INTERNSHIP REPORT

# EEG-Audio based multimodal classification of Major Depressive Disorder leveraging emotional attentional deficits

by


**Stuti Wadhwa**

**102197024**


Under the Guidance of


**Dr. Muskaan Singh, Assistant Professor, Ulster University, UK**

**Dr. Deepshikha Tiwari, Assistant Professor**


THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Submitted to the

**Computer Science & Engineering Department**

**Thapar Institute of Engineering & Technology, Patiala**

In Partial Fulfillment of the Requirements for the Degree of

Bachelor of Engineering in Computer Science and Engineering

at

Thapar Institute of Engineering & Technology, Patiala

**June, 2024**

**EEG based multimodal classification of Major Depressive Disorder using audio and EEG data**

By Stuti Wadhwa
Place of work: Ulster University, Northern Ireland, United Kingdom

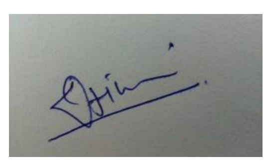
Submitted to the Computer Science & Engineering Department, Thapar Institute of Engineering & Technology

1st July, 2024

In Partial Fulfillment of the Requirements for the Degree of Bachelor of Engineering in Computer Science and Engineering

**Author:** Stuti Wadhwa Stuti

**Certified By:** Dr. Muskaan Singh, Assistant Professor Muskaan

**Certified By:** Dr. Deepshikha Tiwari, Assistant Professor 

# CERTIFICATE

## CERTIFICATE OF INTERNSHIP COMPLETION

This is to certify that **Ms. Stuti Wadhwa** has successfully completed an internship under my supervision at the Cognitive Analytics Research Lab, School of Computing, Engineering and Intelligent Systems, Ulster University, UK. This internship spanned six months, from **January 2nd, 2024 until June 30th, 2024**.

During this period, she has contributed significantly to two research projects:

1. **MEG-based Classification of Mild Cognitive Impairment in a Passive Audiovisual Task:** This project involved leveraging MEG signals to analyse attentional deficits in MCI patients through time-frequency analysis and Event-Related Potentials, ultimately developing classifiers and anomaly detectors to distinguish between healthy individuals and MCI patients.

2. **EEG-based Multimodal Classification of Major Depressive Disorder using Audio and EEG Data:** This project focused on utilizing EEG and audio signals, employing advanced signal processing and deep learning techniques to enhance the diagnostic process for MDD, aiming to develop a robust multimodal diagnostic pipeline.

Ms. Stuti has demonstrated exceptional skills and dedication throughout the internship, contributing to the successful completion of these projects.

Muskaan

Dr. Muskaan Singh

Assistant Professor

Cognitive Analytics Research Lab

Ulster University, Londonderry, UK

Date: 30th June, 2024

# ACKNOWLEDGEMENT

I would like to extend my heartfelt gratitude to those who have provided invaluable support and guidance throughout this project. First and foremost, I am deeply thankful to my faculty mentor, Dr. Deepshikha Tiwari, whose expert advice and continuous encouragement have been instrumental in the successful completion of this project. I am also profoundly grateful to my supervisor, Dr. Muskaan Singh, for her unwavering support and for providing me with this opportunity. Her practical insights and mentorship have greatly contributed to the project's success and my professional development.

I would also like to extend my deepest thanks to my friends and family for their constant support and encouragement. Their belief in me and their unwavering support have been a source of strength and motivation throughout this journey. Thank you all for your invaluable contributions and for making this project a rewarding experience.

Date: 1st July, 2024

# 1. COGNITIVE ANALYTICS RESEARCH LAB, ULSTER UNIVERSITY, UK

## 1.1. Overview

The Cognitive Analytics Research Lab (CARL) at Ulster University is a pioneering initiative aimed at establishing a world-class research capability in cognitive analytics. CARL is designed to attract significant local and international industry engagement, along with foreign direct investment, positioning the region as a global leader in this emerging field.

Ulster University has a longstanding reputation for excellence in data analytics, both in terms of developing advanced machine learning algorithms and applying analytical techniques across various domains. The university boasts over 60 academic staff dedicated to this research, contributing across all subject areas. CARL aims to consolidate and expand this expertise, growing into a 200-person center of excellence over the next five years. This unique center is built from the ground up through collaborative consultation with industry and civic stakeholders, focusing on both economic and societal impacts.

Fig 1: Organization Logo

## 1.2. Research

Ulster University's extensive experience in data analytics is supported by substantial investments in several key facilities and projects. These include:

- **Centre for Stratified Medicine:** A £11 million investment focused on tailored healthcare solutions.
- **Functional Brain Mapping Facility:** A £5 million facility dedicated to understanding brain function.

- **MIDAS (Meaningful Integration of Data, Analytics, and Services):** A €4.5 million project aimed at integrating data analytics services.
- **Capital Markets Collaboration:** A £1 million initiative enhancing data analytics in financial markets.
- **Centre for Precision Medicine:** An €8.6 million project using data to improve clinical decision-making and patient safety.

CARL brings together these investments, totaling around £30 million, to create a consolidated research hub. This integration aims to foster international collaboration and achieve world-leading research outcomes in cognitive analytics.

## 1.3. Cognitive Analytics and Data Science at CARL

Cognitive analytics represents the next paradigm shift in data analytics, combining artificial intelligence (AI) and machine learning with advanced data analytics techniques. As data generation continues to accelerate, with 2.5 quintillion bytes created daily, cognitive analytics harnesses this vast data pool to deliver intelligent, human-like insights.

Cognitive analytics involves several intelligent technologies, including semantics, AI algorithms, deep learning, and machine learning. These technologies enable systems to learn and improve over time, enhancing their effectiveness and intelligence. This approach offers significant benefits, such as mining untapped data sources, providing highly personalized services, improving service consistency and quality, and enhancing knowledge sharing.

The application of cognitive analytics creates a competitive advantage for businesses by offering real-time answers, contextual understanding, and the ability to sift through massive amounts of information to find optimal solutions. This capability opens up new opportunities for businesses to leverage data for strategic gains.

Ulster University, with its rich history of research excellence in neuro-inspired cognitive analytics and a substantial track record in data analytics, is uniquely positioned to lead the way in cognitive analytics research. Through CARL, the university is set to make significant strides in this field, contributing to both academic advancements and practical applications that have a global impact.

