DEPRESSION DETECTION USING ELECTROENCEPHALOGRAM AND AUDIO MODALITIES

A PROJECT REPORT

Submitted by

IBRAHIM NAVAS S - 2021115042

MANI@ARUN VENKATESH P - 2021115060

THIRUMURUGAN K - 2021115117

ADNAN KHAN S - 2021115304

submitted to the faculty of

INFORMATION AND COMMUNICATION TECHNOLOGY

in partial fulfillment

for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY

COLLEGE OF ENGINEERING GUINDY

ANNA UNIVERSITY

CHENNAI 600 025

DECEMBER 2024

ANNA UNIVERSITY

CHENNAI - 600 025

BONAFIDE CERTIFICATE

Certified that this project report titled **“DEPRESSION DETECTION USING ELECTROENCEPHALOGRAM AND AUDIO MODALITIES”** is the bonafide work of **IBRAHIM NAVAS S (2021115042)**, **MANI@ARUN VENKATESH P (2021115060)**, **THIRUMURUGAN K(2021115117)**, **ADNAN KHAN S (2021115304)** who carried out project work under my supervision. Certified further that to the best of my knowledge and belief, the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on this or any other candidate.

PLACE:CHENNAI
DATE:

Dr. P. VARALAKSHMI
PROFESSOR
PROJECT GUIDE
DEPARTMENT OF IST, CEG
ANNA UNIVERSITY
CHENNAI 600025

COUNTERSIGNED

Dr. S. SWAMYNATHAN
HEAD OF THE DEPARTMENT
DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY
COLLEGE OF ENGINEERING GUINDY
ANNA UNIVERSITY
CHENNAI 600025

ABSTRACT

The early and precise detection of depression remains a critical challenge in mental health care, neuroscience, and public health, with significant implications for treatment and recovery. Despite advancements in technology and research, many individuals suffering from depression are diagnosed late or not at all, delaying access to essential interventions. This project aims to address this issue by developing an innovative, multimodal depression detection system. Leveraging diverse data sources such as Electroencephalogram (EEG) signals and audio data. The project integrates state-of-the-art machine learning techniques to enhance the accuracy and reliability of depression detection. The system is divided into three distinct yet interconnected modules. Module 1 focuses on EEG-based detection, utilizing a Comma Separated Values (CSV) dataset containing alpha, beta, gamma, delta, and theta band data across 19 prominent EEG channels. Advanced feature selection methods like ElasticNet and the XGBoost classifier are employed to predict one of 12 mental disorders with high accuracy. Module 2 targets EEG data stored in EDF format, transforming these signals into EEG graphs for image-based classification using a ResNet18 model. This module categorizes individuals as Healthy or Major Depressive Disorder (MDD) based on eye-opened, eye-closed, and task-based EEG recordings. Module 3 centers on audio-based depression detection, employing various sequential models like Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), etc., for spectrogram-based analysis of .wav files. Each module incorporates robust preprocessing pipelines and cutting-edge techniques, ensuring the reliability and generalizability of the models across diverse datasets. By combining these modalities, the project establishes a comprehensive framework for depression detection, offering a nuanced understanding of the disease's multifaceted nature.

.

ABSTRACT

ACKNOWLEDGEMENT

It is our privilege to express our deepest sense of gratitude and sincere thanks to **Dr. P VARALAKSHMI,** Professor, Project Guide, Department of Information Science and Technology, College of Engineering, Guindy, Anna University, for her constant supervision, encouragement, and support in our project work. We greatly appreciate the constructive advice and motivation that was given to help us advance our project in the right direction.

We are grateful to **Dr. S. SWAMYNATHAN,** Professor and Head, Department of Information Science and Technology, College of Engineering Guindy, Anna University for providing us with the opportunity and necessary resources to do this project.

We would also wish to express our deepest sense of gratitude to the Members of the Project Review Committee: **Dr. K. VANI,** Professor, **Dr. T. MALA,** Professor, **Dr. K. VIDYA,** Associate Professor, **Dr. T. J. VIJAYKUMAR,** Teaching Fellow, Department of Information Science and Technology, College of Engineering Guindy, Anna University, for their guidance and useful suggestions that were beneficial in helping us improve our project.

We also thank the faculty member and non teaching staff members of the Department of Information Science and Technology, Anna University, Chennai for their valuable support throughout the course of our project work.

IBRAHIM NAVAS S
MANI@ARUN VENKATESH P
THIRU MURUGAN K
ADNAN KHAN S

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AR	Augmented Reality
CNN	Convolutional Neural Network
Conv2DLSTM	Convolutional Long Short-Term Memory
EDF	European Data Format
EEG	Electroencephalogram
FPR	False Positive Rate
GPU	Graphics Processing Unit
GNN	Graph Neural Network
GUI	Graphical User Interface
KNN	K-Nearest Neighbors
LGBM	Light Gradient Boosting Machine
LSTM	Long Short-Term Memory
MFCC	Mel-Frequency Cepstral Coefficients
ML	Machine Learning
MNE	MNE Python Library for Processing EEG Data
MME	Mean Model Error
PPV	Positive Predictive Value
QEEG	Quantitative Electroencephalography
ResNet	Residual Neural Network
RNN	Recurrent Neural Network
SMOTE	Synthetic Minority Oversampling Technique
SWIN	Shifted Window Transformer
Transformer	
TPR	True Positive Rate
VR	Virtual Reality
XGB	Extreme Gradient Boosting

CHAPTER 1

INTRODUCTION

Depression is a debilitating mood disorder marked by persistent sadness, diminished interest in daily activities, and impaired functioning. It profoundly impacts individuals' emotional, cognitive, and physical health while imposing substantial burdens on families, communities, and healthcare systems worldwide.

1.1 OVERVIEW OF DEPRESSION

Depression is a chronic and complex mental health condition influenced by genetic, biological, environmental, and psychological factors. It manifests in various forms, including Major Depressive Disorder (MDD), dysthymia (persistent depressive disorder), postpartum depression, and Seasonal Affective Disorder (SAD). Symptoms of depression include persistent low mood, fatigue, changes in appetite and sleep patterns, difficulty concentrating, and recurrent thoughts of death or suicide, with the severity and duration of these symptoms disrupting an individual's ability to function effectively. According to the World Health Organization (WHO), depression affects approximately 280 million people worldwide, making it one of the leading causes of disability. The prevalence of depression is about 5% globally, with a higher incidence among women compared to men and alarming rates among adolescents. In high-income countries, the lifetime prevalence of major depressive disorder ranges between 15-20%, whereas underreporting remains a concern in low- and middle-income countries due to stigma and limited healthcare access. Depression is a significant risk factor for suicide, claiming nearly 700,000 lives annually and ranking as the fourth leading cause of death among individuals aged 15-29. It also exacerbates other chronic health conditions such as cardiovascular diseases, diabetes, and cancer, compounding its societal and economic impact. The combined direct and indirect costs of depression, including healthcare expenses, lost productivity, and absenteeism, are estimated to exceed one trillion dollars annually worldwide, emphasizing the urgent need for effective interventions and mental health care strategies.

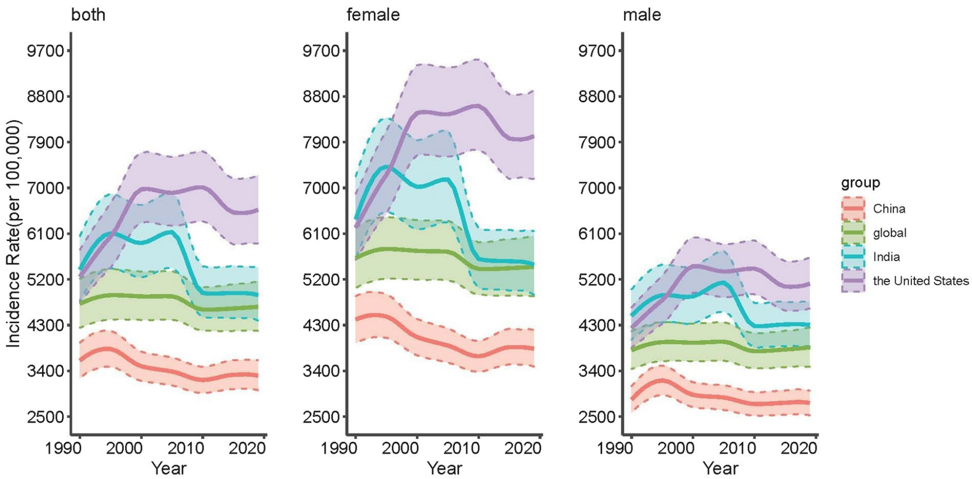

Figure 1.1: Depression Statistics through out the Years

1.2 DOMAIN APPLICATION AREAS

The proposed system has diverse applications across several critical domains. In healthcare and clinical settings, it can enable early intervention and personalized treatment through EEG and audio-based depression detection, providing objective insights into neurobiological markers that enhance diagnostic precision and effectively monitor patient progress. In workplace mental health, advanced screening tools can identify mental health risks early, reducing productivity losses while fostering supportive environments that address depression as both a humanitarian and strategic business concern. Educational institutions can benefit by identifying vulnerable student populations, enabling timely interventions, and enhancing academic performance, thereby addressing stressors and improving student retention. Furthermore, digital health platforms can leverage this technology to provide accessible, non-invasive mood tracking and mental health assessments through wearable devices and AI-driven applications, democratizing mental health management and empowering users with personal monitoring tools for proactive health interventions. Lastly, in research and development, the system can aid in discovering novel depression biomarkers and developing sophisticated diagnostic methodologies, fostering interdisciplinary collaborations in neuroscience, psychology, and engineering to advance the understanding of mood disorders and improve mental health technologies.

1.3 CHALLENGES

Access to Privileged Data: Securing specialized datasets for depression detection research involves navigating extensive institutional processes, including formal applications, confidentiality agreements, and verification delays. For example, repositories like the University of Southern California's collection often require weeks-long approval timelines. Researchers must develop clear research proposals, maintain transparent communication, and adhere to strict data usage protocols to overcome these hurdles effectively.

Ethical and Privacy Concerns: Handling sensitive human subject data, such as EEG recordings, raises critical ethical issues. To maintain privacy and trust, researchers must ensure data anonymization, comply with institutional review board protocols, and secure informed consent from participants. These considerations are vital for preserving the acceptability of research efforts.

Technical Challenges in Preprocessing: EEG data is often stored in formats like EDF, which require specialized preprocessing to make the data suitable for machine learning models. This involves advanced tools and interdisciplinary collaboration to address the complexity of data transformations, ensuring that the data is clean, standardized, and ready for analysis.

Dataset Variability and Standardization: Variations in recording conditions, such as task-based versus resting-state protocols, complicate data integration and analysis. Researchers employ standardization techniques, strategic segmentation, and ensemble learning approaches to manage these inconsistencies and ensure robust model performance across diverse datasets.

Scarcity of Public Datasets: The limited availability of diverse and comprehensive datasets restricts the generalizability of depression detection research. To address this challenge, researchers use strategies such as data augmentation, synthetic data generation, and collaborative partnerships, enhancing the representativeness and applicability of datasets to real-world scenarios.

Ensuring Dataset Relevance: Evaluating the relevance of datasets to specific research objectives is critical for maximizing scientific value. This involves thorough review processes, filtering datasets based on inclusion criteria, and employing adaptable machine-learning architectures that align with the research goals while maintaining methodological rigor.

1.4 MOTIVATION

Depression affects more than 264 million individuals worldwide, leading to significant societal and economic burdens, including reduced productivity and widespread personal suffering. These alarming statistics highlight the urgency of developing advanced detection methods that transcend cultural, geographical, and socioeconomic barriers to address this pressing global health crisis. Conventional diagnostic methods often depend on subjective self-reports, which can introduce patient bias and accessibility challenges. These traditional approaches lack the capability to detect subtle neurophysiological markers, delaying early identification and leaving many individuals, particularly those without access to healthcare, undiagnosed or misdiagnosed. Emerging technologies, such as EEG analysis and machine learning, offer transformative potential for mental health diagnostics. By providing objective insights into neurobiological markers, these data-driven methodologies promise to improve detection accuracy, moving beyond subjective assessments to create precise and reliable diagnostic tools. Furthermore, the integration of advanced techniques enables earlier and more personalized detection of depression, allowing for a nuanced understanding of individual mental health needs. These innovations pave the way for tailored treatment strategies, addressing the unique requirements of each patient to improve outcomes and foster better mental health care.

1.5 RESEARCH OBJECTIVES

Develop a Multimodal Depression Detection Framework: Integrate EEG and audio data using parallel architectures and modality-specific models to enhance depression detection accuracy and handle diverse data characteristics.

