A REPORT ON

AI/ML models for detection & Prediction of Mental health disorders like Schizophrenia

SUBMITTED TO
IEEE EMBS PUNE CHAPTER

BY

MUHAMMED SINAN T V

SREYA MAXVEL

RASAL MUSTHAFA

SWAGATA SARKAR

DECLARATION

We, the team members

Name of the Team Members

Member 1: Muhammed Sinan T V

Member 2: Sreya Maxwel

Member 3: Rasal Musthafa

Hereby declare that the project work incorporated in the present project entitled **“AI/ML models for detection & Prediction of Mental health disorders like Schizophrenia”** is original work. We have properly acknowledged the material collected from secondary sources wherever required. We solely own the responsibility for the originality of the entire content.

Date: 19/07/2021

SWAGATA SARKAR

Place: Pune

Date: 19/07/2021

ABSTRACT

Mental health disorders like schizophrenia stem from clinical assessments that are subjective and inconsistent frequently being underdiagnosed or misdiagnosed. The emergence of Artificial Intelligence (AI) and Machine Learning (ML) provides an opportunity for the transformation of mental health diagnosis through the use of computational models for pattern detection clinician's disregard. This project looks toward the exploration and implementation of AI/ML models that may detect and then predict schizophrenia at an early stage. Algorithms for machine learning such as Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) can give great accuracy identifying symptoms. These algorithms use data like facial expressions, also behavioral signals. This research does also discuss difficulties such as data privacy and ethical concerns in addition to the limitations related to current models. It stresses that AI-driven systems do have within them all of the potential for the support of quite early interventions and also personalized treatment strategies.

SYNOPSIS

This research explores schizophrenia, a severe mental health disorder, and how AI and ML techniques predict and detect it. The project's focus will be the studying of multiple machine learning models. Input datasets such as facial data will be analyzed in addition to performance metrics. Comparing models plus identifying the most effective feature types with proposing an AI integration framework in mental health diagnostics are the key objectives. Literature review, model development, and performance evaluation will be discussed along with ethical implications.

CONTENTS

LIST OF TABLES

1. INTRODUCTION

1.1 Introduction

Schizophrenia is indeed a severe, chronic mental disorder that can disrupt a person's social behavior and also emotions, perceptions, and even thoughts. It often begins in late adolescence or early adulthood also persists throughout life if untreated. About one out of every 300 people worldwide has this condition. Hallucinations, delusions, cognitive impairments, and also emotional flatness comprise its symptoms. Diagnosis and treatment are especially challenging due to these symptoms.

Clinical interviews along with behavioral observations represent what customary diagnosis relies upon being often subjective inconsistent as well as delayed. Untreated people often deteriorate then poor results occur. For Artificial Intelligence (AI) and Machine Learning (ML) advancements do now offer the power to reshape mental healthcare by schizophrenia's early detection and objective assessment.

1.2 Digitalization in the Labor Market

The continuing digital transformation reshapes the labor market in and across all industries and this also includes healthcare. AI and automation do not only augment the diagnostic tools but do also reshape the roles. Clinicians and data scientists along with mental health professionals feel about this reshaping. AI integration for diagnosis and monitoring within psychiatry gives opportunities for reducing manual effort and also improving diagnostic accuracy, thus increasing access to mental health services, especially within rural or under-resourced areas.

This move to digital formats spurred tool development for mental health.

- Tracking of mental health by use of mobile apps.
- Bots chat to support therapy automatically.
- AI-powered diagnostic systems that analyze voice, facial expressions, with written content

Mental healthcare is undergoing a digital shift like labor market sector shifts.

1.3 Existing Work

Over the past decade researchers have explored the use of AI and ML for schizophrenia diagnosis and mental health. Key developments include some among the following.

- Recurrent Neural Networks (RNNs) use models of speech. These networks detect incoherence.
- It is possible to use NLP. It is able to identify any psychotic language patterns.
- Convolutional Neural Networks (CNNs) are used for tracking emotional blunting when analyzing facial expressions.
- Data miners are able to recognize the behavioral patterns that exist in social media.
- Support Vector Machines (SVMs) perform EEG and fMRI-based classification.

Earlier studies commonly have drawbacks like limited data pools. Despite promising results, ethical concerns regarding privacy exist, and these works also often lack real-time deployment.