# EEG based multimodal classification of Major Depressive Disorder using audio and EEG data


## ABSTRACT

This research explores deep learning approaches for multi-modal signal fusion and classification to detect Major Depressive Disorder (MDD) using audio and Electroencephalography (EEG) data. Depression, a prevalent mental illness, is marked by persistent low mood, anhedonia, grief, and cognitive impairment, significantly impacting life quality. The World Health Organization (WHO) reports that over 350 million people globally suffer from depression. Despite its high incidence, depression often goes unrecognized, making accurate diagnostic methods critical. Current diagnosis relies heavily on physician consultations, which can be biased and inaccurate.

Neuroimaging technologies like positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have been used to study depression. However, PET requires radioactive substances, and fMRI is unsuitable for those with claustrophobia. EEG, with its high temporal resolution, non-invasiveness, ease of recording, and low cost, is optimal for studying depression.

Combining data from multiple modalities can provide a comprehensive and accurate diagnosis. In this project, we aim to integrate audio and EEG data using various fusion techniques, including CNN-based fusion and selective dropout, to develop a pipeline for MDD classification. Methods such as concatenation, weighting, and gating integrate these modalities at the input or feature level.

MDD patients often exhibit cognitive dysfunction, especially in attention control. Deficiencies in the cognitive control of emotional information are more pronounced in MDD patients compared to healthy controls (HCs). This study examines emotion-regulated cognitive competence in MDD during a dynamic attentional stage. ERPs were recorded from 24 clinical MDD outpatients and matched HCs using a modified affective priming dot-probe paradigm with various emotional facial expression pairs. Features extracted from P200 and P300 ERPs are used to investigate the neurocognitive mechanisms underlying dysregulated attentional control in MDD. These features will be combined with acoustic, linguistic, and paralinguistic features from audio data in the future to design a classification architecture for MDD.

# 2. INTRODUCTION

## 2.1. Project Overview

The increasing prevalence of Major Depressive Disorder (MDD) underscores the urgent need for effective diagnostic tools and treatment strategies. Traditional diagnostic methods often rely on subjective assessments, which can lead to misdiagnosis or delayed treatment. This project aims to address these challenges by leveraging advanced technologies in signal processing and deep learning to develop a robust multimodal diagnostic pipeline. The primary objective is to utilize EEG and audio signals to classify MDD accurately and explore various deep-learning-based feature fusion techniques to enhance the diagnostic process. By integrating these modalities, we aim to capture the complex interactions between neurological and behavioral markers of depression, providing a more comprehensive understanding of the disorder.

## 2.2. Nature and Scope of the Project

The scope of this project encompasses several critical stages, from data acquisition and preprocessing to feature extraction and deep learning model development. The project utilizes a multi-modal open dataset specifically curated for mental-disorder analysis, incorporating EEG and audio data from clinically depressed patients and healthy controls. The EEG data is processed using MATLAB's EEGLAB toolbox, which includes steps like rereferencing, filtering, and artifact removal using both automatic and manual techniques. Following preprocessing, features are extracted from EEG data using techniques such as Power Spectral Density (PSD) analysis and Event-Related Potential (ERP) analysis, ensuring a rich set of features for classification tasks. Audio data preprocessing involves handling recordings from various speaking tasks, including interviews, reading, and picture descriptions. The audio features, combined with the EEG features, form a comprehensive dataset for developing a deep-learning-based diagnostic model.

The project aims to explore various feature fusion techniques to integrate EEG and audio features effectively, leveraging the strengths of both modalities. Deep learning models, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are planned to be employed to analyze the fused features and classify MDD. The performance of these models will be evaluated to identify the most effective approach for multimodal MDD classification.

## 2.3. Objectives

- **Data Acquisition and Preprocessing:** To collect and preprocess a multi-modal dataset comprising EEG and audio signals from both MDD patients and healthy controls. This involves importing EEG data, rereferencing, filtering, artifact removal, and extracting relevant features from both EEG and audio data.
- **Feature Extraction:** To extract and quantify meaningful features from EEG and audio data. For EEG, this includes PSD and ERP features, while audio features are derived from various speech tasks. These features are crucial for capturing the underlying patterns associated with MDD.
- **Deep Learning Model Development:** To develop and train deep learning models that can effectively classify MDD based on the extracted features. This includes experimenting with different network architectures and hyperparameters to optimize performance.
- **Feature Fusion Techniques:** To explore and implement various deep-learning-based feature fusion techniques to integrate EEG and audio features. This step is critical for leveraging the complementary information provided by both modalities, aiming to improve classification accuracy.
- **Evaluation and Validation:** To evaluate the performance of the developed models using appropriate metrics and validation techniques. This involves assessing the accuracy, precision, recall, and overall robustness of the models in classifying MDD.
- **Research Contribution:** To contribute to the field of mental health diagnostics by developing a novel multimodal diagnostic pipeline. The findings and methodologies from this project are intended to be shared with the broader research community through publications and presentations.

This project is a collaborative effort, with specific responsibilities assigned to each team member based on their expertise. The responsibilities outlined above highlight the comprehensive approach taken to ensure the project's success and its potential impact on improving MDD diagnosis through advanced signal processing and deep learning techniques.

# 3. BACKGROUND AND RELEVANCE

Depression, as defined by the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), is characterized by the occurrence of one or more major depressive episodes lasting at least two weeks. These episodes are marked by symptoms such as a depressed mood, diminished interest or pleasure in most activities, and pervasive feelings of worthlessness or guilt [1]. It is classified as a mood disorder and manifests as persistent or episodic feelings of sadness, reduced enjoyment, and low self-esteem. Additionally, individuals may experience disturbances in sleep and eating patterns, concentration difficulties, and persistent fatigue. These symptoms can persist over time, leading to chronic and recurrent episodes that significantly impair an individual's ability to function in daily life.

Depression is a widespread mental health condition affecting a substantial portion of the global population. Approximately 280 million people, or about 5% of adults worldwide [2], suffer from depression. It is a major contributor to the global burden of disease and is identified as a potential precursor to suicide, with over 700,000 suicide-related deaths reported annually [2]. Depression severely hampers an individual's capacity to engage in professional, academic, and social activities, leading to a global loss or impairment of 50 million years of work annually [3]. Despite the severe impact of depression, mental health services and treatment remain inaccessible to over 75% of individuals in low- and middle-income countries [2].

The etiology of depression remains complex and multifaceted, with social, psychological, environmental, and medical conditions contributing to its development. This complexity necessitates a multidimensional approach to understanding and diagnosing depression. Psychological researchers, medical professionals, and technical experts are working to correlate symptoms with potential detection systems, aiming for early detection and self-assessment mechanisms to mitigate risks. However, the heterogeneity of depression's causes and symptoms poses significant challenges for conventional diagnostic methods, such as questionnaires and clinical assessments, due to their subjective nature and potential for bias [1].