Characterize EEG Biomarkers for Depression: Extract and validate neurological markers of depression using advanced artifact removal, frequency band analysis, and hybrid deep learning architectures.

Analyze Acoustic Features for Vocal Biomarkers: Identify depression-specific vocal characteristics through MFCCs, prosodic feature analysis, and spectrogram-based deep learning, enabling non-invasive diagnostic methods.

Mitigate Data Challenges: Employ data augmentation, SMOTE, and adaptive class balancing to address dataset limitations, reduce bias, and ensure model robustness and generalizability.

1.6 PROBLEM STATEMENT

The increasing complexity and diversity of patient data, there is a critical need for innovative, data-driven solutions to enhance depression detection and classification. This project addresses these challenges by leveraging multimodal data sources, including Electroencephalogram (EEG) signals, audio features, and clinical records, to develop a comprehensive and robust system for depression detection. By integrating advanced machine learning techniques with domain-specific preprocessing workflows, the project aims to overcome existing diagnostic limitations, facilitate early detection, and provide actionable insights for personalized mental health care. This approach not only seeks to improve diagnostic accuracy but also aspires to bridge the gap between mental health research and clinical practice, ultimately reducing the global burden of depression.

1.7 PROPOSED SOLUTIONS

To address the limitations of current depression detection methods, this project proposes a multimodal and machine learning-driven approach for accurate and early diagnosis of depression. The solution is divided into three distinct modules, each leveraging different data types to provide comprehensive insights into an individual's mental health. The first module focuses on EEG-based analysis, utilizing frequency bands and coherence values from EEG signals to predict 12 mental disorders, including depression. By employing ElasticNet for feature selection and XGBoost for classification, this module provides robust predictions from highly dimensional data, ensuring efficient handling of noise and feature redundancy.

The second module extends the approach to image-based analysis by converting EEG data into graphical representations, such as spectrograms, for classification using deep learning models like ResNet18. This enhances the diagnostic process by visually interpreting brain signal patterns associated with Major Depressive Disorder (MDD). The third module focuses on audio-based depression detection, processing speech signals to remove noise and silence, extracting key features, and training a ResNet50 model on spectrograms generated from these signals. Together, these modules form an integrated pipeline that addresses the challenge of depression detection from multiple perspectives. The solution also incorporates advanced preprocessing, data augmentation, and class balancing techniques to ensure accuracy, scalability, and generalizability, making it a reliable tool for real-world clinical and research applications.

1.8 ORGANIZATION OF THE REPORT

This report is organized into 6 chapters, describing each part of the project with detailed illustrations and system design diagrams.

CHAPTER 2: Literature Review reviews existing research, studies, and relevant literature related to depression detection. It discusses the background of depression diagnosis, theories on mental health analysis, and methodologies employed in previous studies to detect depression.

CHAPTER 3: System Architecture which gives insight about the architecture of the entire system and followed by that gives the brief information about the tools and technologies that are used in the project's modules. This portion will provide knowledge about the front end and back end.

CHAPTER 4: System Design describes the design of the proposed solution for depression detection. It explains the architecture, data preprocessing techniques, model pipelines, and algorithms used in each of the three modules: csv-based, edf-based analysis of brain signals, and audio-based analysis of speech signals.

CHAPTER 5: Implementation provides details about how the project was implemented. It elaborates on the development, the preprocessing pipelines, and model training processes.Result and Analysis presents the results of the project, including the performance metrics of the models developed in each module. It analyzes the outcomes, compares them with the expectations, and discusses challenges such as data imbalance, noise handling, and computational limitations faced during implementation.

CHAPTER 6: Conclusion and Future Work summarizes the findings and conclusions drawn from the project. It discusses the significance of the proposed multimodal depression detection system, its potential clinical applications, and future enhancements, such as refining models for higher accuracy and generalizability.

CHAPTER 2

LITERATURE SURVEY

This chapter deals with the existing work carried out in the field of Machine Learning and Deep Learning for Depression Detection. It provides an overview of the challenges and advancements in developing models, focusing on methodologies and their performance.

2.1 PIPELINING PROCESSES AND FEATURES IN MULTICLASS CLASSIFICATION

Park *et al.*[6] introduced a robust pipeline for classifying multiple mental disorders using EEG data. Their pipeline consists of key stages such as EEG signal preprocessing, feature extraction, and classification using machine learning models. The preprocessing stage includes filtering the EEG signals and segmenting them into different frequency bands (alpha, beta, gamma, delta, theta). Feature extraction is performed using statistical and frequency-based methods, ensuring that the most relevant EEG channels are selected. The classification stage employs XGBoost, a popular boosting model, to predict mental disorders based on extracted features. This pipeline significantly improves classification accuracy by combining signal processing with machine learning models. The modularity of the pipeline also allows for easy integration of new data sources or adjustments based on disorder-specific requirements. Ahmed *et al.*[7] similarly emphasized the importance of signal preprocessing and effective classification, highlighting advancements in using deep learning models for improved accuracy and generalizability of predictions.

2.2 EXTRACTION OF CHANNEL-WISE FEATURES USING POWER SPECTRAL DENSITY

PSD represents the distribution of signal power over different frequency bands, which provides deep insights into the brain's electrical activity. As highlighted in Park *et al.*[6] and Ahmed *et al.*[7], the power in specific frequency bands delta, theta, alpha, beta, and gamma holds substantial clinical relevance. Variations in these bands can be indicators of psychiatric disorders like MDD. For instance, studies have shown that depression correlates with reduced alpha activity and increased theta activity. These features, derived through Fast Fourier Transform (FFT), are fundamental in distinguishing depression from other states like anxiety or healthy control conditions. Novel methods focus on extracting features from individual EEG channels, which are then integrated to form a comprehensive feature set. Liu *et al.*[2] explored local oscillations from different brain regions, providing deeper insights into localized brain activity related to depression. For example, frontal lobe activity (related to cognitive functions) and theta waves are often observed as disrupted in depressive states. Aggregating power values across multiple frequency bands contributes to creating a global picture of brain functioning and enables the model to capture holistic neural signatures of depression.

2.3 FUSING MULTI-STATE EEG FOR MDD CLASSIFICATION

Yang *et al.*[17] introduced a method to integrate EEG signals from eyes-open (EO) and eyes-closed (EC) conditions, aiming to enhance the classification accuracy for Major Depressive Disorder (MDD). Liu *et al.*[2] highlighted the distinct neural dynamics exhibited in EO and EC states, noting their potential to provide complementary diagnostic information. By combining features derived from both conditions, the integrated approach leveraged the unique state-dependent variations in brain activity, as demonstrated in prior findings by Mumtaz *et al.*[16]. The proposed fusion framework, compared to single-state analyses, significantly improved classification performance, corroborating evidence from other studies on the advantages of multi-state EEG integration. Furthermore, this method underscored the potential of using EO and EC EEG fusion to develop robust, automated diagnostic systems for mental health disorders, aligning with advancements in psychiatric diagnostics discussed by Bai *et al.* [18]. This multifaceted approach opens new avenues for exploring state-dependent neural biomarkers in the context of clinical applications.

2.4 SIGNAL ANALYSIS FOR ENHANCED CLASSIFICATION AND PREDICTION

State-based EEG analysis, focusing on Eye Open (EO), Eye Closed (EC), and Task states, offers valuable insights for depression detection. Liu *et al.*[2] compared EEG signals during EO and EC states in individuals with first-episode depression, revealing significant differences in power spectra that may serve as reliable biomarkers. Increased theta and decreased alpha power in the EO condition were linked to cognitive and emotional dysregulation in depression. Building on this, Yang *et al.*[17] fused EO and EC EEG data, achieving higher classification accuracy for MDD by leveraging state-dependent brain activity. Bai *et al.*[18] emphasized task-related EEG analysis, showing that distinct brainwave patterns elicited during cognitive tasks complemented resting state analyses, enhancing depression detection. Further advancements in state-based EEG methodologies were proposed by Mumtaz *et al.*[16], which combined EEG and fMRI signals using tensor decomposition to improve neural biomarker precision. These findings underscore the critical role of analyzing mental states like EO, EC, and Task to extract nuanced brain activity patterns, enabling more accurate and reliable diagnostic tools for MDD.

2.5 SPECTROGRAM-BASED FEATURE EXTRACTION

Spectrogram analysis has become a cornerstone in audio-based mental health research, offering comprehensive insights into speech patterns and emotional dynamics. Gao *et al.*[3] utilized Short-Time Fourier Transform (STFT) and Mel-frequency Cepstral Coefficients (MFCCs) to extract temporal and frequency features, highlighting depression-related speech characteristics through graph-based neural networks. Sardari *et al.*[5] emphasized Mel-spectrograms alongside spectral roll-off, zero-crossing rate, and flatness to capture subtle speech irregularities, employing Support Vector Machines (SVM) and Random Forest classifiers for robust prediction. Zhang *et al.*[26] explored hierarchical feature learning using log-Mel spectrograms integrated with convolutional neural networks (CNNs), focusing on pitch shifts and energy contours to detect anxiety and depression markers. Wang *et al.*[4] leveraged chroma spectrograms in conjunction with Recurrent Neural Networks (RNNs) to analyze sequential emotional transitions, effectively addressing variability in speech delivery. Moreover, Zhang *et al.*[26] introduced Constant-Q Transform (CQT) spectrograms for fine-grained frequency resolution and applied hybrid CNN-LSTM architectures to model temporal dependencies. These methodologies collectively underscore the adaptability of spectrogram-based approaches in uncovering biomarkers for mental health conditions, though addressing generalization challenges for diverse populations remains an open avenue for further research.

2.6 AUGMENTATION TECHNIQUES FOR AUDIO

Audio augmentation techniques play a critical role in enhancing model generalization and robustness in mental health diagnostics. Sardari *et al.*[5] implemented time-stretching, pitch-shifting, and random noise injection to diversify speech datasets, significantly improving classification accuracy for anxiety and depression detection. Sardari *et al.*[5] also introduced SpecAugment, a method that masks frequency bands and time segments in spectrograms, optimizing deep learning models to handle variability in speech patterns. Zhang *et al.*[26] employed random cropping and rotation in waveforms to simulate real-world distortions, effectively improving model resilience against noisy environments. Wang *et al.*[4] leveraged synthetic data generation through Generative Adversarial Networks (GANs) to address the underrepresentation of minority speech classes, particularly in detecting schizophrenia-related vocal markers. These augmentation strategies collectively demonstrate the potential to mitigate data imbalances, improve generalization across populations, and ensure robust performance, though balancing synthetic and natural data for clinical reliability remains a key challenge.

2.7 ATTENTION MECHANISMS FOR AUDIO MODELS

Attention mechanisms have revolutionized audio-based mental health diagnostics by enabling models to focus on critical speech features. Wang *et al.*[4] integrated self-attention layers in RNNs to capture temporal dependencies in speech patterns, enhancing depression classification accuracy. Zhang *et al.*[26] utilized multi-head attention in transformer models for schizophrenia detection, effectively emphasizing subtle vocal and linguistic cues critical for diagnosis. These methodologies underscore the potential of attention mechanisms in extracting meaningful patterns from audio data, ultimately improving the precision and reliability of mental health assessments through speech analysis.

2.8 CONCLUSION FROM LITERATURE SURVEY

EEG-based methodologies have achieved high accuracies, with techniques like hybrid models and graph neural networks reaching up to 96.8%, and advanced signal processing methods such as Detrended Fluctuation Analysis (DFA) and machine learning achieving 91% accuracy by fusing eyes-open and eyes-closed EEG data. Neural oscillation studies provide insights into the cognitive and emotional states underlying mental disorders. Speech and audio-based detection methods leverage acoustic patterns, prosody, and linguistic analysis, with ensemble models achieving 94.1% accuracy, though challenges remain in capturing linguistic and cultural diversity, necessitating improved multimodal integration and cross-population validation. Biomarker-driven approaches focus on precision psychiatry, predicting treatment responses and enabling personalized interventions by integrating objective biomarkers with standardized frameworks like ICD-10, DSM-5, and precision medicine initiatives like the NIMH RDoC project. These efforts aim to bridge theoretical advancements with clinical workflows, advancing real-world applicability.

CHAPTER 3

SYSTEM ARCHITECTURE

This chapter presents the system architecture of the depression detection framework, outlining the preprocessing of multimodal data (EEG, audio, and images), the implementation of advanced machine learning and deep learning techniques, and the integration of various modules to achieve accurate depression diagnosis.

3.1 ARCHITECTURE OF THE SYSTEM

The depression detection system integrates multiple advanced components to process multimodal data (EEG, audio, and images) and make accurate predictions regarding depression. The system processes raw data through a series of steps, including preprocessing, feature selection, model training, and evaluation. Each component plays a vital role in ensuring accurate predictions for depression diagnosis.