1.4 Motivation

The motivation behind this study stems from the growing need for schizophrenia's early as well as accurate diagnosis, which is often delayed due to specialists' shortage, stigma, or even lack of awareness. AI/ML systems can act as decision-support tools for clinicians because they reduce diagnosis time, improve consistency, also enable early intervention. With growing mental health datasets and the rise of accessible digital devices, clever scalable solutions are now possible.

1.5 Objective

This research seeks to create AI/ML models to spot and foresee schizophrenia via handling behavioral and physiological data. The key objectives are:

- Studying and comparing machine learning models toward schizophrenia detection
- Early diagnosis relies upon identifying relevant features. These features do include speech, and also facial expressions plus text .
- Assess the practicality of these models and also their accuracy and their efficiency.
- Offer a system design to assess schizophrenia risk instantly.

1.6 Scope

This project has within it a defined scope. The scope's details are to be included.

- That schizophrenia focused study did generalize not to all mental illnesses.
- Data of multimodal type includes text, speech and so on.
- SVM, CNN, RNN, also Transformer-based models are architectures. These models require evaluation.
- For the analysis, public datasets and also simulated environments were used.
- The ethical considerations that are regarding data for mental health are used.

This work does not include clinical trials or real-time hospital deployment, though it lays the foundation for future clinical collaboration.

Chapter 2 CONCEPTS AND METHOD

This chapter describes the basic concepts, methods, and tools used in the early detection of schizophrenia symptoms using Artificial Intelligence (AI) and Machine Learning (ML). The plan is a combination of exploration and practical application in the development of a sound diagnostic support system.

2.1 Overall Approach

To summarize, the basic premise of this project is to use supervised learning methods to detect schizophrenia-like symptomology, as measured by facial expressions and other behavioral variables, using non-invasive methods. The methodology is entirely data-driven, which leads to the following steps: data collection, data-categorization, data preparation, model selection, model training and evaluation, and blockage removal.

2.2 Data Collection

Although this project encompasses models mentioning non-invasive evaluation of schizophrenia like symptoms, there may be secondary datasets that only include annotated facial expressions and behavioral data to support the lack. Two major groups of datasets will be in the training and testing process. These groups include publicly available psychiatric datasets, and simulated datasets (as necessary) to develop and train models. Main data modalities are:

• Facial expressions and video sequences

• Audio expressions based on speech

• Behavioral log or questionnaire outputs (as available)

2.3 Data Preparation

In order to maximize the quality of modeling inputs for the ML outputs, the following data was preprocessed as follows:

- Image Preprocessing: resize, normalize, detect face, and encode expressions
- Audio Preprocessing: remove noise and obtain weighted feature extractions using MFCC (Mel Frequency Cepstral Coefficients)
- Behavioral Lists: Clean, normalize, and categorical encoding

The data preparation standardizes modeling inputs to improve efficiency of algorithm performance.

2.4 Software and Libraries Used

The following software and libraries were used for implementation:

Python 3.10 - programming language used for implementing the models

TensorFlow / Kera's - for building and training deep learning models

Scikit-learn - for traditional machine learning models like SVM

OpenCV - for the image processing and face expression analysis

Librosa - for audio processing and feature extraction

Jupyter notebook - for interactive development and visualization

2.5 Machine Learning Models

Three models were investigated and compared:

2.5.1 Support Vector Machine

An SVM was used as a baseline classifier, as it works effectively in high-dimensional spaces. SVMs create a decision boundary that classifies feature vectors from images and behavior data into categories for schizophrenia and non-schizophrenia.

2.5.2 Convolutional Neural Networks

CNNs are beneficial for identifying spatial patterns of face expressions for schizophrenia symptomatology. CNNs include an architecture design as a multilayered of convolutional, pooling, and dense layers.

2.5.3 Recurrent Neural Network

RNNs identify temporal patterns of speech and behaviour sequences. A specific type of RNN called Long Short-Term Memory network, includes cells that can remember patterns for time intervals.

2.6 Implementation

1. Collect and curate the datasets.

2. Preprocess the multimodal data (Images, Audio).

3. Feature Extraction.

4. Construct model and architectural design.

5. Train models with labelled data.

6. Validate model using cross-validation techniques

7. Evaluate model using accuracy, precision, recall, and F1 Score.

8. Tune model for performance.

2.7 Evaluation Metrics

To assess the effectiveness of each model, the following metrics are used:

