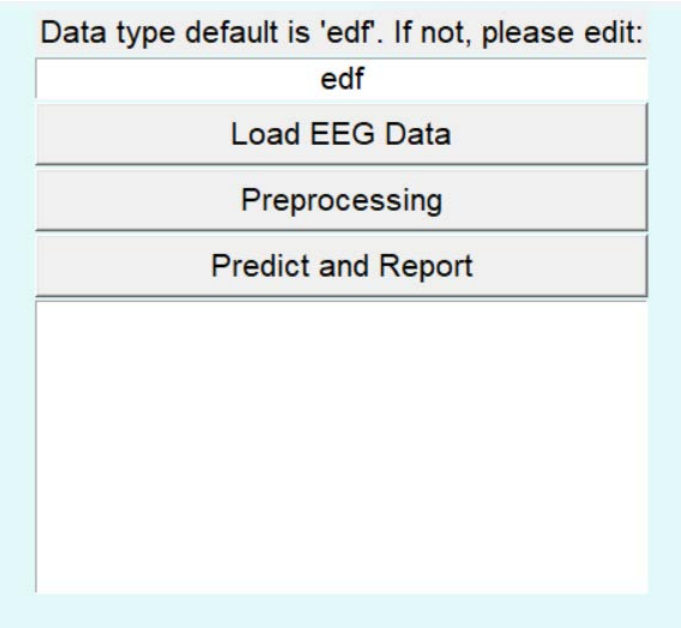


Figure 5-2. The initial interface while executing the python file.

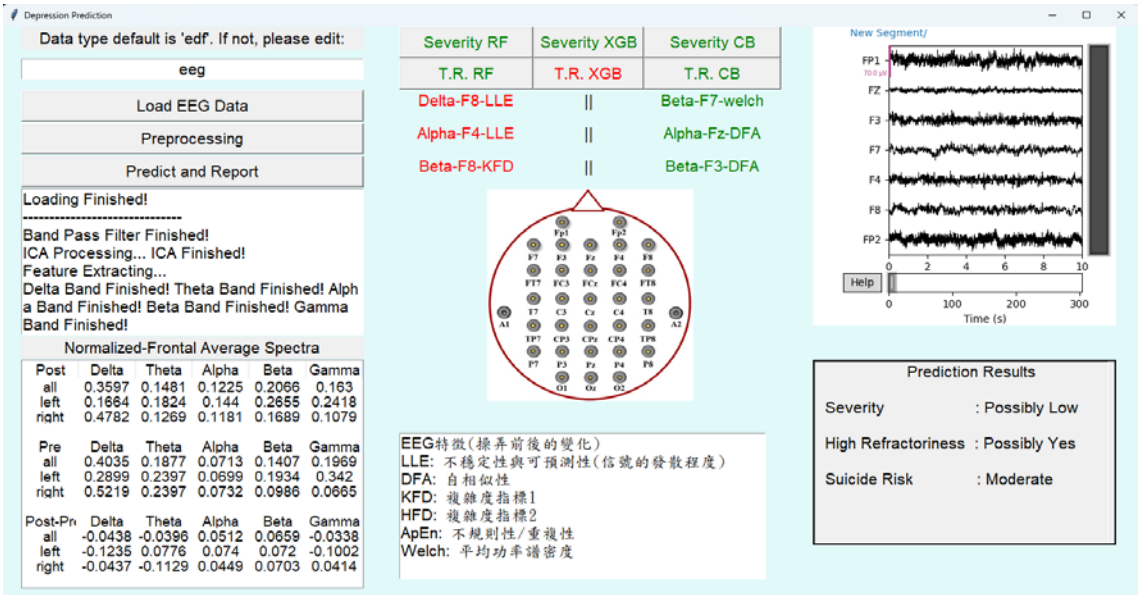


Figure 5-3. The prediction results detailed in the interface.

doi:10.6342/NTU202301910

Chapter 6 Discussion and Conclusion

For classifying severity, the Random Forest model identified the most important feature with p-value less than 0.05, specifically in alpha band of FP2 channel. Similar finding from previous research has concluded that individuals with depression have much lower alpha wave in the frontal brain regions compared to healthy individuals. Therefore, measuring alpha waves using EEG may serve as a potential biomarker for distinguishing between healthy and depressed individuals [68].