PREPROCESSING: The preprocessing module is responsible for cleaning and preparing the raw data for analysis. This involves handling missing values, standardizing data, and converting data formats as required for each modality. It also includes removing background noise from audio data and resizing and normalizing image data. This step ensures that the data is in a consistent format, ready for subsequent analysis.

FEATURE SELECTION: Feature selection aims to improve the model's efficiency and accuracy by identifying the most important features from the raw data. By focusing on the most relevant information, this step

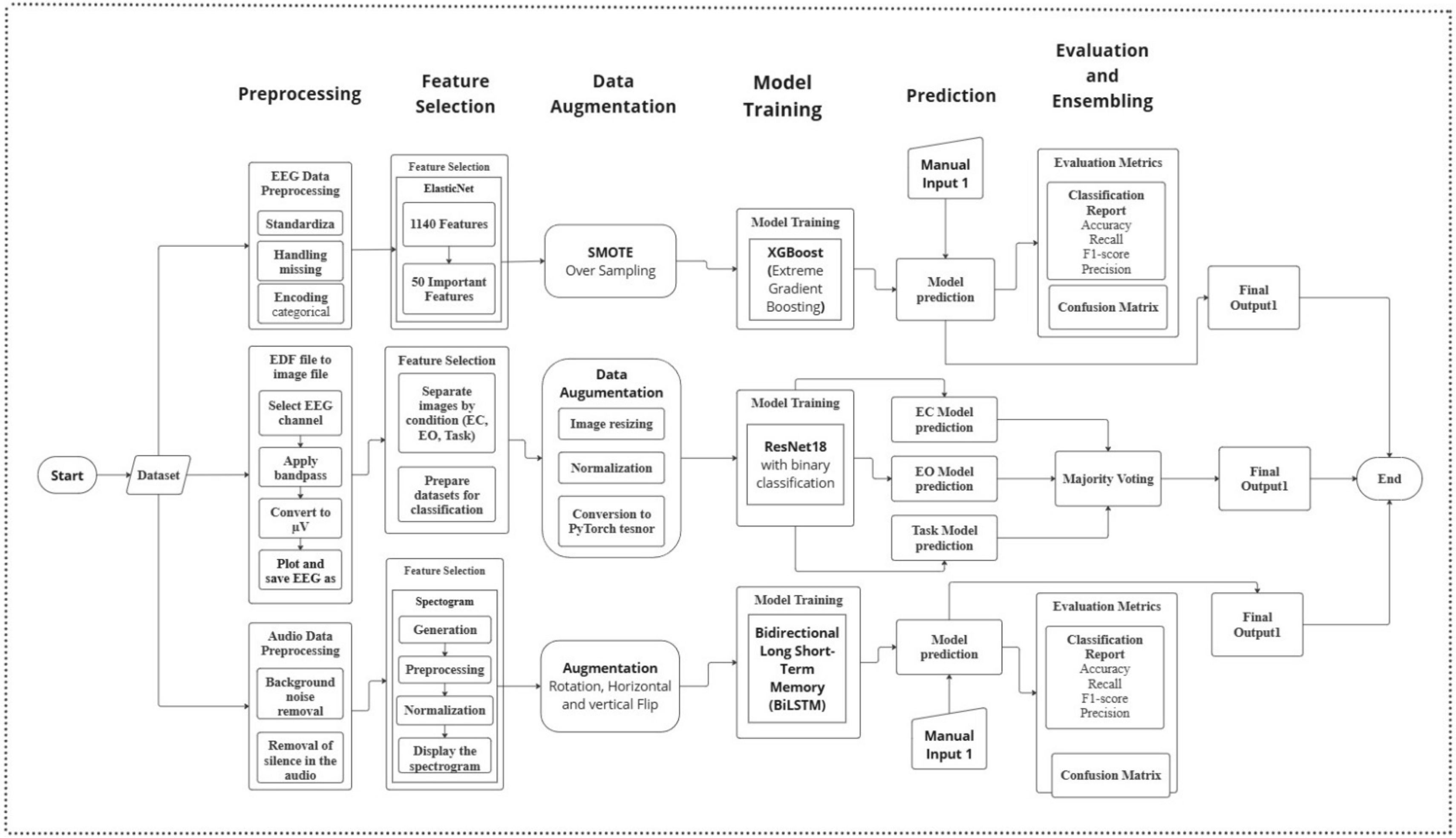

Figure 3.1: Depression Detection System Architecture

reduces the dimensionality of the data, enabling faster model training and better performance in depression detection.

DATA AUGMENTATION: To address issues like data imbalance and overfitting, data augmentation techniques are applied. These techniques increase the diversity of the training dataset by artificially expanding it with transformations such as rotations, noise addition, and pitch shifts. This helps improve model robustness and ensures better generalization on unseen data.

MODEL TRAINING: During the model training phase, the preprocessed data is used to train machine learning and deep learning models for classification tasks. These models are trained on each modality of data (EEG, audio, and images) to accurately detect depression. The models are fine-tuned to achieve optimal performance on the data, ensuring accurate predictions.

PREDICTION: The prediction module uses the trained models to classify an individual's mental health status based on the processed data. It combines predictions from EEG, audio, and image-based models to generate a final diagnosis, ensuring a comprehensive evaluation of the individual's mental health.

EVALUATION: Once predictions are made, the system evaluates model performance using metrics such as accuracy, recall, precision, and F1-score. These metrics help assess the reliability of the models. An ensembling technique is also applied to combine predictions from different models, ensuring a robust and reliable final decision.

3.2 TOOLS AND TECHNOLOGIES USED

The depression detection system utilizes a combination of technologies to streamline its development and enhance functionality. **HTML** structures the content on the web pages, ensuring an organized layout for the user interface. **CSS** is used to style the web pages, providing a responsive and visually appealing design that adapts across devices. **JavaScript** adds interactivity to the application, enabling dynamic updates and smooth navigation between system modules. **Python** serves as the backbone of the system, handling data preprocessing, feature selection, and model training. **Flask**, a lightweight Python web framework, facilitates backend development, enabling easy interaction with models and prediction delivery through the web interface. **Git** and **GitHub** ensure effective version control, collaborative development, and stable management of the project's codebase.

3.2.1 Deep Learning Frameworks

The depression detection system utilizes several deep learning frameworks to enhance model performance and efficiency. **PyTorch** is used for implementing models such as ResNet18 and ResNet50 for classifying EEG images and audio spectrograms. It provides flexibility and GPU acceleration, enabling faster training of deep learning models. **Torchvision** complements PyTorch by facilitating image preprocessing tasks like normalization and resizing, ensuring that the data is compatible with the deep learning models. **TensorFlow** serves as a comprehensive open-source deep learning framework, offering a rich ecosystem for developing both simple and complex neural networks with efficient GPU support. **Keras**, which runs on top of TensorFlow, provides a high-level API for building deep learning models. It simplifies the process of model creation, training, and evaluation with its user-friendly interface. The prominent deep learning frameworks lays the foundation for the project.

3.2.2 Data Preprocessing Libraries

The depression detection system leverages various data preprocessing libraries to handle different data modalities effectively. **Scikit-learn** provides utilities for scaling, normalization, encoding categorical variables, and splitting datasets into training and testing sets. It also supports feature selection techniques like ElasticNet. **EDFlib** or **MNE** are used to process EEG data stored in EDF format, including tasks like channel selection and bandpass filtering, which are essential for EEG-specific preprocessing. **Pandas** is utilized for handling missing values, encoding categorical variables, and cleaning datasets, simplifying tasks such as dropping unnecessary columns. **NumPy** aids in efficient numerical computations, working with multi-dimensional arrays and matrices, which are vital for feature extraction and transformation. **LibROSA** is employed to process audio data, removing background noise and silence, as well as extracting features like spectrograms to ensure clean, structured audio data. Finally, **Wavelet Transform Libraries** such as **PyWavelets** are used to apply wavelet transforms on EEG signals, capturing both time and frequency information for valuable insights in depression detection.

3.2.3 Plotting Libraries

Matplotlib is used to visualize EEG signals, spectrograms, and model performance metrics like confusion matrices, helping in the clear representation of data insights and model results. **Seaborn** enhances the visualizations by offering aesthetically pleasing plots, aiding in exploring data distributions and evaluating model performance. Seaborn comes with libraries that provide various plots to gain deeper insights into the data and the results of the model.

3.2.4 Evaluation Metrics

Accuracy measures the proportion of correctly classified instances out of all instances in the dataset. It is the most straightforward metric for classification tasks. **Precision**, also known as Positive Predictive Value (PPV), measures the proportion of positive predictions that are actually correct, providing insight into how many of the instances predicted as positive are truly positive. **Recall**, or Sensitivity, measures the proportion of actual positives correctly identified by the model, indicating how many of the actual positive instances were correctly predicted. **F1-Score** is the harmonic mean of precision and recall, offering a balanced measure that accounts for both false positives and false negatives, making it particularly useful in situations of class imbalance. **Confusion Matrix** is a summary table that assesses model performance, showing counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), making it easier to visualize classification results. **AUC-ROC Curve**, which stands for Area Under the Receiver Operating Characteristic Curve, is a graphical representation of model performance, plotting the True Positive Rate (Recall) against the False Positive Rate (FPR) at various threshold settings. **Log Loss** measures the performance of a classification model by evaluating the accuracy of predicted probabilities, comparing predicted probability distributions to the actual labels.

CHAPTER 4

SYSTEM DESIGN

In this chapter, we will discuss the system design for the Depression Detection project along with the algorithms implemented in each of the modules.

4.1 QUANTITATIVE EEG CHANNEL DATA (QEEG)

4.1.1 Data Description

The input for this module consists of a .csv file containing 1140 EEG features, including alpha, beta, gamma, delta, and theta waves across 19 EEG channels. Additionally, demographic and categorical features, such as gender, age, and disorders, are included.

4.1.2 Models used for Classification

1. XGBoost (Extreme Gradient Boosting)

XGBoost is a highly efficient implementation of gradient boosting algorithms designed for speed and performance. The overall working model architecture of XGBoost is given in Figure 4.1. It builds an ensemble of decision trees in a sequential manner, optimizing a differentiable loss function. XGBoost employs advanced features like regularization, sparsity awareness, and parallel processing, making it one of the most popular algorithms for structured data tasks.

Strengths: XGBoost excels in handling large datasets with high-dimensional features. It is robust against overfitting due to built-in regularization (L1 and L2), and its distributed computing support enables efficient scalability.

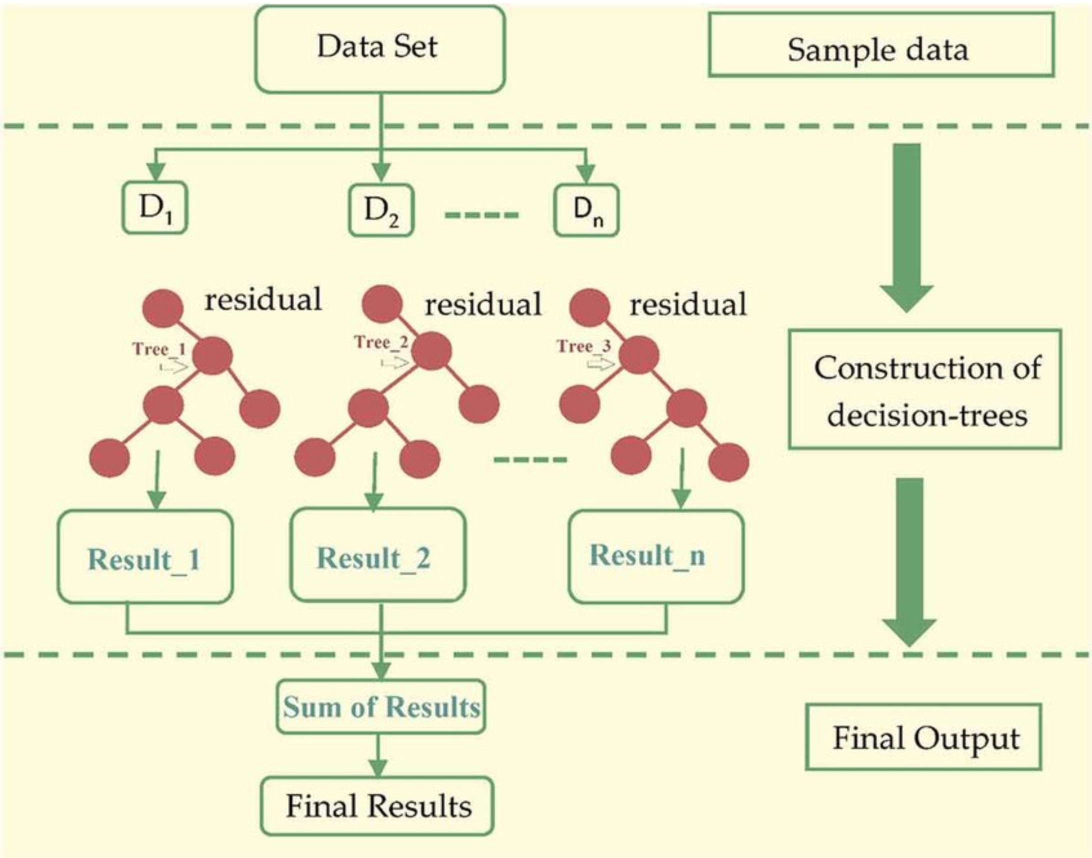

Figure 4.1: XGBoost Architecture

2. LGBM Classifier (Light Gradient Boosting Machine)

LightGBM is a fast, distributed, and efficient gradient-boosting framework that uses decision tree algorithms. It focuses on leaf-wise tree growth instead of level-wise growth, making it highly efficient for large datasets.