Numerous studies have demonstrated that depression can be detected through the observation and analysis of subconscious states using a variety of methods and tools. Psychological approaches [4] often employ standardized questionnaires, interviews, or scales to assess the symptoms, severity, and impact of depression. Despite their widespread use, these methodologies have limitations, including subjectivity, potential bias, low sensitivity, and cultural differences. In contrast, machine learning (ML) approaches [5], [6], [7] utilize computational algorithms to analyze diverse modalities such as facial expressions, speech, text, and physiological signals. These techniques offer more objective and data-driven insights into the manifestation of depression.

Neuroimaging techniques, including electroencephalography (EEG), magnetic resonance imaging (MRI), and positron emission tomography (PET), have been employed to measure structural and functional changes in the brain associated with depression [8]. These methods provide valuable insights into the neurobiological mechanisms and potential biomarkers of depression. However, they are often limited by high computational costs, invasiveness, low accessibility, and technical challenges. Most existing methods rely on a single domain or modality of data, which may not fully capture the complexity and heterogeneity of depression. Different modalities can provide complementary or even contradictory information about the disorder.

To address these limitations, researchers have explored multimodal approaches that combine multiple data sources to provide a more comprehensive and accurate diagnosis of depression. Techniques such as concatenation, weighting, and gating are used to integrate multiple modalities at the input or feature level. Advanced methods like the multimodal deep learning framework (MDLF), cross-modal attention network (CMAN), deep convolutional neural network (DCNN), and bi-directional long short-term memory (BiLSTM) have shown promise in this area [9], [10], [11]. These methods utilize both feature (early) and dense (late) level concatenation with existing and custom algorithms. Despite their potential, a common drawback is the need for training on specific modalities, which may limit their generalizability. For example, a model trained on text and speech features may not effectively detect depression using images or physiological data.

Our project builds on these previous efforts by developing a robust multimodal diagnostic pipeline that integrates EEG and audio signals to classify Major Depressive Disorder (MDD) accurately. By leveraging advanced signal processing and deep learning techniques, our approach aims to capture the complex interactions between neurological and behavioral markers of depression. Unlike single-modality approaches, our method integrates multiple modalities to provide a more comprehensive understanding of the disorder. This integration is expected to enhance the diagnostic process, potentially leading to more accurate and timely identification of MDD. Additionally, our project explores various deep learning-based feature fusion techniques to maximize the complementary strengths of EEG and audio features, setting our work apart from previous studies. Through this innovative approach, we aim to contribute significantly to the field of mental health diagnostics and offer new insights into the detection and understanding of depression.

# 4. METHODOLOGY

## 4.1. Data Description

We utilized a multi-modal open dataset specifically curated for mental-disorder analysis, incorporating EEG and audio data from clinically depressed patients and healthy controls. The dataset consists of participants aged 18 to 55 with at least a primary education level. For Major Depressive Disorder (MDD) patients, inclusion criteria included meeting MINI diagnostic criteria for depression, a Patient Health Questionnaire-9 (PHQ-9) score of ≥5, and no psychotropic drug treatment within the last two weeks. Exclusion criteria for MDD patients included other mental disorders, brain damage, severe physical illness, and extreme suicidal tendencies. For healthy controls, exclusion criteria were a personal or family history of mental disorders.

### 4.1.1. Participants Information:

- **EEG Data:** 53 subjects (24 MDD patients: 13 males, 11 females, aged 16–56; 29 healthy controls: 20 males, 9 females, aged 18–55).
- **Audio Data:** 52 subjects (23 MDD patients: 16 males, 7 females, aged 16–56; 29 healthy controls: 20 males, 9 females, aged 18–55).

### 4.1.2. EEG Data:

EEG signals were recorded using a 128-channel HydroCel Geodesic Sensor Net and Net Station acquisition software at a sampling frequency of 250 Hz. Participants completed two tasks: a resting state task and a dot-probe task.

1. **Resting State:** 5 minutes of eyes-closed resting-state EEG were recorded. Participants were instructed to stay awake and still, avoiding unnecessary movements and eye blinks.
2. **Dot-Probe Task:** Participants focused on emotional-neutral face pairs displayed on a monitor and responded to the appearance of a dot by pressing a button. The task consisted of three blocks (Fear-Neutral, Sad-Neutral, and Happy-Neutral), each with 160 trials.

   In each trial:

   - A fixation cross appeared for 300 ms.
   - An emotional-neutral face pair was shown for 500 ms.
   - After a 100-300 ms interval, a dot appeared on either side of the cross for 150 ms.
   - Participants pressed '1' if the dot appeared on the left and '4' if it appeared on the right.
   - Each trial had a response window of 2000 ms, followed by a 600 ms black screen.
   - The entire task took approximately 25 minutes to complete.

From the MODMA dataset, the BIDS format EEG data was utilised, which specifies various aspects from the way in which EEG files should be stored and organized. BIDS specifies the file structure and metadata where, for each subject, there is a directory with the following elements:

- EEG raw data for each session in European Data format or .edf
- File listing the channels and electrodes saved as “channels.tsv” and “electrodes.tsv” file respectively
- Json file “coordsystem.json” providing the coordinate system file.
- Additional metadata file(s) which exhaustively specifies all the details of the stimuli in a .tsv file.

### 4.1.3. Audio Data:

Audio data were recorded in a controlled environment with ambient noise below 60 dB during three speaking tasks: an interview, reading, and picture description.

1. **Interview:** Participants answered 18 questions with positive, neutral, and negative implications based on DSM-IV and depression scales.

2. **Reading:** Participants read a short story and three groups of words categorized as positive, neutral, and negative.
3. **Picture Description:** Participants described four pictures with varying emotional content including images from the Chinese Facial Affective Picture System and the Thematic Apperception Test.

This comprehensive dataset provides a robust foundation for developing and evaluating a deep learning-based multimodal classification pipeline for Major Depressive Disorder detection using EEG and audio data.

## 4.2. Data Preprocessing

For processing EEG datasets, both in resting and task states, the EEGLAB toolbox in MATLAB was utilized. EEGLAB is an interactive MATLAB toolbox designed for processing continuous and event-related EEG, MEG, and other electrophysiological data. It offers a user-friendly graphical user interface (GUI) that allows users to flexibly and interactively process high-density EEG and other dynamic brain data using independent component analysis (ICA) and/or time-frequency analysis (TFA), in addition to standard averaging methods.