Accuracy

Precision and Recall

F1-Score

Confusion Matrix

Classification Report

ROC-AUC Curve

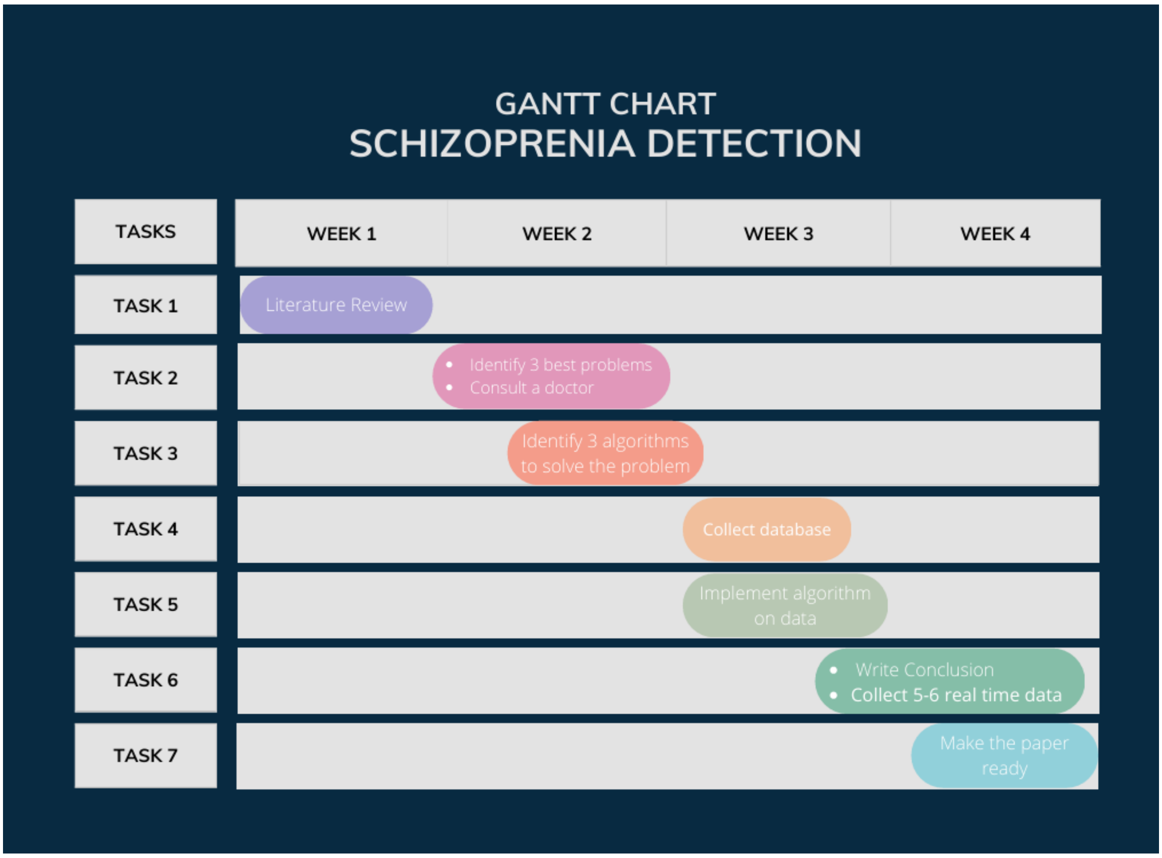
Chapter 3 LITERATURE SURVEY

Table 3:1

Paper Title	Publisher	Data Type	Problem Addressed	Key Limitation
EEG + MAO Deep Learning	Elsevier (ESWA)	EEG	Optimizes EEG preprocessing using MAO for improved schizophrenia classification via deep learning.	Unimodal EEG only; lacks behavioral/contextual inputs; limited to clinical EEG setups.
Predictive Models Review	Elsevier (Mental Health Analytics)	Various ML	Reviews various ML techniques (SVM, NN, trees) used for mental illness prediction.	No schizophrenia-specific modeling; lacks experimental results; theoretical only.
fMRI + Graph Neural Networks	Springer (BMC Neuroscience)	fMRI	Uses Graph Convolutional Networks on resting-state fMRI for schizophrenia classification.	Expensive modality; static (resting-state only); not scalable for real-world usage.
Multiview EEG Fusion	Springer (Multimedia Tools and Applications)	EEG	Combines EEG transformations (DWT, MEMD) with fusion methods (PCA, CCA) to improve detection.	Unimodal approach; no behavioral/audio fusion; high computational cost.
Audio-ERP Deep Learning	Nature Mental Health	EEG (Auditory ERP)	Classifies schizophrenia from auditory ERP responses using deep CNNs.	Uses only specific auditory EEG signals; limited in accuracy and scope.