Strengths: LGBM is well-suited for large-scale datasets and excels in handling categorical features. It is faster and more memory-efficient compared to traditional gradient boosting methods.

3. Bi-LSTM (Bidirectional Long Short-Term Memory)

Bi-LSTM extends the traditional LSTM model by adding a second layer that processes input sequences in reverse. This allows the model to capture both past (backward) and future (forward) context information, making it powerful for sequence data.

Strengths: Bi-LSTM is excellent for tasks requiring an understanding of sequential dependencies, such as natural language processing (NLP) and time-series data.

4. RNN (Recurrent Neural Network)

RNNs are a class of neural networks designed to process sequential data by maintaining a hidden state that captures information about previous elements in the sequence.

Strengths: RNNs are ideal for time-series and sequence prediction tasks, where contextual information from earlier inputs is necessary.

5. Gradient Booster (Gradient Boosting Machines)

Gradient Boosting Machines are an ensemble method that builds models sequentially by optimizing a differentiable loss function. Each new model corrects the errors of the previous ones.

Strengths: Gradient boosting is highly flexible and effective for both regression and classification tasks. It handles complex data distributions well and is robust against overfitting with appropriate hyperparameter tuning.

Algorithm 4.1 Quantitative EEG-based Multi-class Mental Disorder Detection

Input: EEG CSV dataset (features and target labels)
Output: Predicted disorder label (one of 12 classes)
Step 1: Data Preprocessing
Load dataset and fill missing values with column means:

$$\text{data} \leftarrow \text{pd.read\_csv().fillna(mean)}$$

Apply ordinal encoding to categorical features:

$$\text{data} \leftarrow \text{OrdinalEncoder().fit\_transform()}$$

Drop unnecessary columns and split into X (features) and y (target).

Step 2: Train-Test Split
Split data into 80%-20% for training and testing.

Step 3: Feature Selection and Scaling
Standardize features and use ElasticNet for feature selection:

$$\text{ElasticNet}(\alpha = 0.1, l1\_ratio = 0.5)$$

Step 4: Handle Class Imbalance
Balance classes using SMOTE:

$$X, y = \text{SMOTE().fit\_resample}(X\_train, y\_train)$$

Step 5: Model Training and Saving
Train XGBoost within a pipeline:

$$\text{Pipeline}(\{\text{StandardScaler}, \text{'feature\_sel'}, \text{ElasticNet}, \text{XGBClassifier()}\})$$

Save pipeline using joblib:

$$\text{joblib.dump(pipeline, 'xgb\_pipeline.pkl')}$$

Step 6: Prediction and Mapping
Predict on test data and map predictions to class names:

$$\hat{y} = \text{pipeline.predict}(X\_test)$$

Output corresponding disorder names:

$$\text{class\_names}[\hat{y}]$$

4.1.3 Computational Complexity Analysis

The preprocessing phase demonstrates linear complexity as its performance is directly proportional to the size of the dataset, making it highly scalable for large datasets. Following this, the ElasticNet feature selection method introduces polynomial complexity, which is primarily influenced by the number of features and data points in the dataset. This step is pivotal for identifying the most relevant features while maintaining computational efficiency. Subsequently, the XGBoost algorithm contributes additional polynomial complexity, especially during the training phase. The iterative boosting mechanism of XGBoost, which refines weak learners through multiple iterations, enhances model accuracy but increases computational demands. Despite its intensive nature, XGBoost is recognized for its effectiveness in handling complex classification tasks, making it an integral component of the pipeline.

4.2 EDF SIGNAL BASED DETECTION OF HEALTHY AND MDD INDIVIDUALS

4.2.1 Data Description

EEG data in EDF format, categorized into three conditions: Eye Opened (EO), Eye Closed (EC), and Task.

4.2.2 Models used for Classification

1. ResNet18 (Residual Networks)

ResNet18 is a lightweight variant of the ResNet family, designed with 18 layers as shown in Figure 4.2. It introduces residual connections to mitigate the vanishing gradient problem, ensuring stable training in deep

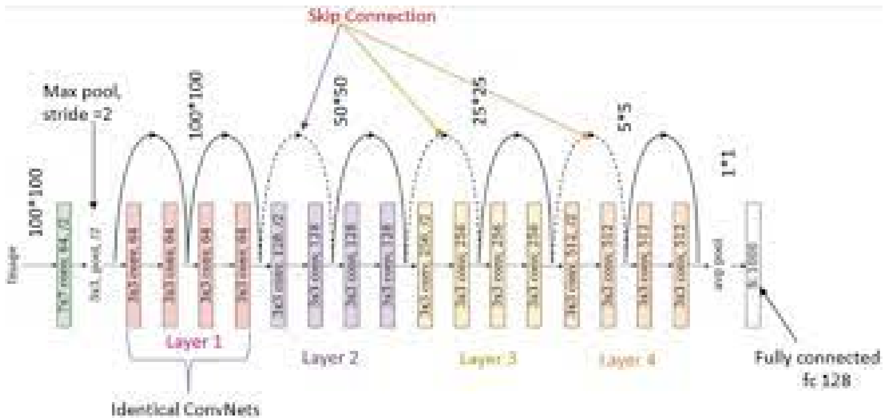

Figure 4.2: Resnet-18 Architecture

networks. Residual connections allow gradients to ”shortcut” across layers, improving learning efficiency and reducing training time.

Strengths: ResNet18 is well-suited for tasks requiring efficient feature extraction and classification with moderate computational resources. It provides strong generalization performance and is ideal for smaller to medium-scale image datasets.

2. DenseNet (Densely Connected Convolutional Networks)

DenseNet introduces dense connections between all layers, ensuring maximum feature reuse by connecting each layer to every other layer. This eliminates redundancy, reduces overfitting, and promotes efficient learning.

Strengths: DenseNet is highly parameter-efficient and effective for tasks requiring high feature extraction while keeping computational requirements low. It achieves excellent performance on image recognition tasks.

3. EfficientNet (Efficient Neural Networks)

EfficientNet introduces a compound scaling method that uniformly scales depth, width, and resolution of the network, optimizing accuracy and computational efficiency. This ensures a balance between performance and resource usage.

Strengths: EfficientNet achieves state-of-the-art performance in image classification tasks while requiring fewer parameters and FLOPs compared to other models. It is ideal for applications where computational resources are limited.

4. Swin Transformer (Shifted Window Transformer)

Swin Transformer is a hierarchical vision transformer that processes image patches using shifted windowing mechanisms. It excels in capturing both local and global information while maintaining computational efficiency.

Strengths: Swin Transformer is highly versatile and efficient, making it suitable for image recognition, object detection, and segmentation tasks. It outperforms traditional CNNs in scalability and accuracy on large-scale datasets.

4.2.3 System Complexity Overview

The preprocessing phase of the system demonstrates linear complexity, as its performance is directly proportional to the size of the dataset and the number of EEG channels being processed. This ensures efficient handling of large EDF files, making the system scalable for extensive datasets. On the other hand, the training phase involving the ResNet18 model introduces

Algorithm 4.2 EEG-EDF based Depression Detection

Input: Raw EEG data in EDF format (Healthy and MDD subjects)
Output: Predicted class label (Healthy or MDD)
Step 1: Data Preprocessing
Load EEG EDF files using MNE library:

$$\text{raw\_eeg} \leftarrow \text{mne.io.read\_raw\_edf()}$$

Retain EEG channels and filter signals to $0.5 - 45\,\text{Hz}$:

$$\text{raw\_eeg} \leftarrow \text{raw\_eeg.filter}(0.5, 45)$$

Convert signal amplitudes from $\text{V} \rightarrow \mu\text{V}$.
Step 2: Feature Extraction
Segment EEG data based on states (EC, EO, Task) and extract time-series signals.
Generate spectrograms for each segment using Short-Time Fourier Transform (STFT):

$$\text{spectrogram} \leftarrow \text{STFT(raw\_eeg)}$$

Save spectrograms as PNG images
Step 3: Dataset Preparation
Load spectrogram images and create a dataset for binary classification.
Split dataset into 80% training and 20% testing:

$$\text{train\_set}, \text{test\_set} \leftarrow \text{train\_test\_split(data, test\_size=0.2)}$$

Step 4: Model Selection and Training
Fine-tune ResNet18 for binary classification:

$$\text{model} \leftarrow \text{ResNet18(pretrained=True)}$$

Replace the final layer with a binary classification head.
Step 5: Training Process
Define data augmentation and normalization transformations:

$$\text{transforms} \leftarrow \{\text{Resize, Normalize}\}$$

Train model using CrossEntropyLoss and Adam optimizer:

$$\text{loss} \leftarrow \text{CrossEntropyLoss()}, \quad \text{optimizer} \leftarrow \text{Adam(model.parameters())}$$

Train for 6 epochs with GPU acceleration, if available.
Step 6: Model Testing and Evaluation
Evaluate the model on the test set and compute accuracy and plot confusion matrix for classification results.
Use the trained model to predict the class (Healthy or MDD).

high computational complexity. The deep learning architecture of ResNet18 demands significant computational resources due to its multiple layers and intricate operations. Despite the intensive nature of this process, it results in highly accurate classification outcomes, making it a critical component of the overall system for effective depression detection. .

4.3 AUDIO BASED DEPRESSION DETECTION

4.3.1 Data Description

Audio data in .wav format, containing speech samples from individuals for depression detection.

4.3.2 Models used for Classification

1. Bi-LSTM (Bidirectional Long Short-Term Memory)

Bi-LSTM extends the LSTM architecture shown in Figure 4.3 by processing input sequences in both forward and backward directions, enabling the model to capture past and future context. This makes it highly effective for tasks involving sequential data.

Strengths: Well-suited for text, speech, and time-series data. It excels at handling long-term dependencies and context-sensitive tasks.

2. ConvLSTM2D (Convolutional LSTM)

ConvLSTM2D combines convolutional layers with LSTMs to handle spatiotemporal data, making it suitable for tasks such as video analysis and weather prediction. It captures spatial patterns and temporal dependencies simultaneously.

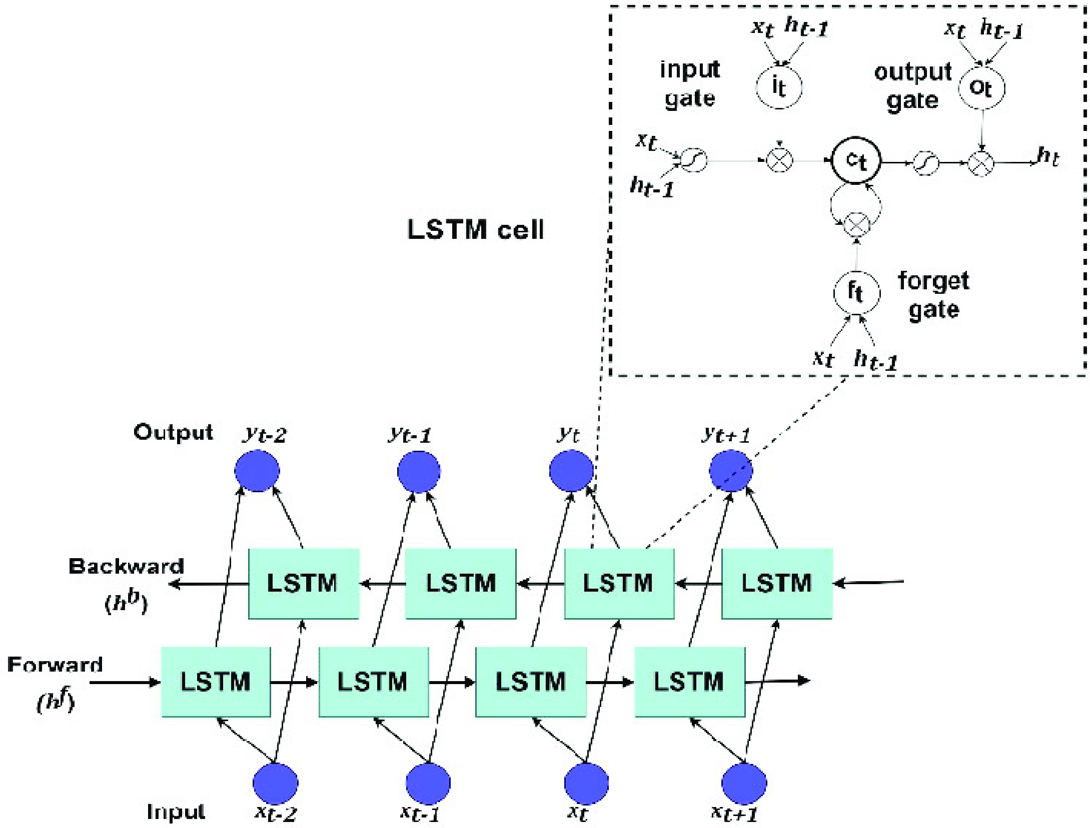

Figure 4.3: Bi-LSTM Architecture

Strengths: Efficient for applications requiring spatiotemporal feature extraction, such as video classification or medical imaging.