The following preprocessing steps were applied to each subject's EEG data for both the resting state and the dot probe task state:

### 4.2.1. Initial Data Preparation:

1. **Importing EEG Data:** The EEG data, stored in .edf files, were imported into EEGLAB. Channel locations were then set according to the Fieldtrip GSN-Hydrocel-128 format, matching the configuration of the EEG headset used for data collection.

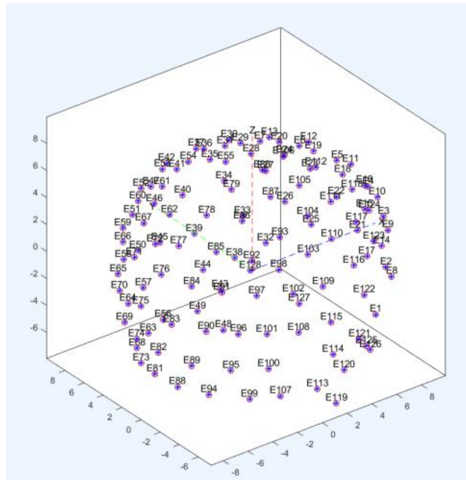


Fig 6: 3D layout of 128 EEG channels according to the GSN-Hydrocel 128 format

2. **Rereferencing:** The data were rereferenced to the average of all channels. The primary reason for using an average reference is to mitigate model errors inherent in the forward model of source projection to channels. Using a single common reference can introduce model error from that reference channel into all other channels. An average reference helps to distribute this error across all channels, reducing its impact. This approach assumes no specific directional bias in the model error, thus averaging out potential errors and improving the forward solution for each channel.
3. **Filtering:** The EEG recordings were bandpass filtered between 0.5 Hz and 40 Hz using an FIR filter.

   Reasons for high-pass filtering at 0.5 Hz:

   - **Removing Slow Drifts:** EEG data often contain low-frequency noise or slow drifts caused by factors such as electrode impedance changes, sweat, or subject movement. These slow drifts can distort ICA results by introducing large-scale variations unrelated to neural activity. High-pass filtering at 0.5 Hz removes these slow components, improving the signal-to-noise ratio. This allows ICA to better identify the underlying sources of brain activity by focusing on the frequency range where most meaningful EEG signals reside.
   - **Improving ICA Performance:** ICA algorithms assume that the data is centered and have a roughly zero mean over time. Low-frequency drifts can violate this assumption, leading to suboptimal separation of independent components. By applying a high-pass filter, the data becomes more stationary, aligning better with ICA algorithm assumptions.

Reasons for low-pass filtering at 40 Hz:

- **Noise Reduction:** Frequencies above 40 Hz often contain noise from muscle artifacts, electrical noise from equipment, and environmental interference. Low-pass filtering at 40 Hz removes these high-frequency noises, enhancing signal quality.
- **Focusing on Relevant Frequencies:** Most clinically relevant EEG signals related to brain activity and mental health, including those used in identifying MDD, fall within the 0.5 to 40 Hz range. Filtering out higher frequencies allows the analysis to focus on the most relevant data without the distraction of irrelevant high-frequency components.
- **Data Simplification:** Simplifying the data by removing high-frequency components makes subsequent analysis and interpretation easier.
- **Aliasing Prevention:** If the EEG data is sampled at a frequency insufficient to accurately capture frequencies above 40 Hz, these high frequencies can cause aliasing, misrepresenting higher frequencies as lower ones. Low-pass filtering at 40 Hz before downsampling prevents this issue.

By following this preprocessing pipeline, the EEG data is prepared for subsequent analysis with improved signal quality and reduced noise, facilitating the identification of biomarkers for Major Depressive Disorder.

### 4.2.2. Data Cleaning:

To ensure the quality and integrity of the EEG data before applying ICA, a rigorous data cleaning process was implemented for each subject. This process involved both automated and manual techniques to remove artifacts and other unwanted signals from the EEG recordings.

#### 9.2.2.1 Automatic Continuous Rejection

The initial phase of data cleaning utilized the Automatic Continuous Rejection method available in the EEGLAB toolbox. This method operates based on spectrum thresholding to identify and reject continuous portions of data that exhibit significant deviations from expected spectral characteristics.

- **Spectrum Thresholding:** Contiguous data epochs were extracted from the EEG data. A standard spectrum thresholding algorithm was then applied to these epochs. This algorithm identifies regions where the spectral power deviates beyond a predefined threshold, indicative of potential artifacts.

- **Artifact Identification:** Regions of contiguous epochs larger than a specified size, where the spectral power exceeded the threshold, were labeled as artifactual. These regions were marked for rejection to prevent the inclusion of spurious signals in the subsequent analysis.

Despite the effectiveness of the automatic rejection method, certain types of artifacts, particularly those resulting from eye blinks and muscle activity, often persisted. To address the se residual artifacts, a manual inspection and rejection step was conducted.

#### 9.2.2.2 Manual Inspection and Rejection

Following the automatic rejection process, the data was further cleaned through manual inspection using the “Inspect/Reject Data by Eye” feature in the EEGLAB GUI.

- **Visual Inspection:** Each EEG recording was carefully examined visually to identify any remaining artifacts that were not detected by the automatic rejection method. Particular attention was paid to identifying eye blinks, muscle artifacts, and other non-neural signals that could compromise the quality of the data.
- **Manual Rejection:** Identified artifacts were manually marked and removed. This step involved rejecting segments of data that showed clear evidence of contamination, ensuring that only clean and reliable data were retained for analysis.

EEGLAB facilitates this manual inspection process by inserting boundary events to mark discontinuities in the data resulting from epoch rejection or when the data is imported. These boundary events help in tracking and managing the portions of data that have been excluded, maintaining the integrity of the dataset.