For classifying high refractoriness, the most important features were identified in Random Forest model and XGBoost model, specifically welch in theta band of FP1 channel and beta band of F7 channel, respectively. Band power from beta band may be helpful for diagnosis using classifiers [39][69], which is like the present study's findings.

For classifying suicide risk, the most important features were identified in all three models, specifically KFD and HFD in beta and gamma band of FP1, F3, and FZ channel. Previous research has found that the greater activation of the left hemisphere that is more typical for the EEG of individuals with an increased risk for suicide [70]. Also, some studies have found that the frontal lobe is related to suicide behaviors in depressed individuals [71][72], which is in line with the present study's findings. However, the latest study has reported that suicide ideation and suicide attempt may involve separate neural circuits, indicating that different factors contribute to these behaviors, and individuals at high risk for suicide can be found even in non-depressed populations [73].

We utilized all the features extracted from the EEG data, and despite conducting statistical analysis to select relevant features, either none or very few features were identified. As a result, the model could not be trained successfully. Therefore, we decide to use all the features in our analysis. However, it is important to note that although we

doi:10.6342/NTU202301910

use all the features, there is still room for improvement in the results. One of the potential reasons for this could be attributed to the fact that the brain regions associated with the severity, high refractoriness, and suicide, which we focus on in this study, are more extensive compared to the brain regions considered in previous works that we referenced. In other words, there is a greater complexity and diversity in the neural correlates of suicide, and incorporating a broader range of brain regions and associated features may lead to further improvements in our model's performance.

For the part dealing with the imbalanced data, we use SMOTE technique to generate synthetic data to balance the number of each class. SMOTE generates synthetic data through a linear approach when synthesizing data. However, our features consist of both linear and nonlinear components. Therefore, while using SMOTE in this paper does improve the model's performance, it is not the most appropriate approach. It does not fully capture the complexities of the nonlinear relationships present in the data. As a result, the synthetic data generated by SMOTE may not accurately represent the underlying distribution of the feature space.

To determine the final prediction results, we employ a voting mechanism using the best model of bagging and boosting models. Specifically, we take the majority vote from the three models to classifying the severity and high refractoriness. When the predictions of the three models are in a ratio of 2 to 1, the predicted result will include the term "possibly." However, in classifying suicide risk, due to the low accuracy of the boosting models, we opt to use eight bagging models (RFs) to determine the result. The decision-making process involves averaging the predictions of the eight models. If the average value is greater than the mode, it indicates that the non-mode classification is more severe, considering the influence of outliers. Therefore, the result is the mode plus one. Conversely, if the average value is smaller than the mode, it suggests that the suicide level

doi:10.6342/NTU202301910

may not be as high. Thus, the result is the mode minus one.

In conclusion, first, with the bagging and boosting machine learning algorithms using linear and nonlinear features extracted from the raw frontal EEG signals, we obtained excellent accuracy on all classification tasks (Severity RF accuracy=82.8%, High Refractoriness RF accuracy=86.7%, and Suicide Risk RF accuracy=97.6%). Second, Random Forest performs the best among the three models. However, based on the top five important features from each model, we found not only Random Forest but also the other models show features with a p-value less than 0.05, which means the other models also find some important biomarkers on the classification task. When differentiating severity, the model tends to prefer alpha, delta, and gamma bands. It also favors the channels, including FP1, FP2, F7, and F8, with an equal representation of linear and nonlinear features. Among these features, DFA in alpha band of FP2 channel stands out as it reflects the activity of the prefrontal cortex. The reduction in activity in this region is associated with the severity of depression. When differentiating treatment refractoriness, the model tends to prefer delta, theta, and gamma bands, with theta being the most prominent. It shows a preference for FP2 channel, especially on the right side, and an equal representation of linear and nonlinear features. Among these features, the ones that exhibit significant differences are theta band in FP1 channel and beta band in F7 channel, as identified by Welch analysis. The linear features can predict treatment refractoriness and are in line with previous research findings. After manipulation using RECT, frontal theta can predict the effectiveness of antidepressant treatment. When differentiating suicide risk, the model tends to prefer the alpha, delta, and gamma bands, with a higher emphasis on gamma band. It shows a preference for FP1, F3, F7, and FZ channels, mostly concentrated on the left side and midline. The prominent features are centered around KFD and HFD. Among the selected significant features, both KFD and

doi:10.6342/NTU202301910

HFD represent self-similarity and are concentrated in gamma frequency band. Prior research has utilized gamma band to classify the intensity of negative emotions and combined it with self-similarity features, suggesting that patients may be immersed in negative emotions, potentially increasing the likelihood of suicidal behavior. Finally, an interface is designed to enhance the transparency and interpretability of our work and promote effective communication between doctors and patients.