3. LSTM (Long Short-Term Memory)

LSTM is a type of recurrent neural network (RNN) designed to overcome the vanishing gradient problem. It uses memory cells to retain important information over long time intervals.

Strengths: Effective for time-series prediction, language modeling, and speech recognition. Handles long-term dependencies better than standard RNNs.

4. Dense Layered Sequential Model

This model consists of densely connected layers arranged in a sequential manner. Each neuron is connected to every neuron in the preceding layer, enabling feature extraction and learning.

Strengths: Flexible and straightforward to implement for classification and regression tasks. It can model nonlinear relationships in the data.

5. XGBoost Classifier (Extreme Gradient Boosting)

XGBoost is an ensemble learning method based on gradient boosting that uses decision trees as base learners. It incorporates advanced regularization and optimization techniques for enhanced performance.

Strengths: Highly efficient and effective for structured data. It provides excellent results in machine learning competitions.

6. KNN (K-Nearest Neighbors)

KNN is a simple, instance-based learning algorithm that classifies data based on the majority vote of its k nearest neighbors in the feature space.

Strengths: Easy to understand and implement. Suitable for small datasets with a well-defined distance metric.

Complexity: Linear complexity for preprocessing and feature extraction, high computational complexity for training the model.

Algorithm 4.3 Audio-based Depression Detection

Input: Audio data (e.g., .wav files)
Output: Predicted depression label (0: Non-depressed, 1: Depressed)
Step 1: Feature Extraction
for each audio file x **do**
  $x, sr \leftarrow$ librosa.load($audio\_path$) ▷ Load audio file
  $chroma\_stft \leftarrow$ librosa.feature.chroma_stft($y = x, sr = sr$)
  $rmse \leftarrow$ librosa.feature.rms($y = x$)
  $spectral\_centroids \leftarrow$ librosa.feature.spectral_centroid($y = x, sr = sr$)
  $spectral\_bandwidth \leftarrow$ librosa.feature.spectral_bandwidth($y = x, sr = sr$)
  $spectral\_rolloff \leftarrow$ librosa.feature.spectral_rolloff($y = x, sr = sr$)
  $zero\_crossing\_rate \leftarrow$ librosa.feature.zero_crossing_rate($y = x$)
  $mfccs \leftarrow$ librosa.feature.mfcc($y = x, sr = sr$)
  Store extracted features: Chroma, RMSE, Spectral Centroid, Spectral Bandwidth, Spectral Rolloff, Zero Crossing Rate, MFCCs
end for
Step 2: Data Preprocessing
Normalize the features using StandardScaler:

$$\text{X\_train}, \text{X\_test} = \text{train\_test\_split}(X, y, \text{test\_size} = 0.2)$$

Apply SMOTE to handle class imbalance (if applicable):

$$X\_balanced, y\_balanced = \text{SMOTE}(X, y)$$

Step 3: Model Training
Train ML models (e.g., Random Forest, SVM, or a Neural Network) using the balanced features:

$$\text{model} = \text{train\_model}(X\_train, y\_train)$$

Step 4: Prediction
Predict the depression status for a given audio file using the trained model:

$$\hat{y} = \text{model.predict}(X\_test)$$

Step 5: Output
if $\hat{y} = 1$ **then**
  Output: Depressed
else
  Output: Non-depressed
end if

4.3.3 System Complexity Overview

The preprocessing phase exhibits linear complexity, with its performance largely determined by the size and duration of the audio samples. The operations of noise removal and conversion of signals into microvolt units introduce minimal computational overhead while ensuring that the data is clean and suitable for effective feature extraction. In contrast, the training phase utilizing the Bi-LSTM model demands significant computational resources. This is due to the complexity of the deep learning architecture, which is designed to process high-dimensional feature data efficiently. Despite its computational intensity, the Bi-LSTM model excels at providing accurate predictions for depression detection, making it a pivotal component of the system.

CHAPTER 5

IMPLEMENTATION RESULTS AND DISCUSSION

A deep down analysis on the implementation process in Depression detection using various modalities along with a detailed analysis on the project's performance over various performance metrics.

5.1 IMPLEMENTATION OF VARIOUS MODALITIES

5.1.1 Quantitative EEG Channel Data

Input: The input for this module is a .csv file containing EEG data, which includes 1140 features across different brain waves (alpha, beta, gamma, delta, and theta) from 19 EEG channels. Additionally, the dataset contains demographic and categorical features like gender, age, and disorders.

Step1: Data Preprocessing:

Handle Missing Values: The first step is to check for any missing values in the dataset. If any are found, they will be imputed using the column mean or median to prevent any loss of information. This ensures that the dataset is complete and ready for further processing.

Ordinal Encoding: Categorical features like gender, age group, and disorders need to be encoded into numerical values. This is done using Ordinal Encoding, which assigns a unique number to each category (e.g., ”Male” = 1, ”Female” = 2). This makes the data usable by machine learning algorithms.

Normalization of Continuous Variables: Continuous variables such as the EEG signal values are normalized using StandardScaler, which scales the values so that they have a mean of 0 and a standard deviation of 1. This helps in speeding up the convergence of the machine learning algorithm.

Step 2: Feature Selection Using ElasticNet:

Apply ElasticNet: ElasticNet is used for feature selection, where it applies both L1 (lasso) and L2 (ridge) penalties to regularize the model. This technique helps to identify and retain the most significant features (in this case, the top 50 EEG features). ElasticNet strikes a balance between Lasso and Ridge regression, helping to eliminate irrelevant features while preserving important ones.

Retain Top 50 Features: After applying ElasticNet, only the top 50 most significant features are retained for further analysis. These features are chosen based on their ability to contribute to the prediction task, thus improving model performance.

Step 3: Balancing the Dataset with SMOTE:

Check for Class Imbalance: It's important to check for class imbalance in the dataset, where some mental disorders may have fewer instances than others. To address this, the Synthetic Minority Over-sampling Technique (SMOTE) is applied.

Apply SMOTE: SMOTE generates synthetic data points for the minority class by interpolating between existing samples. This helps in balancing the dataset, ensuring that the classifier is not biased toward the majority class and performs better on all classes.

Step 4: Model Selection and Model Training with XGBoost:

Model Selection: Several models, such as XGBoost, LGBM Classifier, Bi-LSTM, RNN, Gradient Boosting, AdaBoost, and KNN Classifier, were tested. Based on the performance metrics, XGBoost outperformed all other models. As a result, XGBoost was selected as the best model for this task.

Model Training with XGBoost: After selecting XGBoost as the best model, the next step was to train the model using the selected features (top 50 features) and the balanced dataset. XGBoost, a gradient boosting algorithm, is known for its high performance in classification tasks and is suitable for the complexity of the mental disorder prediction task. The model was trained on the data and used to make predictions for classifying one of the 12 mental disorders based on EEG features.

5.1.2 Challenges faced in QEEG module

Handling missing data in EEG datasets due to sensor failures or corruption required careful imputation, with mean or median methods chosen to maintain data integrity and minimize bias. Feature selection using ElasticNet, while reducing dimensionality, posed risks of overfitting or selecting irrelevant features, highlighting the need for cross-validation to optimize hyperparameters. Addressing class imbalance through SMOTE helped balance underrepresented mental disorder classes, but careful tuning was essential to prevent overly similar synthetic data. Selecting and tuning models, particularly XGBoost, involved significant computational challenges, requiring meticulous adjustment of hyperparameters like learning rate and max depth to achieve optimal performance on complex EEG data. These steps ensured the system's robustness and reliability across diverse scenarios.

5.1.3 Expected Output

A trained XGBoost model capable of predicting one of 12 mental disorders from the EEG data. The model is capable of providing consistent and accurate predictions throughout various user inputs, thereby proving its accurate predictions.

5.2 EEG-EDF BASED DEPRESSION DETECTION

Input: The input for this module consists of EEG data in the EDF (European Data Format) format, recorded from 30 healthy and 34 MDD (Major Depressive Disorder) participants. The data is split across three EEG recording states: Eyes Opened (EO), Eyes Closed (EC), and Task. Each state is associated with EEG signals used for classification.

Step 1: Data Preprocessing:

Handle Missing Values: The first step is to check for any missing values in the dataset. If missing data is found, it will be imputed using the column mean or median to avoid losing valuable information and ensure the dataset is complete for model training.

Bandpass Filtering and Signal Conversion: The EEG signals are bandpass filtered between 0.5 and 45 Hz to remove any noise outside this range, which is optimal for analyzing EEG data related to mental health. The signal values are converted from Volts (V) to Microvolts (V) for better accuracy in further processing.

Data Visualization: The EEG signals are visualized and saved as PNG images categorized by the EEG state (EC, EO, Task) and label (Healthy, MDD). These images will be used for training deep learning models.

Step 2: Model Selection and Model Training with ResNet18:

Model Selection: Several models were evaluated based on their performance in detecting depressive (MDD) and healthy individuals. The following models were tested: ResNet18, DenseNet, EfficientNet, and Swin Transformer. Based on the performance metrics, ResNet18 outperformed all other models in terms of train and test accuracy, with a relatively low loss. Therefore, ResNet18 was selected as the best model for this task.

Model Training with ResNet18: After selecting ResNet18 as the best model, it was trained on the processed EEG images (from the different states: EC, EO, Task) using the corresponding labels (Healthy, MDD). ResNet18 was fine-tuned for binary classification, where it learns to differentiate between depressive (MDD) and healthy states based on the EEG features. The model was trained for 6 epochs, and CrossEntropyLoss was used as the loss function with the Adam optimizer to minimize the error during training.

5.2.1 Challenges faced in EEG-EDF module

Handling noise in EEG signals due to artifacts like muscle activity or eye movement required preprocessing, including bandpass filtering (0.5–45 Hz), with careful parameter tuning to preserve useful information. The small dataset of 30 healthy and 34 MDD participants necessitated data augmentation and techniques like SMOTE to balance and expand the dataset for deep learning. Selecting models like ResNet18, DenseNet, and EfficientNet for classification demanded significant computational resources and precise tuning of architecture, epochs, and learning rate to avoid overfitting or underfitting, with ResNet18 emerging as optimal after extensive adjustments. The presence of EEG data across three states (EO, EC, and Task) added complexity, requiring separate consideration of states and cross-validation to ensure robust performance across all conditions.

5.2.2 Expected Output

A trained ResNet18 model capable of accurately classifying EEG data into Healthy or MDD categories across the different EEG states (EC, EO, Task). The model provides high accuracy, ensuring reliable predictions for detecting depressive individuals from the EEG signals.

5.3 AUDIO BASED DEPRESSION DETECTION

Input: The input for this module consists of audio recordings from 189 interview sessions collected during Wizard-of-Oz interactions with an animated virtual interviewer (Ellie). Each session ranges from 7 to 33 minutes, with an average duration of 16 minutes.

Step 1: Data Preprocessing

Audio File Preparation: The first step involves loading the audio files using specialized audio processing libraries. The audio is standardized to a sampling rate of 16 kHz, converted to mono channel, and normalized to ensure consistent amplitude across all recordings. Any incomplete or corrupted audio segments are removed to ensure data integrity.