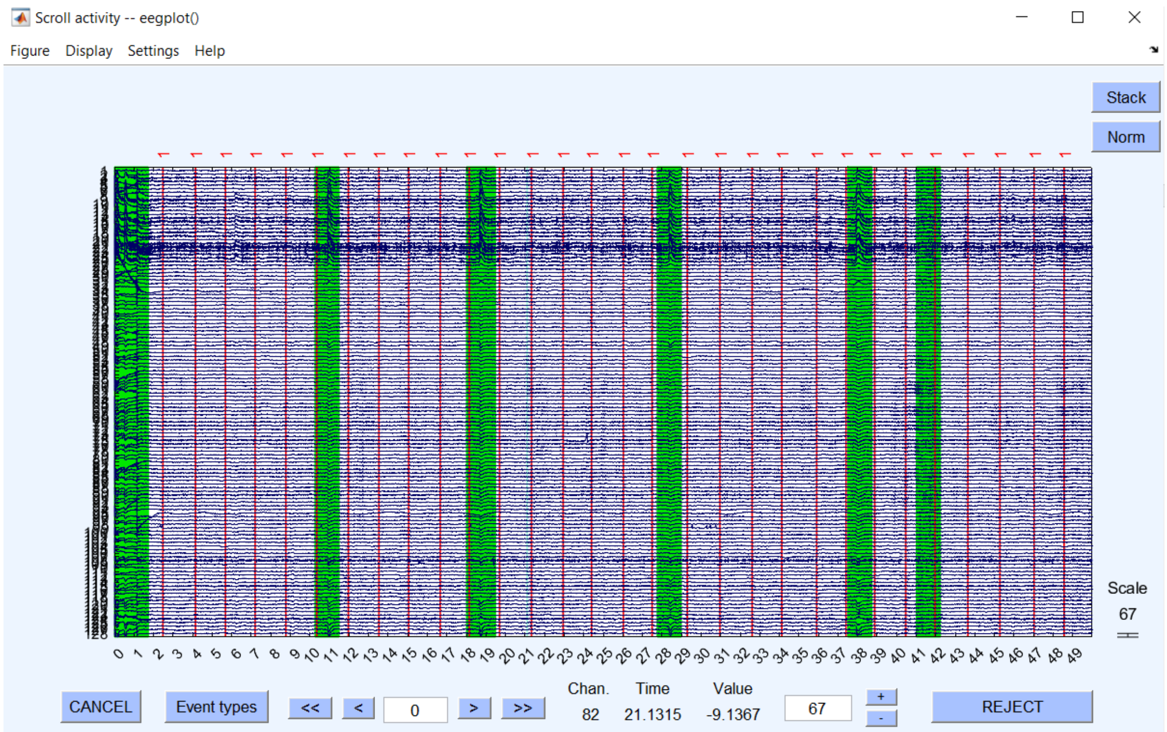



Fig 7: The “Reject data by eye” window with green parts marked for rejection

By combining automatic spectrum thresholding with meticulous manual inspection, this comprehensive data cleaning approach ensures the highest possible quality of EEG data. Removing these artifacts prior to ICA is crucial, as it enhances the ability of ICA to accurately separate independent sources of brain activity, leading to more reliable and meaningful results in the subsequent analysis of EEG signals.

## 4.3. Independent Component Analysis (ICA)

Independent component analysis (ICA) has been shown to be useful when applied to electroencephalographic (EEG) data. By unmixing the channel recordings into statistically independent component processes, the components can capture anatomically and functionally distinct brain source processes and can also separate out non-brain artifacts in the data. However, using ICA for EEG analysis can seem quite difficult at first. Its effective use requires overcoming two main problems: first, cleaning the data sufficiently, and then correctly interpreting the results.

### 9.3.1 Need of ICA:

Independent Component Analysis is used to remove artifacts embedded in the data (muscle, eye blinks, or eye movements) without removing the affected data portions. ICA may also be used to find brain sources.

Decomposing data by ICA (or any linear decomposition method, including PCA and its derivatives) involves a linear change of basis from data collected at single scalp channels into a spatially transformed “virtual channel” basis. In the original scalp channel data, each row of the data recording matrix represents the time course of summed in voltage differences between source projections to one data channel and one or more reference channels (thus itself constituting a linear spatial filter). After ICA decomposition, each row of the data activation matrix gives the time course of the activity of one component process spatially filtered from the channel data.

In the case of ICA decomposition, the independent component filters are chosen to produce the maximally temporally independent signals available in the channel data. These are, in effect, information sources in the data whose mixtures have been recorded at the scalp channels. These information sources may represent synchronous or partially synchronous activity within one (or possibly more) cortical patch(es), else activity from non-cortical sources (e.g., potentials induced by eyeball movements or produced by single muscle activity, line noise, etc.).

Since there is strong line noise visible in the data, the option “extended, 1’’ in the “decompose data by ICA” feature of EEGLAB GUI is used, which calls the function runica.m, so the algorithm can also detect subgaussian sources of activity, such as line current and/or slow activity.

### 9.3.2. Automatic component labelling with ICLabel:

EEGLAB's ICLabel plugin is an advanced tool designed to automatically classify Independent Components (ICs) extracted via ICA. This plugin leverages machine learning techniques to provide accurate and efficient labelling of ICs, facilitating the identification of neural and non-neural components. The classification helps researchers distinguish between various sources of EEG signals, including brain activity, eye movements, muscle activity, heartbeats, line noise, and channel noise.

Functionality and Workflow of ICLabel is as follows:

1. **Training Dataset:** The ICLabel classifier is trained on a large dataset of ICs manually labeled by experts. This dataset includes a diverse array of components from various studies, ensuring the classifier is robust and generalizable
2. **Feature Extraction:** When applied to new data, ICLabel first extracts relevant features from each IC. These features capture the spatial, spectral, and temporal characteristics of the components.
3. **Classification:** The extracted features are fed into the ICLabel's machine learning model, which assigns probabilities to each IC corresponding to different component classes. The primary classes include:

   - Brain: Neural activity reflecting genuine brain signals.
   - Eye: Artifacts resulting from eye movements, such as blinks or saccades.
   - Muscle: Artifacts from muscle activity, often seen as high-frequency noise.
   - Heart: Cardiac-related artifacts, typically rhythmic and low-frequency.
   - Line Noise: Electrical noise at a specific frequency (e.g., 50 or 60 Hz).
   - Channel Noise: Noise specific to certain electrodes, often due to poor contact or movement.

4. **Probabilistic Output:** ICLabel provides a probabilistic output for each IC, indicating the likelihood of belonging to each class. Researchers can review these probabilities to make informed decisions about which components to retain or reject.

### 9.3.3. Manual component Analysis:

While automatic component labeling in EEGLAB using ICLabel is a powerful tool for classifying independent components (ICs), it can sometimes leave out components with ambiguous probabilities that belong to multiple artifactual classes. For instance, consider an IC with a 35% probability of being a muscular artifact, a 39% probability of being a cardiac artifact, and a 25% probability of being an eye blink artifact. If our goal is to retain only brain components, this IC should be removed, as its probability of being a brain component is negligible.