doi:10.6342/NTU202301910

Chapter 7 Future Work

To continue this work in the future, firstly, we may collect more EEG data. The data samples of this thesis are still too small for machine learning. Increasing sample sizes can make the model to learn more information and enhance the robustness and stability of the model. Second, we can combine other neural signal time series or neuroimaging data, such as computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), PET-CT, PET-MRI, or functional near infrared spectroscopy (fNIRS), to obtain more insight brain mechanism and receive more solid multi-model predictions. Third, we can include non-frontal regions associated with these classification tasks may more precisely link to the pathophysiology of different aspects MDD and find more accurate biomarkers. Fourth, we can try the other artificial intelligence approaches, such as deep learning, semi-supervised learning, meta learning, few-shot learning, or zero-shot learning to use concepts and similarities to infer and predict new information when utilizing unlabeled data for the learning process and enhance the interpretability and explanation of prediction results.

Finally, addressing the limitation of SMOTE, alternative methods that can handle both linear and nonlinear relationships in the data should be considered. These methods can include nonlinear oversampling techniques such as Kernel SMOTE or approaches that utilize generative models like Variational Autoencoders (VAEs) or Generative Adversarial Networks (GANs). These techniques can better capture the underlying distribution and preserve the nonlinear characteristics of the data, leading to more accurate and reliable synthetic samples.

doi:10.6342/NTU202301910

Reference

[1] C. Otte *et al.*, “Major depressive disorder,” *Nat. Rev. Dis. Primers*, vol. 2, no. 1, p. 16066, 2016.

[2] R. H. Belmaker and G. Agam, “Major depressive disorder,” *N. Engl. J. Med.*, vol. 358, no. 1, pp. 55–68, 2008.

[3] “Depressive disorder (depression),” *Who.int*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression.

[4] M. S. Reddy, “Depression: the disorder and the burden,” *Indian J. Psychol. Med.*, vol. 32, no. 1, pp. 1–2, 2010.

[5] Research leaders, “Depression,” *Edu.au*, 08-Dec-2016. [Online]. Available: https://qbi.uq.edu.au/brain/brain-diseases/depression.

[6] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders, fifth edition (DSM-5)*. Barking, England: Lulu.com, 2020.

[7] M. Hamilton, “A rating scale for depression,” *J. Neurol. Neurosurg. Psychiatry*, vol. 23, no. 1, pp. 56–62, 1960.

[8] S. Olbrich and M. Arns, “EEG biomarkers in major depressive disorder: discriminative power and prediction of treatment response,” *Int. Rev. Psychiatry*, vol. 25, no. 5, pp. 604–618, 2013.

[9] J. H. Meyer, S. Cervenka, M.-J. Kim, W. C. Kreisl, I. D. Henter, and R. B. Innis, “Neuroinflammation in psychiatric disorders: PET imaging and promising new targets,” *Lancet Psychiatry*, vol. 7, no. 12, pp. 1064–1074, 2020.

[10] M. L. Phillips *et al.*, “Identifying predictors, moderators, and mediators of antidepressant response in major depressive disorder: neuroimaging approaches,” *Am. J. Psychiatry*, vol. 172, no. 2, pp. 124–138, 2015.

doi:10.6342/NTU202301910

[11] N. Kubitz, M. Mehra, R. C. Potluri, N. Garg, and N. Cossrow, “Characterization of treatment resistant depression episodes in a cohort of patients from a US commercial claims database,” *PLoS One*, vol. 8, no. 10, p. e76882, 2013.

[12] “Committee for medicinal product for human use (chmp) concept paper on the need for revision of note for guidance on clinical investigation of medicinal products in the treatment of depression with regard to treatment resistant depression,” *Europa.eu*, 2009. [Online]. Available: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/10/WC500004861.pdf.

[13] M. Fava and K. G. Davidson, “Definition and epidemiology of treatment-resistant depression,” *Psychiatr. Clin. North Am.*, vol. 19, no. 2, pp. 179–200, 1996.