Silence Removal and Segmentation: Using advanced silence

filename	chroma_st	rmse	spectral_c	spectral_b	rolloff	zero_cross	mfcc1	mfcc2
/kaggle/wo	0.308947	0.014098	1371.868	1503.371	2701.625	0.057336	-407.445	70.3146
/kaggle/wo	0.262135	0.007088	1148.674	1303.299	2331.071	0.044797	-480.702	57.23192
/kaggle/wo	0.274031	0.004838	1089.079	1417.03	2226.028	0.036741	-567.431	53.57769
/kaggle/wo	0.339243	0.003493	1286.012	1594.187	2764.472	0.043829	-485.097	45.5759
/kaggle/wo	0.288583	0.005527	1184.221	1419.303	2371.016	0.046802	-491.42	56.01936
/kaggle/wo	0.256882	0.033055	1097.963	1289.244	2197.737	0.044893	-367.263	77.29315
/kaggle/wo	0.213138	0.008536	1159.278	1271.978	2217.249	0.052819	-475.211	63.31031
/kaggle/wo	0.27645	0.005688	1183.718	1399.597	2385.929	0.046206	-478.53	57.77628

Figure 5.1: Features Extracted After Preprocessing

detection algorithms, non-speech portions of the audio are removed. The remaining audio is segmented into continuous speech regions where the participant's responses are captured.

Audio Data Augmentation: To enhance model generalizability, several augmentation techniques are applied. One of the techniques is noise injection, where random white noise is added to the audio with an amplitude of 5% of the maximum signal amplitude.

Modifying the audio playback speed and pitch: The playback speed is altered between 0.8x and 1.4x the original speed, and pitch characteristics are also modified. This helps the model become robust to variations in speaking pace and pitch, which can vary across different speakers.

Time stretching is applied, where the duration of the audio is stretched or compressed with a stretch rate of approximately 0.8. This increases variability in the training data and helps the model generalize better by simulating different speech rates.

Step 2: Feature Extraction

Extract relevant acoustic features from the audio that capture depression-related characteristics. These features include Chroma STFT, which represents the pitch class profile, RMS Energy, which measures the signal loudness, and Spectral Centroid, which indicates the brightness of the sound.

Spectral Bandwidth describes the distribution of the sound spectrum, while Zero Crossing Rate captures the frequency content of the signal. Additionally, Mel-Frequency Cepstral Coefficients (MFCCs) are used; these 20 coefficients capture timbral characteristics and are widely used in speech and emotion analysis.

Step 3: Data Preparation

Class Balancing: To address any potential class imbalance (depressive vs. non-depressive), the Synthetic Minority Over-sampling Technique (SMOTE) is applied to generate synthetic samples for the minority class, ensuring the model is not biased towards the majority class.

Feature Normalization: Features are normalized using StandardScaler, which transforms the features to zero mean and unit variance, ensuring consistency in the scaling of the input features.

Step 4: Model Development and Training

Model Selection: Several machine learning models were evaluated for their performance in detecting depressive individuals based on the extracted audio features. The models used are Bi-LSTM, ConvLSTM2D, LSTM, Dense layered sequential model, XGBoost Classifier, KNN, SVC Model, Bagging Classifier, Decision Tree, and Logistic Regression. Based on the performance metrics, Bi-LSTM appeared to be the most suitable for further processes.

Model Training with Bi-LSTM: After selecting Bi-LSTM as the best model, it is trained on the preprocessed audio features using the following architecture: **Input layer:** Corresponding to the dimensions of the extracted audio features. **Hidden layers:** Bi-directional LSTM layers that capture temporal dependencies in the audio sequences. **Output layer:** A single neuron with a sigmoid activation function for binary classification (Depressive vs. Non-Depressive).The model is trained using Binary Cross-Entropy as the loss function and Adam optimizer with a learning rate of 0.001. The training process is monitored for accuracy and loss over the specified number of epochs.

Step 5: Model Deployment

Model Persistence: Once the model is trained and evaluated, it is saved using serialization techniques (e.g., joblib or pickle). The saved model includes both the model weights and configuration, allowing for future inference and fine-tuning.

5.3.1 Challenges faced in Audio module

Ensuring audio data quality and consistency was challenging due to variability in recording conditions and durations, which required preprocessing steps like normalization, mono conversion, and silence removal. Extracting relevant acoustic features such as MFCCs, Spectral Centroid, and Zero Crossing Rate demanded careful selection to capture depression-related patterns effectively. Class imbalance between depressive and non-depressive samples posed another issue, addressed by SMOTE and comprehensive performance evaluation metrics like F1-score. Training complex models like Bi-LSTM on high-dimensional audio data risked overfitting, mitigated through regularization, cross-validation, and hyperparameter tuning. Deployment challenges, including model persistence and handling real-world variability, were resolved using serialization techniques and a robust inference pipeline for practical application.

5.3.2 Expected Output

A trained Bi-LSTM model capable of predicting whether a subject is Depressive or Non-Depressive based on their audio responses. Performance metrics, including accuracy, precision, recall, and F1 score. Visualizations such as the confusion matrix and performance metric graphs. A saved model for future use, allowing for easy inference and further model improvements.

5.4 QUANTITATIVE EEG CHANNEL DATA BASED DEPRESSION DETECTION

5.4.1 Explanation of Metrics Employed

The following metrics were used to evaluate the models:

- **Accuracy**: The ratio of correctly predicted observations to the total observations.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

- **Precision**: The ratio of correctly predicted positive observations to the total predicted positive observations.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

- **Recall**: The ratio of correctly predicted positive observations to all observations in the actual class.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

- **F1-Score**: The weighted average of Precision and Recall.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

5.4.2 Evaluation Metrics

The performance of various models based on accuracy, precision, recall, F1 score, and AUC-ROC values is summarized in Table 5.2. Among the evaluated models, **XGBoost** achieved the best performance, with 92% accuracy, precision, recall, and F1 score. Its excellent AUC-ROC values further validate its ability to handle the complex, high-dimensional EEG dataset effectively. The performance of each model is visualized using a confusion matrix shown in Figure 5.3.

Model	Accuracy	Precision score	Recall	F1 score
XG Booster	92%	92%	92%	92%
LGBM Classifier	91%	91%	91%	91%
Bi-LSTM	90%	89%	87%	88%
RNN	83.98%	81.98%	79.83%	80.22%
Gradient Booster	89%	89%	89%	89%
Ada Booster	25%	8%	25%	12%
KNN Classifier	70%	67%	70%	66%

Figure 5.2: Model Evaluation metrics for QEEG Channel Data

5.4.3 Performance Analysis

The **XGBoost** model emerges as the top performer, achieving an impressive 92% across all evaluation metrics, including accuracy, precision, recall, and F1 score. These results demonstrate its capability to handle the complexity of the EEG data and distinguish between the different mental disorders effectively. The model's excellent AUC-ROC values further reinforce its robustness in depression detection. In comparison, **LGBM** showed strong performance but lagged slightly behind **XGBoost**, particularly in distinguishing subtle differences between disorder classes. Despite this, **LGBM** exhibited competitive results, making it a reliable option for depression detection tasks. The **Bi-LSTM** model demonstrated decent performance, but it lacked the precision-recall insights offered by the tree-based models. Although the model showed promise, it struggled with fine-grained classification tasks. The **Gradient Booster** model, while still accurate with an 89% accuracy rate, demonstrated lower computational efficiency compared to **XGBoost**, making it less practical for real-time applications. Traditional machine learning models such as **Logistic Regression** and **SVM** performed poorly, especially in distinguishing between multiple mental disorders, which underscores the superior capability of **XGBoost** and other tree-based models in handling the high-dimensional EEG data.

EEG EDF Models Performance Metrics

Model	Train Accuracy	Test Accuracy	Loss
ResNet18	97.20%	86.11%	0.2555
DenseNet	88%	75%	0.0819
EfficientNet	87%	80%	0.0519
Swin Transformer	52.45%	55.56%	0.7174

Figure 5.3: Model Performance metrics for EEG-Signal EDF Data

5.5 EEG-EDF BASED DEPRESSION DETECTION

5.5.1 Evaluation Metrics

The performance of various models based on accuracy, loss, and other evaluation metrics is summarized in Table 5.3. Among the evaluated models, **ResNet18** achieved the best performance, with 86.11% accuracy and a loss value of 0.2555, indicating its superior generalization and optimization capabilities. The model's performance is validated using a confusion matrix shown in Figure 5.4.

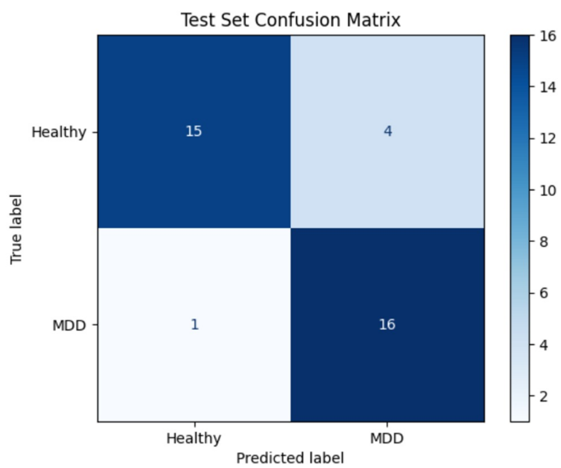

Figure 5.4: Confusion Matrix for EEG-EDF Modality

5.5.2 Performance Analysis

The **ResNet18** model demonstrated an optimal balance between high test accuracy (86.11%) and low loss (0.2555), making it the most reliable model for detecting Major Depressive Disorder (MDD) and healthy states. Its ability to learn spatial and temporal features from the EEG signals made it superior in handling the complexities of EEG data. The **EfficientNet** model, while performing well with an 80% accuracy, showed better loss minimization (0.0519), but at the cost of a slightly lower overall accuracy. This highlights a trade-off between achieving better loss reduction and optimizing accuracy. The **DenseNet** model exhibited a significant overfitting issue, with an accuracy of 88% on the training set but only 75% on the test set, indicating that it struggled to generalize well on unseen data. In contrast, the **Swin Transformer** underperformed with a test accuracy of only 55.56% and a high loss of 0.7174, suggesting that it may not be suited for this task in its current form. Overall, **ResNet18** outperformed other models due to its ability to effectively capture and process the key features of EEG data, making it the most reliable model for detecting depression from EEG-EDF signals.

5.6 AUDIO BASED DEPRESSION DETECTION

5.6.1 Evaluation Metrics

The performance of various models based on accuracy, precision, recall, and F1 score is summarized in Table 5.5. Among the evaluated models, the **Bi-LSTM + Attention** model achieved the best performance, with 90.5% accuracy, 89.5% precision, 89% recall, and 89% F1 score. Therefore, **Bi-LSTM + Attention** is employed as the primary model for audio-based performance due to its superior metrics. The performance is validated using a confusion matrix shown in Figure 5.6.

Audio Models Performance Metrics

Model	Accuracy	Precision score	Recall	F1 score
Bi- LSTM + Attention	90.5%	89.5%	89%	89%
Bi- LSTM	89%	87%	87%	87.5%
ConvLSTM2D	88%	89%	88%	88%
LSTM	87%	88%	87%	87%
Dense layered sequential model	87%	87%	87%	97%
XG Boost Classifier	74%	73%	62%	63%
KNN	73%	73%	60%	60%
SVC Model	70%	73%	53%	48%
Bagging Classifier	70%	64%	58%	58%
Decision Tree	69%	61%	57%	57%
Logistic Regression	68%	59%	56%	55%

Figure 5.5: Model Performance metrics for Audio

5.6.2 Performance Analysis

The **Bi-LSTM + Attention** model achieved the best performance, with 90.5% accuracy, 89.5% precision, 89% recall, and 89% F1 score, effectively capturing intricate temporal patterns in audio data. The **Bi-LSTM** (89%) and **ConvLSTM2D** (88%) models also demonstrated strong results, showcasing their robustness in sequential data processing. The **Dense Layered Sequential Model** excelled in precision-recall balance, achieving an impressive F1 score of 97%, despite having a similar accuracy (87%) to the **LSTM**. In contrast, traditional machine learning models like **XGBoost**, **KNN**, and **Bagging Classifier** underperformed, with accuracies between 70% and 74%. The weakest results came from **Logistic Regression**, with 68% accuracy and 55% F1 score, and models like **SVC** and **Decision Tree** struggled with low recall. Overall, deep learning models, especially **Bi-LSTM + Attention**, significantly outperformed traditional approaches, proving their superiority for audio-based depression detection.