However, if the criteria for discarding non-neural ICs are set to high probability thresholds (e.g., 80-90%), this component would not be flagged for removal. To address this issue, it is crucial to perform a manual re-evaluation of each component, focusing on their activations, 2D component maps, ERP images, and activity power spectrums. This detailed manual analysis ensures a more accurate classification, distinguishing between neural and non-neural components.

The detailed process of manual re-evaluation is as follows:

- **Component Activations:** Each IC's activation pattern is examined to identify temporal characteristics indicative of neural or artifactual sources. Neural components typically exhibit consistent activation patterns across trials, while artifactual components show irregular or sporadic activity.
- **2D Component Maps:** The spatial distribution of each IC is inspected using 2D component maps. Neural components often present distinct and localized patterns corresponding to known neural sources, whereas artifacts like muscle or eye movements display diffuse or atypical distributions.
- **ERP Images:** ERP images are analyzed to evaluate the temporal dynamics of each IC. Neural components usually align well with task-related events, showing clear ERP features. Artifacts, on the other hand, appear as noise or unrelated fluctuations.
- **Activity Power Spectrums:** The power spectrum of each IC is scrutinized to identify frequency characteristics. Neural components generally exhibit power in specific frequency bands (e.g., alpha, beta), whereas artifacts show power distributions indicative of non-neural sources (e.g., high-frequency noise from muscle activity).

By manually reviewing these aspects, each component is classified more accurately, ensuring that only those truly related to neural activity are retained. This process enhances the reliability of the data by minimizing the inclusion of artifactual components that could otherwise compromise the analysis.

The components identified as artifacts through manual re-evaluation are flagged in conjunction with those identified by the ICLabel plugin, followed by their removal. This comprehensive approach ensures that all non-neural components are effectively removed, leaving a cleaner dataset for subsequent analysis.

In conclusion, while automatic component labeling is an essential first step, manual component analysis is crucial for achieving the highest data quality. This methodical approach allows for a more precise distinction between neural and non-neural components, thereby enhancing the robustness and validity of the EEG data analysis.

## 9.4. Feature Extraction

For each subclass of the data, Power Spectral Density (PSD) and Event-Related Potentials (ERPs) were computed and utilized for feature extraction. The PSD of EEG data represents the distribution of power across different frequencies within the recorded neural signals. Extracting features from the PSD captures essential aspects of neural dynamics and can be crucial in distinguishing between different brain states or conditions. For example, features extracted from specific frequency bands, such as theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (>30 Hz), reflect distinct neural processes related to attention, memory, or motor function.

ERPs in EEG data refer to synchronized neural responses elicited by specific events or stimuli, such as sensory stimuli or cognitive tasks. ERPs are time-locked to the onset of the event, enabling the isolation and analysis of neural activity associated with particular cognitive or sensory events. In our study, these events consisted of emotional-neutral face pairs pseudo-randomly displayed on a monitor and a dot appearing on the screen soon after on either the left or right side, when the subject has to identify the side as emotional or neutral. This is done to record the accuracy as well as rection times (RTs) of the two classes of subjects. Features such as peak amplitude, latency, duration, and spatial distribution of ERP components encode essential information about cognitive processing, sensory perception, and neural dynamics. Features derived from PSD and ERP provide valuable insights into the underlying neural mechanisms, aiding in the identification of biomarkers or neural signatures associated with specific cognitive tasks or neurological disorders.

In this study, ERP features include peak amplitude, latency, area under the curve (AUC), slope, peak-to-peak amplitude, mean absolute amplitude, root mean square (RMS) amplitude, standard deviation, skewness, kurtosis, number of zero crossings, adaptive autoregressive coefficients (AAR), and zero-crossing rate over kurtosis (ZORK). Similarly, PSD features include power in the alpha band, power in the beta band, and spectral entropy in the alpha and beta bands.

### 9.4.1 PSD Feature Extraction

This process is a critical step in the analysis pipeline, aimed at quantifying the frequency-domain characteristics of the EEG signals. It is carried out using MNE toolbox in Python and the workflow of steps is as follows:

- **Data Preparation:** After preprocessing, the EEG data was stored in a .set file for each subject. Then, the EEG data for all subjects were loaded and stored in a list. The data were read using the `mne.io.read_raw_eeglab` function. Each raw object contains the continuous EEG data for a single subject, and this data is subsequently used for PSD and ERP computation.
- **PSD Computation:** The Power Spectral Density (PSD) of the EEG signals was computed using the multitaper method, which is known for providing a robust estimate of the power spectrum. The frequency range of interest was set from 0 to 40 Hz to encompass the most relevant EEG frequency bands while excluding high-frequency noise. The `compute_psd` method from MNE was used for this purpose.

- **Feature Extraction:** The extracted PSD data were then used to compute a set of statistical and frequency-domain features for each EEG channel. The features were computed as follows:

  **1. Statistical Features:**

  - **Mean Power:** The average power across the entire frequency range.
  - **Variance of Power:** The variability of the power across the frequency range.

  **2. Frequency-Domain Features:**

  - **Maximum Power Frequency:** The frequency at which the power is maximized.
  - **Maximum Power:** The peak power value in the spectrum.
  - **Band Power in Alpha Range (8-13 Hz):** The total power within the alpha band, which is associated with relaxation and cognitive processes.
  - **Band Power in Beta Range (13-30 Hz):** The total power within the beta band, which is related to active thinking and focus.
  - **Spectral Entropy in Alpha Range:** The entropy of the power distribution within the alpha band, indicating the complexity of the signal.

- **Spectral Entropy in Beta Range:** The entropy of the power distribution within the beta band, reflecting the signal's complexity in this frequency range.

For each channel, these features were computed and stored in a structured format. The process was repeated for all channels and subjects, resulting in a comprehensive dataset of PSD features.

### 9.4.2 ERP Feature Extraction

In our study, feature extraction from Event-Related Potentials (ERPs) of EEG data was conducted using a comprehensive approach that involves multiple signal processing techniques.

- **Data Preparation and Epoch Extraction:** The preprocessing steps began by importing raw EEG data and identifying event markers. Annotations in the data were used to extract events, which were subsequently mapped to specific event IDs. These IDs corresponded to different experimental conditions. Epochs were then defined within a time window of 0 to 1.5 seconds relative to the event onset, with a baseline correction applied from the beginning of the epoch to the event onset.
- **ERP Averaging and Features Extraction:** Averaging the epochs resulted in ERP waveforms for each channel, providing a time-locked representation of neural responses to the stimuli. The ERP data matrix, representing the averaged signal across all channels and time points, was then used for feature extraction.