[14] T. Petersen *et al.*, “Empirical testing of two models for staging antidepressant treatment resistance,” *J. Clin. Psychopharmacol.*, vol. 25, no. 4, pp. 336–341, 2005.

[15] “Suicide,” *Who.int*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/suicide.

[16] K. Hawton, C. Casañas I Comabella, C. Haw, and K. Saunders, “Risk factors for suicide in individuals with depression: a systematic review,” *J. Affect. Disord.*, vol. 147, no. 1–3, pp. 17–28, 2013.

[17] M. K. Nock, G. Borges, E. J. Bromet, C. B. Cha, R. C. Kessler, and S. Lee, “The epidemiology of suicide and suicidal behavior,” in *Suicide: Global perspectives from the WHO World Mental Health Surveys , (pp*, vol. 391, M. K. Nock, Ed. New York, NY, US: Cambridge University Press, xi, 2012, pp. 5–32.

[18] H. Cai *et al.*, “Prevalence of suicidal ideation and planning in patients with major depressive disorder: A meta-analysis of observation studies,” *J. Affect. Disord.*, vol.

doi:10.6342/NTU202301910

293, pp. 148–158, 2021.

[19] F. S. de Aguiar Neto and J. L. G. Rosa, “Depression biomarkers using non-invasive EEG: A review,” *Neurosci. Biobehav. Rev.*, vol. 105, pp. 83–93, 2019.

[20] N. Bajaj, “Wavelets for EEG analysis,” in Wavelet Theory, IntechOpen, 2021.

[21] C.-H. Lin, “Using rs-fMRI Signals for Prediction of Major Depressive Disorder Based on Machine Learning and Deep Learning,” *Master Thesis*, National Taiwan University, 2022.

[22] M.-Y. Chen, “Using FDG-PET Brain Signals for Diagnosing Major Depressive Disorder Based on Machine Learning and Deep Learning,” *Master Thesis*, National Taiwan University, 2022.

[23] C.-S. Chen, “EEG Analysis for Prediction of Antidepressant Responses of Transcranial Magnetic Stimulation in Major Depressive Disorder Based on Attentional Convolution Time Series Prototypical Neural Network Model and Classical/Quantum Machine Learning Approaches,” *Master Thesis*, National Taiwan University, 2021.

[24] C.-T. Li, *重覆經顱磁刺激 rTMS 與抗鬱藥物治療頑固型憂鬱症 TRD*. 2022.

[25] A. L. Brody, M. W. Barsom, R. G. Bota, and S. Saxena, “Prefrontal-subcortical and limbic circuit mediation of major depressive disorder,” *Semin. Clin. Neuropsychiatry*, vol. 6, no. 2, pp. 102–112, 2001.

[26] G. Domes *et al.*, “The neural correlates of sex differences in emotional reactivity and emotion regulation,” *Hum. Brain Mapp.*, vol. 31, no. 5, pp. 758–769, 2010.

[27] R. J. Davidson, W. Irwin, M. J. Anderle, and N. H. Kalin, “The neural substrates of affective processing in depressed patients treated with venlafaxine,” *Am. J. Psychiatry*, vol. 160, no. 1, pp. 64–75, 2003.

[28] N. S. Lawrence *et al.*, “Subcortical and ventral prefrontal cortical neural responses

doi:10.6342/NTU202301910

to facial expressions distinguish patients with bipolar disorder and major depression," *Biol. Psychiatry*, vol. 55, no. 6, pp. 578–587, 2004.

[29] C.-T. Li, T.-P. Su, S.-J. Wang, P.-C. Tu, and J.-C. Hsieh, "Prefrontal glucose metabolism in medication-resistant major depression," *Br. J. Psychiatry*, vol. 206, no. 4, pp. 316–323, 2015.

[30] C.-T. Li et al., "Cognition-modulated frontal activity in prediction and augmentation of antidepressant efficacy: A randomized controlled pilot study," Cereb. Cortex, vol. 26, no. 1, pp. 202–210, 2016.

[31] P. Zimmermann and B. Fimm, "Test for Attentional Performance (TAP)," Psytest Press, 1997.

[32] C.-T. Hsu, "探討經由認知作業程式驅動後的前額葉 Theta 波預測憂鬱症病患未來之療效以及認知作業程式開發與驗證," National Taiwan University, 2017.