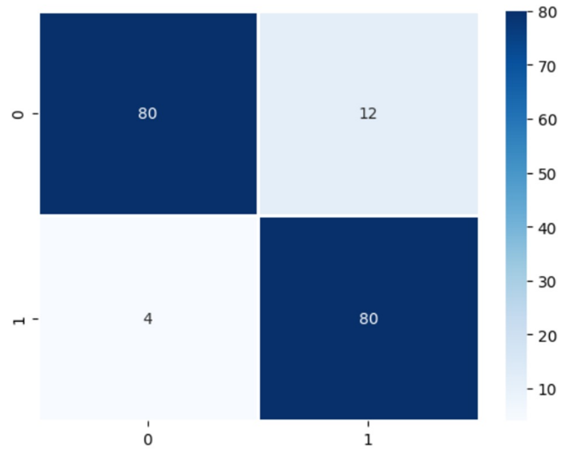

Figure 5.6: BiLSTM ConfusionMatrix for Audio Modality

5.7 CONCLUSION FROM ANALYSIS

The performance analysis across EEG and audio modalities for depression detection demonstrates the superiority of advanced models. XGBoost achieved the highest accuracy (92%) for EEG detection, while ResNet18 showed an optimal balance with 86.11% accuracy. For audio, the Bi-LSTM + Attention model outperformed others with 90.5% accuracy. Traditional models like Logistic Regression and KNN underperformed in both modalities. Overall, deep learning models, especially XGBoost and Bi-LSTM + Attention, proved most effective for depression detection.

5.8 VISUAL GUIDE TO WEBSITE FEATURES

5.8.1 Website Overview

The web application is designed for depression detection with a user-friendly interface. It includes three modules: QEEG-based, EEG signal-based, and audio-based depression detection. The workflow is divided into a series of pages for an intuitive experience.

5.8.2 Landing Page

The landing page given in Figure 5.7 is the user's first interaction with the application. It prominently displays the project name and provides a simple, clean design. A navigation menu directs users to the different modules.

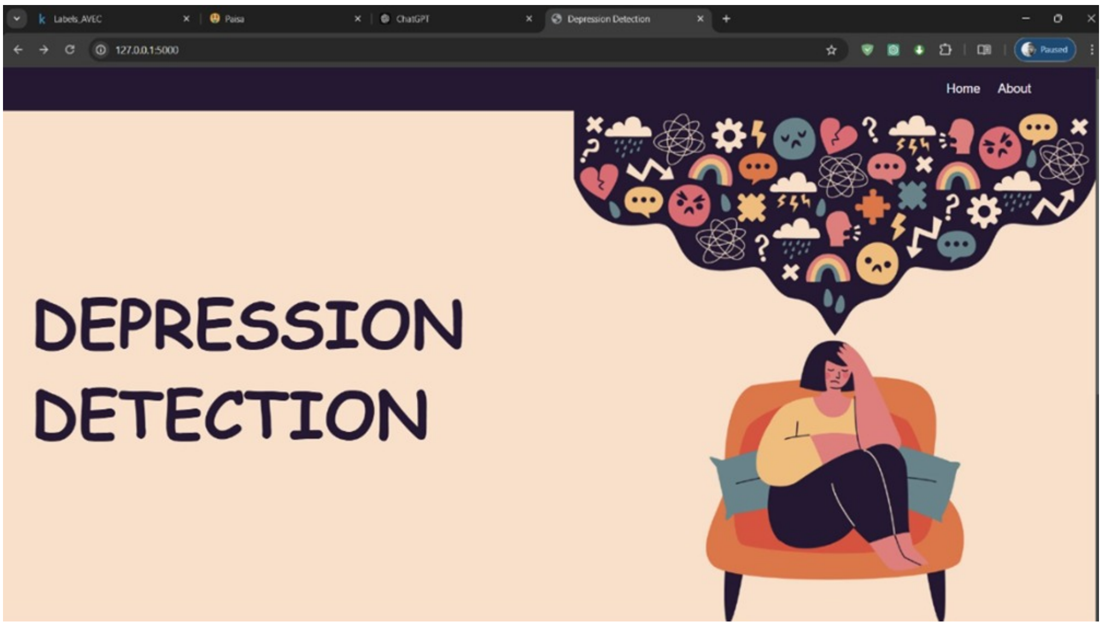

Figure 5.7: Landing Page

5.8.3 Menu Page

The menu page acts as the central hub for accessing the three modules. Each module is represented by a clickable button. Clicking on a module redirects the user to the corresponding interface which has been shown in the Figure 5.8.

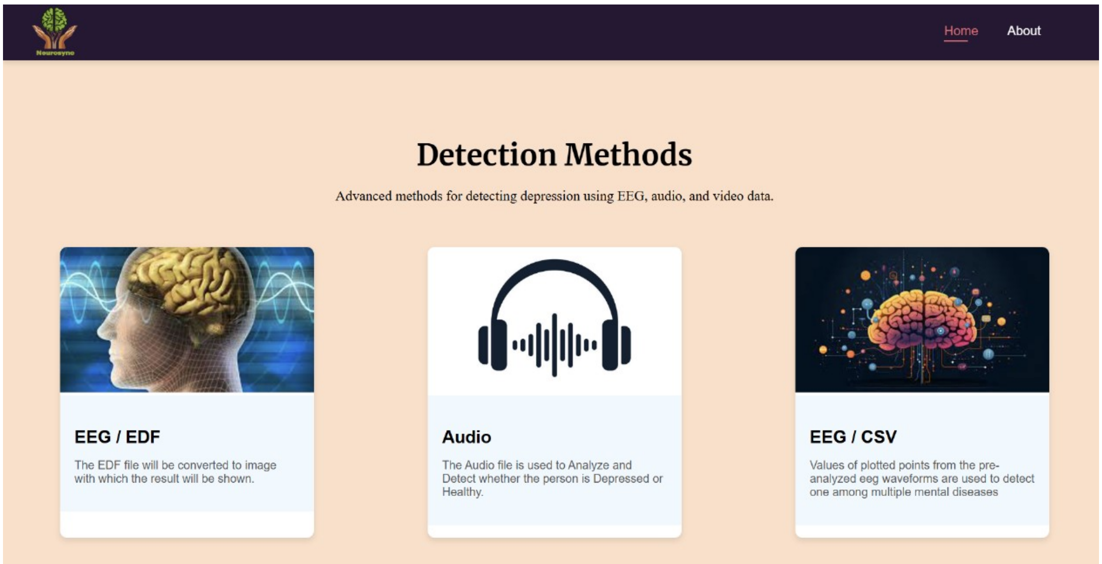

Figure 5.8: Menu Page

The modules are: **QEEG Depression Detection (CSV-based), EEG Signal-based Depression Detection (EDF file-based), and Audio-based Depression Detection (WAV file-based)**. Navigation is seamless for easy access.

5.8.4 QEEG Depression Detection

Upon selecting this module, users upload a CSV file containing EEG data. The system preprocesses the file by extracting features, normalizing data, and handling missing values.The process is displayed in the Figure 5.9

Figure 5.9: Interface for QEEG Depression Detection

5.8.5 EEG Signal-based Depression Detection

Users upload an EDF file containing raw EEG signal data as shown in the Figure 5.10. The backend processes the data using MNE to filter, extract features, and apply a deep learning model for depression detection.

5.8.6 Audio-based Depression Detection

Users upload a WAV file with audio data, the user interface for this process is mentioned in the Figure 5.11. Acoustic features like MFCCs and spectral centroid are extracted, passed through a Bi-LSTM model, and results are displayed.

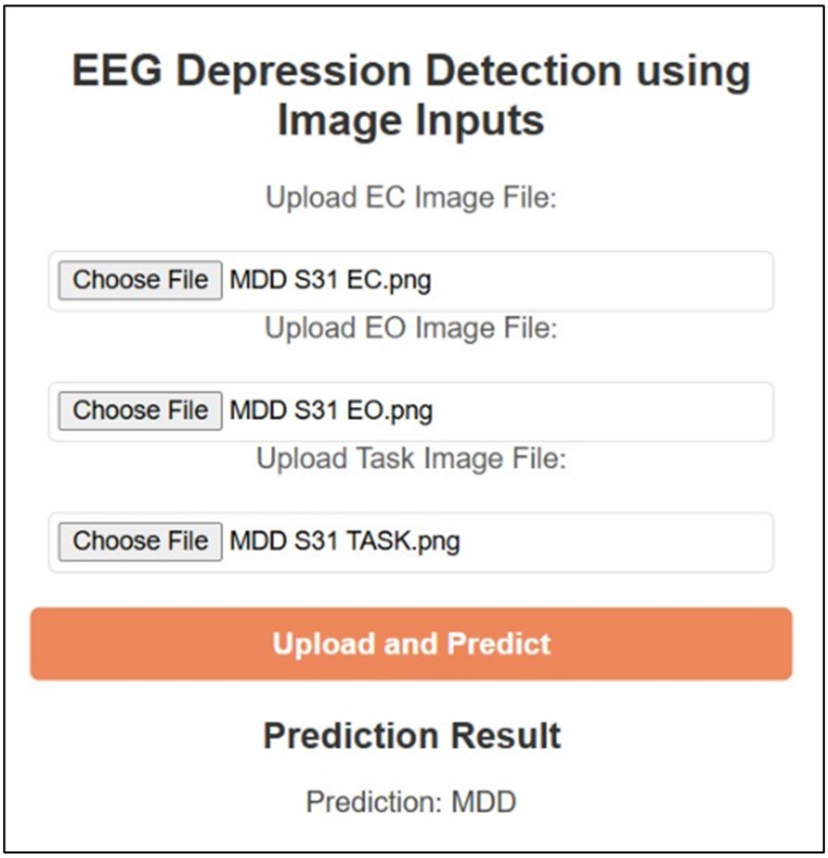

Figure 5.10: Interface for EEG Signal Based Depression Detection

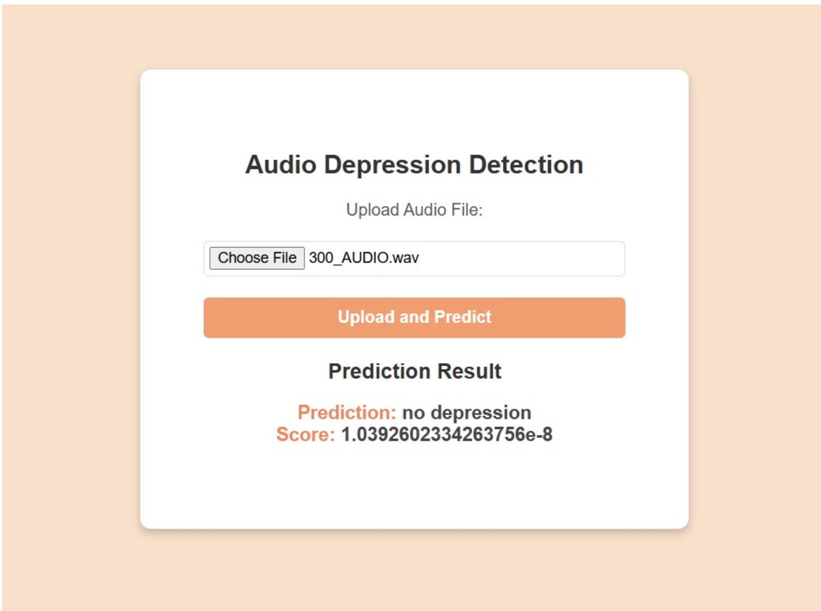

Figure 5.11: Interface for Audio Based Depression Detection

CHAPTER 6

CONCLUSIONS AND FUTURE WORKS

A deep down analysis on the implementation process in depression detection using various modalities along with a detailed analysis on the project's performance over various performance metrics.

6.1 CONCLUSION

The project successfully demonstrates the application of Machine Learning and Deep Learning techniques to the field of depression detection. By leveraging EEG and audio data, this multi-phase system enables early detection of depressive symptoms and provides actionable insights to aid clinical assessment. The project is structured into three modules: detecting depression through EEG signals, binary classification using demographic and categorical features, and spectrogram-based audio analysis for depression classification. Each of these modules is designed to provide comprehensive insights into depressive patterns from diverse data sources, ensuring robust detection capabilities across multiple dimensions of analysis. Each module showcased impressive performance metrics, with accuracy levels exceeding 85%, validating the methodologies and algorithms employed. The integration of advanced neural network architectures, such as ResNet-34 and EfficientNet, along with transfer learning, contributed significantly to the detection accuracy. Challenges related to data quality, class imbalance, and computational constraints were systematically addressed through preprocessing techniques, feature selection, data augmentation, and hyperparameter tuning. These optimizations ensured that the system was both accurate and scalable, highlighting the necessity of meticulous data preparation in enhancing model performance in real-world scenarios.