  To capture the critical characteristics of the ERP signals, the following features were extracted for each channel:

  - **Peak Amplitude and Latency:** The maximum amplitude of the ERP waveform and its corresponding latency were identified, indicative of the neural response strength and timing.
  - **Peak-to-Peak Amplitude:** This measure captures the difference between the maximum and minimum values of the ERP signal, reflecting the overall signal amplitude variation.
  - **Mean Absolute Amplitude:** The average absolute value of the ERP waveform was computed to provide a measure of the signal's average strength.

- **Root Mean Square (RMS) Amplitude:** The RMS value of the ERP signal was calculated, offering another perspective on the signal's amplitude by considering both positive and negative deflections.
- **Standard Deviation:** This feature quantifies the variability of the ERP signal, providing insight into the consistency of the neural responses.
- **Skewness:** The asymmetry of the ERP waveform was assessed using skewness, which can indicate the presence of predominant positive or negative deflections.
- **Kurtosis:** The peakedness of the ERP signal distribution was measured using kurtosis, helping to identify waveforms with sharp peaks or flat tops.
- **Zero Crossing Rate:** The number of times the ERP waveform crosses the zero axis was counted, providing information on the signal's oscillatory nature.
- **Area Under the Curve (AUC):** The AUC was computed using numerical integration to quantify the overall magnitude of the ERP signal over the specified time window.
- **Slope:** The slope of the ERP signal was determined by fitting a linear regression model, capturing the general trend of the waveform over time.
- **Adaptive Autoregressive Parameters (AAR):** AAR parameters capture the dynamic aspects of EEG signals, providing insights into the temporal dynamics of brain responses that help in identifying cognitive and emotional processing abnormalities in MDD.
- **Zero Crossing to Kurtosis Ratio (ZORK):** The ratio of zero crossings to kurtosis was computed, combining information on the signal's oscillatory behavior and peakedness.

This extensive feature extraction process ensures that a wide range of signal characteristics are captured, enabling robust and nuanced analysis of ERP data. By combining both time-domain and statistical features, our approach provides a detailed representation of the neural responses, essential for identifying biomarkers associated with MDD.

### 9.4.3 Feature Aggregation

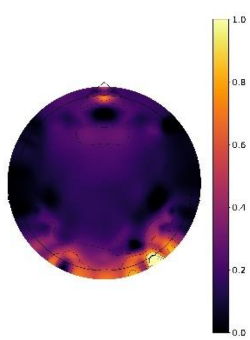
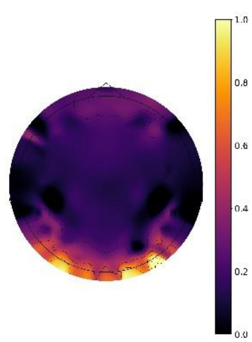
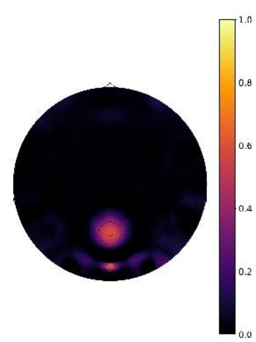
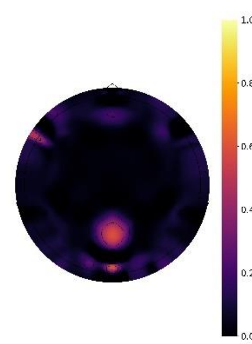
The extraction of these features is essential for subsequent analysis and classification tasks. Statistical features provide insights into the overall power distribution of the EEG signals, while frequency-domain features offer specific information about the brain's activity within different frequency bands. The spectral entropy measures add an additional layer of information regarding the complexity and variability of the EEG signals within these bands.

The PSD and ERP features are then combined such that for each subject’s EEG data, there are 128 EEG channels, each of which is characterized by 13 ERP features and 8 PSD features, forming the feature sets of resting state and dot probe task state EEG data.

## 10. OBSERVATIONS AND FINDINGS

In the preliminary phase of data exploration, we computed the band powers in the alpha and beta frequency bands for 4 categories of data, namely, (1) resting state MDD, (2) resting state healthy, (3) task state MDD, and (4) task state healthy.

Figure 8 illustrates the marked differences in band powers across these categories. These figures reveal the potential of using ERP and PSD features for distinguishing between classes from a machine-learning perspective, thus guiding the methodology adopted in this research.

| | | | |
|---|---|---|---|
|  |  |  |  |
| Topomap of resting state data for healthy subjects in alpha frequency range | Topomap of resting state data for MDD subjects in alpha frequency range | Topomap of resting state data for healthy subjects in beta frequency range | Topomap of resting state data for MDD subjects in beta frequency range |

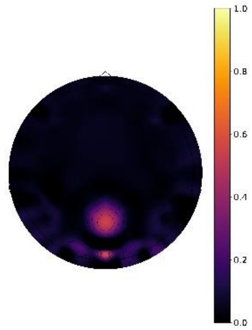
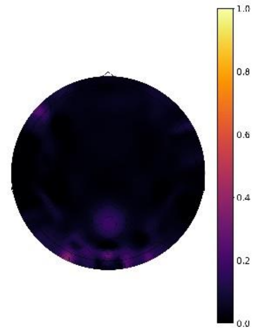
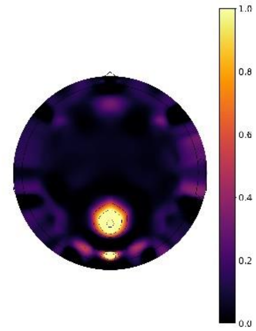
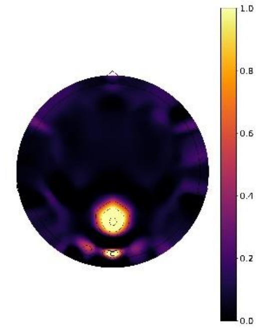
| | | | |
|---|---|---|---|
|  |  |  |  |
| Topomap of task state data for healthy subjects in alpha frequency range | Topomap of task state data for MDD subjects in alpha frequency range | Topomap of task state data for healthy subjects in beta frequency range | Topomap of task state data for MDD subjects in beta frequency range |

Fig 8: Band Powers of healthy and MDD subject data in Alpha and Beta frequency regions for both resting as well as task states.