[33] W. Mumtaz, S. S. A. Ali, M. A. M. Yasin, and A. S. Malik, "A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD)," *Med. Biol. Eng. Comput.*, vol. 56, no. 2, pp. 233–246, 2018.

[34] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, 2014.

[35] H. U. Amin *et al.*, "Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques," *Australas. Phys. Eng. Sci. Med.*, vol. 38, no. 1, pp. 139–149, 2015.

[36] X. Li *et al.*, "Depression recognition using machine learning methods with different feature generation strategies," *Artif. Intell. Med.*, vol. 99, no. 101696, p. 101696, 2019.

[37] H. Akbari *et al.*, "Depression recognition based on the reconstruction of phase

doi:10.6342/NTU202301910

space of EEG signals and geometrical features," *Appl. Acoust.*, vol. 179, no. 108078, p. 108078, 2021.

[38] H. Peng *et al.*, "Multivariate pattern analysis of EEG-based functional connectivity: A study on the identification of depression," *IEEE Access*, vol. 7, pp. 92630–92641, 2019.

[39] D. K. Prasad, S. Liu, S.-H. A. Chen, and C. Quek, "Sentiment analysis using EEG activities for suicidology," *Expert Syst. Appl.*, vol. 103, pp. 206–217, 2018.

[40] C. Jiang, Y. Li, Y. Tang, and C. Guan, "Enhancing EEG-based classification of depression patients using spatial information," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 566–575, 2021.

[41] O. Faust, P. C. A. Ang, S. D. Puthankattil, and P. K. Joseph, "Depression diagnosis support system based on EEG signal entropies," *J. Mech. Med. Biol.*, vol. 14, no. 03, p. 1450035, 2014.

[42] B. Hosseinifard, M. H. Moradi, and R. Rostami, "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal," *Comput. Methods Programs Biomed.*, vol. 109, no. 3, pp. 339–345, 2013.

[43] T. Althobaiti, S. Katsigiannis, D. West, and N. Ramzan, "Examining human-horse interaction by means of affect recognition via physiological signals," *IEEE Access*, vol. 7, pp. 77857–77867, 2019.

[44] Y. Mohammadi, M. Hajian, and M. H. Moradi, "Discrimination of depression levels using machine learning methods on EEG signals," in *2019 27th Iranian Conference on Electrical Engineering (ICEE)*, 2019.

[45] A. Pigoni, G. Delvecchio, D. Madonna, C. Bressi, J. Soares, and P. Brambilla, "Can Machine Learning help us in dealing with treatment resistant depression? A

doi:10.6342/NTU202301910

review,” *J. Affect. Disord.*, vol. 259, pp. 21–26, 2019.

[46] Y.-P. Lin, Y.-H. Yang, and T.-P. Jung, “Fusion of electroencephalographic dynamics and musical contents for estimating emotional responses in music listening,” *Front. Neurosci.*, vol. 8, p. 94, 2014.

[47] Wikipedia contributors, “10–20 system (EEG),” *Wikipedia, The Free Encyclopedia*, 16-Mar-2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=10%E2%80%9320_system_(EEG)&oldid=1144970353.

[48] “Nuamps,” *Compumedicsneuroscan.com*. [Online]. Available: https://compumedicsneuroscan.com/tag/nuamps/.

[49] H. J. Landau, “Sampling, data transmission, and the Nyquist rate,” Proc. IEEE Inst. Electr. Electron. Eng., vol. 55, no. 10, pp. 1701–1706, 1967.

[50] E. Oja, A. Hyvarinen, and J. Karhunen, Independent Component Analysis, 1st ed. Nashville, TN: John Wiley & Sons, 2004.

[51] H. Kantz and T. Schreiber, Nonlinear Time Series Analysis, 2nd ed. Cambridge, England: Cambridge University Press, 2010.

[52] L. F. Márton, S. T. Brassai, L. Bakó, and L. Losonczi, “Detrended fluctuation analysis of EEG signals,” Procedia Technol., vol. 12, pp. 125–132, 2014.

[53] M. L. Steyn-Ross, D. A. Steyn-Ross, J. W. Sleigh, and D. T. J. Liley, “Theoretical electroencephalogram stationary spectrum for a white-noise-driven cortex: Evidence for a general anesthetic-induced phase transition,” Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics, vol. 60, no. 6, pp. 7299–7311, 1999.