6.2 FUTURE WORK

Future work on this project will focus on advancing the diagnostic system into a comprehensive VR-based therapeutic platform, complementing depression detection with mental health interventions. This includes integrating 70 immersive Virtual Reality environments to offer therapeutic solutions for depression, anxiety, and fear management. Proposed features include exposure therapy scenarios for phobias, interactive relaxation spaces for stress relief, and cognitive games to stimulate mental activity and promote emotional well-being. Biofeedback mechanisms will utilize real-time EEG signals during VR sessions to dynamically adjust environments based on user responses, ensuring a personalized therapeutic experience. Additionally, the system will be enhanced by integrating multimodal data sources, such as video recordings and textual data (interview transcripts or chat interactions), to gain a more comprehensive understanding of depressive symptoms. A real-time monitoring system will be developed using wearable EEG devices and mobile recording tools, ensuring continuous detection and instant feedback on depressive symptoms, which is crucial in clinical and remote healthcare settings. Personalized diagnostic models will also be created to adapt to individual differences in brain activity and audio profiles, offering tailored and more accurate detection of depression patterns. To ensure practical usability, extensive validation studies will be conducted in collaboration with mental health professionals and clinical facilities, alongside the development of user-friendly interfaces for healthcare practitioners, facilitating seamless integration into existing diagnostic workflows and mental health screening protocols. By collaborating with mental health professionals, this project aims to validate and integrate the system in clinical settings. This approach bridges the gap between AI-driven diagnostics and real-world healthcare 71 applications, integrating advanced AI, data science, and VR technology to create scalable, accessible, and cost-effective mental health solutions that can support patients, therapists, and healthcare institutions worldwide.

REFERENCES

[1] X. Liu, H. Zhang, Y. Cui, T. Zhao, B. Wang, X. Xie, S. Liang, S. Sha, Y. Yan, X. Zhao, and L. Zhang. EEG-based major depressive disorder recognition by neural oscillation and asymmetry. *Frontiers in Neuroscience*, Feb 2024.

[2] Shuang Liu, Xue Liu, Dong Yan, Shiyan Chen, Yong Liu, Xiaoyu Hao, Weilin Ou, Zhi Huang, Fang Su, Fang He, and Daqing Ming. Alterations in Patients With First-Episode Depression in the Eyes-Open and Eyes-Closed Conditions: A Resting-State EEG Study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Feb 2024.

[3] Chenjian Sun, Min Jiang, Linlin Gao, Yu Xin, and Yihong Dong. A novel study for depression detecting using audio signals based on graph neural network. *Biomedical Signal Processing and Control*, 2024.

[4] V. Ravi, J. Wang, J. Flint, and A. Alwan. Enhancing accuracy and privacy in speech-based depression detection through speaker disentanglement. *Computer Speech and Language*, Jun 2024.

[5] Sara Sardari, Bahareh Nakisa, Mohammed Naim Rastgoo, and Peter Eklund. Audio based depression detection using Convolutional Autoencoder. *Expert Systems with Applications*, 189, 2022.

[6] Seong Min Park, Byung Kyu Jeong, Dong Yun Oh, Chang Hyun Choi, Hae Young Jung, Jin Yong Lee, Daehyun Lee, and Jong Sam Choi. Identification of Major Psychiatric Disorders From Resting-State Electroencephalography Using a Machine Learning Approach. *Frontiers in Psychiatry*, Aug 2021.

[7] Zahid Ahmed, Ali Wali, Shahid Shahid, Shafqat Zikria, Jawad Rasheed, and Tugce Asuroglu. Psychiatric disorders from EEG signals through deep learning models. *IBRO Neuroscience Reports*, Sep 2024.

[8] Vandana, Nikhil Marriwala, and Deepti Chaudhary. A hybrid model for depression detection using deep learning. *Measurement: Sensors*, 25, 2023.

[9] Xia Li, Rong La, Ying Wang, Bin Hu, and Xiang Zhang. A Deep Learning Approach for Mild Depression Recognition Based on Functional Connectivity Using Electroencephalography. *Frontiers in Neuroscience*, Apr 2020.

[10] Jing Shen, Xiang Zhang, Xian Huang, Min Wu, Jian Gao, Dan Lu, Zhiguang Ding, and Bin Hu. An Optimal Channel Selection for EEG-Based Depression Detection via Kernel-Target Alignment. *IEEE Journal of Biomedical and Health Informatics*, Jul 2021.

[11] Erik Cambria, Yuhong Li, Fei Zong Xing, Soujanya Poria, and Kam-Fai Kwok. SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*, Oct 2020.

[12] Brendan H. Hidaka. Depression as a disease of modernity: explanations for increasing prevalence. *Journal of Affective Disorders*, Nov 2012.

[13] Colin D. Mathers and Dejan Loncar. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine*, Nov 2006.

[14] Melissa Kulesza, Eric Pedersen, Patrick Corrigan, and Grant Marshall. Help-Seeking Stigma and Mental Health Treatment Seeking Among Young Adult Veterans. *Military Behavioral Health*, 2015.

[15] Moumita Chatterjee, Piyush Kumar, Poulomi Samanta, and Dhrubasish Sarkar. Suicide ideation detection from online social media: A multi-modal feature based technique. *International Journal of Information Management Data Insights*, 2(2), 2022.

[16] Wajeeha Mumtaz, Aamir Saeed Malik, Shujaat Syed Ali, Muhammad Asif Yasin, and Hussain Amin. Detrended fluctuation analysis for major depressive disorder. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015.

[17] Muhammad Yousufi, Robertas Damaševičius, and Raimundas Maskeliūnas. Multimodal Fusion of EEG and Audio Spectrogram for Major Depressive Disorder Recognition Using Modified DenseNet121. *Brain Sciences*, Oct 2024.

[18] Rui Bai, Yuxing Guo, Xiaodong Tan, Lingfeng Feng, and Huaming Xie. An EEG-based depression detection method using machine learning model. *International Journal of Pharma Medicine and Biological Sciences*, 2021.

[19] Abir Baskaran, Roumen Milev, and Roger S. McIntyre. The neurobiology of the EEG biomarker as a predictor of treatment response in depression. *Neuropharmacology*, Sep 2012.

[20] Mathias Benedek, Richard J. Schickel, Emanuel Jauk, Andreas Fink, and Aljoscha C. Neubauer. Alpha power increases in right parietal cortex reflects focused internal attention. *Neuropsychologia*, Apr 2014.

[21] Gerald E. Bruder, Jonathan W. Stewart, Charles E. Tenke, Patricia J. McGrath, Paulo Leite, Nilanjan Bhattacharya, and Frederick M. Quitkin. Electroencephalographic and perceptual asymmetry differences between responders and nonresponders to an SSRI antidepressant. *Biological Psychiatry*, Mar 2001.

[22] Yong Cao, Chao Han, Xiaoqin Peng, Zhiguo Su, Guangwei Liu, Yun Xie, Yumin Zhang, Jian Liu, Peijia Zhang, Wei Dong, Min Gao, Shanshan Sha, and Xuelian Zhao. Correlation Between Resting Theta Power and Cognitive Performance in Patients With Schizophrenia. *Frontiers in Human Neuroscience*, Apr 2022.

[23] World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. World Health Organization, 1992.

[24] Vaibhav N. Vahia. Diagnostic and statistical manual of mental disorders 5: A quick glance. *Indian Journal of Psychiatry*, Jul 2013.

[25] Thomas R. Insel. The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. *American Journal of Psychiatry*, Apr 2014.

[26] Xiaoxiao Zhang, Urs Braun, Heike Tost, and Danielle S. Bassett. Data-Driven Approaches to Neuroimaging Analysis to Enhance Psychiatric Diagnosis and Therapy. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, Aug 2020.

[27] Quentin J. M. Huys, Tiago V. Maia, and Michael J. Frank. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, Mar 2016.

[28] Ronald C. Kessler, Patricia Berglund, Olga Demler, Rong Jin, Daniel Koretz, Kathleen R. Merikangas, A. John Rush, Ellen E. Walters, and Philip S. Wang; National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*, Jun 2003.

[29] Jian Pan, Hui Lin, Yihong Dong, Yang Wang, and Yan Ji. MAMF-GCN: Multi-scale adaptive multi-channel fusion deep graph convolutional network for predicting mental disorder. *Computers in Biology and Medicine*, Sep 2022.

[30] Sumit Soni, Anirban Seal, Subrat Kumar Mohanty, and Kouichi Sakurai. Electroencephalography signals-based sparse networks integration using a fuzzy ensemble technique for depression detection. *Biomedical Signal Processing and Control*.

[31] Ayan Seal, Rishabh Bajpai, Mohan Karnati, Jagriti Agnihotri, Anis Yazidi, Enrique Herrera-Viedma, and Ondrej Krejcar. Benchmarks for machine learning in depression discrimination using electroencephalography signals. *Applied Intelligence*, 2022.

[32] Sumit Soni, Ayan Seal, Anis Yazidi, and Ondrej Krejcar. Graphical representation learning-based approach for automatic classification of electroencephalogram signals in depression. *Computers in Biology and Medicine*, Jun 2022.

[33] Lei Zhao, Donglin Wang, Shao-Wei Xue, Zhonglin Tan, Yan Wang, and Zhenzhen Lian. Aberrant state-related dynamic amplitude of low-frequency fluctuations of the emotion network in major depressive disorder. *Journal of Psychiatric Research*, 133, 2021.

[34] Lei Zhao, Donglin Wang, Shao-Wei Xue, Zhonglin Tan, Yan Wang, and Zhenzhen Lian. Aberrant state-related dynamic amplitude of low-frequency fluctuations of the emotion network in major depressive disorder. *Journal of Psychiatric Research*.

[35] Gopikrishna Deshpande, Deepak Rangaprakash, Laura Oeding, Andrzej Cichocki, and Xiao-Peng Hu. A New Generation of Brain-Computer Interfaces Driven by Discovery of Latent EEG-fMRI Linkages Using Tensor Decomposition. *Frontiers in Neuroscience*, Jun 2017.

[36] Mike X. Cohen. Where Does EEG Come From and What Does It Mean? *Trends in Neurosciences*, Apr 2017.

[37] U. Rajendra Acharya, V. K. Sudarshan, Hojjat Adeli, J. Santhosh, J. E. Koh, and Amir Adeli. Computer-Aided Diagnosis of Depression Using EEG Signals. *European Neurology*, 73, 2015. Epub 2015 May.

[38] U. Rajendra Acharya, V. K. Sudarshan, Hojjat Adeli, J. Santhosh, J. E. Koh, and Amir Adeli. Computer-aided diagnosis of depression using EEG signals. In *European Neurology*, 2015.

[39] Afshan Afshan, Jing Guo, Seung-Jae Park, V. Ravi, John Flint, and Abeer Alwan. Effectiveness of voice quality features in detecting depression. In *Proceedings of Interspeech 2018*, 2018.

[40] Relja Arandjelovic, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. NetVLAD: CNN Architecture for Weakly Supervised Place Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jun 2018.

[41] Jinhan Wang, Vijay Ravi, Jonathan Flint, and Abeer Alwan. Speechformer-CTC: Sequential modeling of depression detection with speech temporal classification. *Speech Communication*, 163, 2024.

[42] Kamal Berahmand, Farzad Daneshfar, Ehsan Shahmoradi Salehi, et al. Autoencoders and their applications in machine learning: a survey. *Artificial Intelligence Review*, 57(28), 2024.

[43] İbrahim Emre, Çağatay Erol, Cengiz Taş, and Nevzat Tarhan. Multi-class classification model for psychiatric disorder discrimination. *International Journal of Medical Informatics*, Feb 2023.

[44] Alex Esteva, Adrien Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Ken Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeffrey Dean. A guide to deep learning in healthcare. *Nature Medicine*, Jan 2019.

[45] Muhammad Sadiq Fareed, Shahid Zikria, Gulnaz Ahmed, Muzzud-din, Saqib Mahmood, Muhammad Aslam, Syeda Jilani, Ahmad Moustafa, and Muhammad Asad. ADD-Net: An Effective Deep Learning Model for Early Detection of Alzheimer Disease in MRI scans. *Applied Intelligence*, 2022.

[46] Sandra Garcia-Ponsoda, Jorge Garcia-Carrasco, Miguel Teruel, Alejandro Maté, and Juan Trujillo. Feature engineering of EEG applied to mental disorders: a systematic mapping study. *Applied Intelligence*, 53, 2023.

[47] Samanwoy Ghosh-Dastidar and Hojjat Adeli. Improved spiking neural networks for EEG classification and epilepsy and seizure detection. *Integrated Computer-Aided Engineering*, 2007.

[48] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. KNN Model-Based Approach in Classification. In *Lecture Notes in Computer Science*, 2003.

[49] Zhiheng Huang, Wei Xu, and Kai Yu. Bidirectional LSTM-CRF Models for Sequence Tagging. In *Proceedings of the 2015 Conference on Computational Natural Language Learning (CoNLL)*, 2015.