These preliminary results demonstrate significant variations in neural activity patterns between healthy and MCI subjects, as well as between different states (resting, task-related). These differences provide a foundational basis for the methodologies designed for this project and emphasize the feasibility of using deep learning techniques to differentiate between these conditions.

# 11. CONCLUSIONS AND FUTURE WORK

## 11.1. Conclusions

This project aimed to enhance the identification and analysis of Major Depressive Disorder (MDD) using EEG and audio signals. The key findings and outcomes of this project are summarized below:

- **Data Preprocessing:** The preprocessing pipeline effectively managed the initial challenges of importing and preparing EEG data, utilizing robust referencing and filtering techniques to ensure high-quality data for analysis.

- **Data Cleaning:** Combining automatic continuous rejection with meticulous manual inspection ensured the removal of significant artifacts. This dual approach enhanced the reliability of the EEG data by addressing both spectrum thresholding and manual artifact identification.
- **Independent Component Analysis (ICA):** The integration of ICLabel for automatic component labeling and subsequent manual re-evaluation provided a precise classification of neural and non-neural components. This ensured that only relevant neural data were retained, minimizing the risk of including artifacts in the analysis.
- **Feature Extraction:** The extraction of PSD and ERP features captured essential frequency-domain and time-domain characteristics of the EEG signals. These features are critical for understanding the neural dynamics associated with MDD.

The results demonstrate that the methodologies employed in this project are effective in preprocessing, cleaning, and extracting meaningful features from EEG data. These steps are crucial for ensuring the accuracy and reliability of subsequent analyses and interpretations in the context of MDD research.

## 11.2. Learnings

Throughout the course of this project, several key learnings emerged:

- **Importance of Data Quality:** High-quality preprocessing and cleaning are essential for reliable EEG analysis. Combining automatic and manual techniques ensures a comprehensive approach to artifact removal.
- **Effective Use of ICA:** While ICA is a powerful tool for separating neural and non-neural components, its effectiveness is greatly enhanced by incorporating both automatic and manual review processes.
- **Robust Feature Extraction:** The extraction of diverse features, including PSD and ERP metrics, provides a comprehensive view of the neural dynamics, facilitating more accurate and insightful analyses.
- **Integration of Toolboxes:** Utilizing specialized toolboxes such as EEGLAB and MNE for different stages of the analysis pipeline streamlines the workflow and ensures the use of state-of-the-art methods.

- **Deep Learning Applications:** The robust feature set developed through this methodology is well-suited for deep learning applications, enabling the development of predictive models for MDD.

## 11.3. Future Work

While significant progress has been made in the EEG analysis component of the project, several avenues for future work remain, particularly in the exploration of advanced fusion techniques:

- **Fusion of EEG and Audio Signals:** Future work will focus on developing and implementing fusion techniques to combine EEG and audio signals. This multimodal approach is expected to provide a more comprehensive understanding of MDD by capturing both neural and auditory dynamics.
- **Advanced Machine Learning Models:** Leveraging the rich feature set, advanced machine learning models, such as deep learning and ensemble methods, will be explored to improve the accuracy and robustness of MDD detection and classification.
- **Real-time Analysis:** Developing real-time analysis pipelines for EEG and audio data can facilitate immediate interventions, which are particularly valuable in clinical settings.
- **Longitudinal Studies:** Conducting longitudinal studies with repeated measurements over time will help in understanding the progression of MDD and the impact of various treatments, providing deeper insights into the disorder.
- **Broader Clinical Applications:** Expanding the application of these methodologies to other neurological and psychiatric disorders can validate the generalizability and effectiveness of the approaches developed in this project.

By addressing these future directions, the project aims to build on the foundational work completed thus far, advancing the field of EEG and audio signal analysis for mental health research and clinical applications.

# BIBLIOGRAPHY


[1] *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, vol. 5, no. 5, Amer. Psychiatric Assoc., Washington, DC, USA, 2013.

[2] (Mar. 31, 2023). *Depressive Disorder (Depression)*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[3] S. Dattani, H. Ritchie, and M. Roser, 2021, ‘‘Mental health,’’ Our World in Data. [Online]. Available: https://ourworldindata.org/mental-health

[4] K. Kroenke, R. L. Spitzer, and J. B.W.Williams, ‘‘The PHQ-9: Validity of a brief depression severity measure,’’ *J. Gen. Internal Med.*, vol. 16, no. 9, pp. 606–613, Sep. 2001.

[5] F. M. Shah, F. Ahmed, S. K. Saha Joy, S. Ahmed, S. Sadek, R. Shil, and Md. H. Kabir, ‘‘Early depression detection from social network using deep learning techniques,’’ in *Proc. IEEE Region 10 Symp. (TENSYMP)*, Jun. 2020, pp. 823–826.

[6] K. A. Govindasamy and N. Palanichamy, ‘‘Depression detection using machine learning techniques on Twitter data,’’ in *Proc. 5th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, May 2021, pp. 960–966.

[7] B. Yalamanchili, N. S. Kota, M. S. Abbaraju, V. S. S. Nadella, and S. V. Alluri, ‘‘Real-time acoustic based depression detection using machine learning techniques,’’ in *Proc. Int. Conf. Emerg. Trends Inf. Technol. Eng. (ic-ETITE)*, Feb. 2020, pp. 1–6.

[8] J. Zhu, Z.Wang, T. Gong, S. Zeng, X. Li, B. Hu, J. Li, S. Sun, and L. Zhang, ‘‘An improved classification model for depression detection using EEG and eye tracking data,’’ *IEEE Trans. Nanobiosci.*, vol. 19, no. 3, pp. 527–537, Jul. 2020.

[9] L. Yang, D. Jiang, X. Xia, E. Pei, M. C. Oveneke, and H. Sahli, ‘‘Multimodal measurement of depression using deep learning models,’’ in *Proc. 7th Annu. Workshop Audio/Visual Emotion Challenge*, Oct. 2017, pp. 1–8.

[10] J. P. Gray, V. I. Müller, S. B. Eickhoff, and P. T. Fox, ‘‘Multimodal abnormalities of brain structure and function in major depressive disorder: A meta-analysis of neuroimaging studies,’’ *Amer. J. Psychiatry*, vol. 177, no. 5, pp. 422–434, May 2020.

[11] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Hyett, G. Parker, and M. Breakspear, ‘‘Multimodal depression detection: Fusion analysis of paralinguistic, head pose and eye gaze behaviors,’’ *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 478–490, Oct. 2018.