[54] M. J. Katz, “Fractals and the analysis of waveforms,” Comput. Biol. Med., vol. 18, no. 3, pp. 145–156, 1988.

[55] T. Higuchi, “Approach to an irregular time series on the basis of the fractal theory,”

doi:10.6342/NTU202301910

Physica D, vol. 31, no. 2, pp. 277–283, 1988.

[56] P. Welch, “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,” IEEE Trans. Audio Electroacoust., vol. 15, no. 2, pp. 70–73, 1967.

[57] S.-J. Yen and Y.-S. Lee, “Under-sampling approaches for improving prediction of the minority class in an imbalanced dataset,” in Intelligent Control and Automation, Springer Berlin Heidelberg, 2006, pp. 731–740.

[58] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.

[59] C. Cortes and V. Vapnik, “Support-vector networks,” Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[60] “非線性支持向量機器 (Non-linear SVMs),” 逍遙文工作室, 20-May-2012. [Online]. Available: https://cg2010studio.com/2012/05/20/%E9%9D%9E%E7%B7%9A%E6%80%A7%E6%94%AF%E6%8C%81%E5%90%91%E9%87%8F%E6%A9%9F%E5%99%A8-non-linear-svms/.

[61] M. R. N. Kousarrizi, A. A. Ghanbari, M. Teshnehlab, M. A. Shorehdeli, and A. Gharaviri, “Feature extraction and classification of EEG signals using wavelet transform, SVM and artificial neural networks for brain computer interfaces,” in 2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, 2009.

[62] L. Breiman, “Random Forests,” Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[63] Wikipedia contributors, “Random forest,” Wikipedia, The Free Encyclopedia, 06-May-2023. [Online]. Available:

doi:10.6342/NTU202301910

https://en.wikipedia.org/w/index.php?title=Random_forest&oldid=1153432547.

[64] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.

[65] R. Guo, Z. Zhao, T. Wang, G. Liu, J. Zhao, and D. Gao, “Degradation state recognition of piston pump based on ICEEMDAN and XGBoost,” Appl. Sci. (Basel), vol. 10, no. 18, p. 6593, 2020.

[66] A. V. Dorogush, V. Ershov, and A. Gulin, “CatBoost: gradient boosting with categorical features support,” arXiv preprint arXiv: 1810.11363, 2018.

[67] S. Narkhede, “Understanding Confusion Matrix,” Towardsdatascience.com, 2018. [Online]. Available: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62.

[68] D. P. X. Kan and P. F. Lee, “Decrease alpha waves in depression: An electroencephalogram (EEG) study,” in *2015 International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, 2015.

[69] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, and D. P. Subha, “Automated EEG-based screening of depression using deep convolutional neural network,” *Comput. Methods Programs Biomed.*, vol. 161, pp. 103–113, 2018.

[70] A. F. Iznak, E. V. Iznak, E. V. Damyanovich, and I. V. Oleichik, “Differences of EEG frequency and spatial parameters in depressive female adolescents with suicidal attempts and non-suicidal self-injuries,” *Clin. EEG Neurosci.*, vol. 52, no. 6, pp. 406–413, 2021.

[71] K. van Heeringen, G.-R. Wu, M. Vervaet, M.-A. Vanderhasselt, and C. Baeken, “Decreased resting state metabolic activity in frontopolar and parietal brain regions is associated with suicide plans in depressed individuals,” *J. Psychiatr. Res.*, vol.

doi:10.6342/NTU202301910

84, pp. 243–248, 2017.

[72] M. A. Oquendo et al., “Positron emission tomography of regional brain metabolic responses to a serotonergic challenge and lethality of suicide attempts in major depression,” *Arch. Gen. Psychiatry*, vol. 60, no. 1, pp. 14–22, 2003.

[73] C.-T. Li and M.-B. Lee, “Evidence-Based Somatic Treatments and Brain Abnormalities for Suicidality in Patients with Depression,” *Journal of Suicidology*, vol. 17, no. 1, 2022.

[74] S.-H. Hung, I. Wu, Y.-C. Li, C.-T. Li, and C. C.-P. Chen, “A real time EEG analysis system for the prediction of clinical antidepressant responses,” in 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), 2018.

doi:10.6342/NTU202301910