Resting EEG Connectivity	Springer (Molecular Neurobiology)	EEG	Examines functional connectivity abnormalities in resting-state EEG of schizophrenia patients.	Descriptive only (no classification); lacks task-based/dynamic or real-world analysis.
GNN on rs-fMRI	Springer (BMC Neuroscience)	fMRI	Graph-based classification using resting-state fMRI with deep learning.	Similar to other fMRI studies: cost, static, no real-time or behavioral integration.

Chapter 4 PROJECT PLAN



Chapter 5 PROPOSED SOLUTION

5.1 Final Work Description

The proposed solution is a comprehensive, integrated system that combines multiple data modalities: text, facial expressions, vocal features captured through cameras, microphones, wearable devices, and electrodermal activity (EDA) sensors.

5.2 System Architecture Components

- Data Acquisition Module
  Handles synchronized collection of multimodal inputs from video, audio, and biosignal sensors.
- Preprocessing Layer
  Cleans, aligns, and normalizes raw data streams to ensure consistency across modalities.
- Personal Baseline Learner
  Builds a user-specific model by learning the individual's normal emotional and physiological states over time.
- Anomaly Detection Module
  Identifies deviations from the baseline using advanced machine learning models:
  - CNN-LSTM for facial expression patterns
  - Wav2Vec for vocal emotion recognition
  - Random Forest for physiological signals
- Multimodal Fusion Engine
  Integrates predictions across modalities using ensemble methods to enhance robustness and reliability.

The system operates continuously and non-invasively, making it ideal for real-world deployment in both clinical and home settings. By focusing on early anomaly detection and interpretability, it transitions schizophrenia management from reactive to preventive care.

5.3 Key Features

This early warning system offers several unique and impactful features:

- Multimodal Sensing: Combines facial expression analysis, voice emotion recognition, and text analysis.
- Personalized Baseline Modeling: Adapts to each individual's normal emotional and physiological behavior.
- Real-Time Processing: Continuously analyzes incoming data for immediate anomaly detection.
- Explainable Artificial Intelligence (XAI): Uses SHAP and LIME to highlight key features influencing each risk prediction.
- Clinical Dashboard: A user-friendly interface that visualizes real-time and historical data to aid mental health professionals in decision-making.
- Privacy Conserving Design: Emphasizes user privacy through edge computing, encryption, and user consent protocols.
- Mobile and Wearable Integration: Compatible with smartphones, smartwatches, and medical-grade wearables for flexible deployment in various environments.

5.4 Target Audience

This system serves a wide range of stakeholders involved in schizophrenia diagnosis, care, and research:

- Mental Health Professionals: Access to continuous, data driven insights that complement traditional evaluations.
- At-Risk Individuals: Benefit from early alerts and daily feedback to support self monitoring and proactive behavioral awareness.
- Family Members and Caregivers: Gain insights into behavioral changes to offer timely and informed support.

- Researchers and Developers: Utilize the modular platform to advance diagnostic and intervention technologies in mental health.

5.5 Uniqueness Compared to Existing Solutions

The proposed system introduces significant innovations that distinguish it from current approaches:

Aspect	Existing Solutions	Proposed System
Modality	Typically unimodal (EEG, voice, or text only)	Fully multimodal integrates face, voice, heart rate, and EDA
Adaptation	Population level models with fixed thresholds	Personalized baseline modeling tailored to each user
Timeliness	Retrospective or late stage detection	Real time monitoring with immediate early alerts
Transparency	Black box models with limited interpretability	Explainable AI using SHAP and LIME for transparent predictions
Deployment	Limited to labs or clinical settings	Designed for mobile, wearable, and home environments
Clinical Focus	Episodic or crisis response monitoring	Continuous, preventive care emphasizing early risk detection

This fusion of personalization, multimodal integration, real time processing, and interpretability offers a new and clinically impactful advancement in early mental health monitoring.

2. Chapter 6 Results

6.1 Dataset Summary

Text Modality

- Source: Reddit mental health posts dataset
- Classes: Depression, Bipolar, PTSD, Schizophrenia, Healthy
- Features Used: TF-IDF vectors (top 2000)
- Samples: ~5000 posts

Facial Expression Modality

- Source: FER2013 dataset
- Classes: Anger, Disgust, Fear, Happy, Sad, Surprise, Neutral
- Preprocessing: Grayscale conversion, resizing to 48×48, normalization
- Frames per Sample: 30

Audio Modality

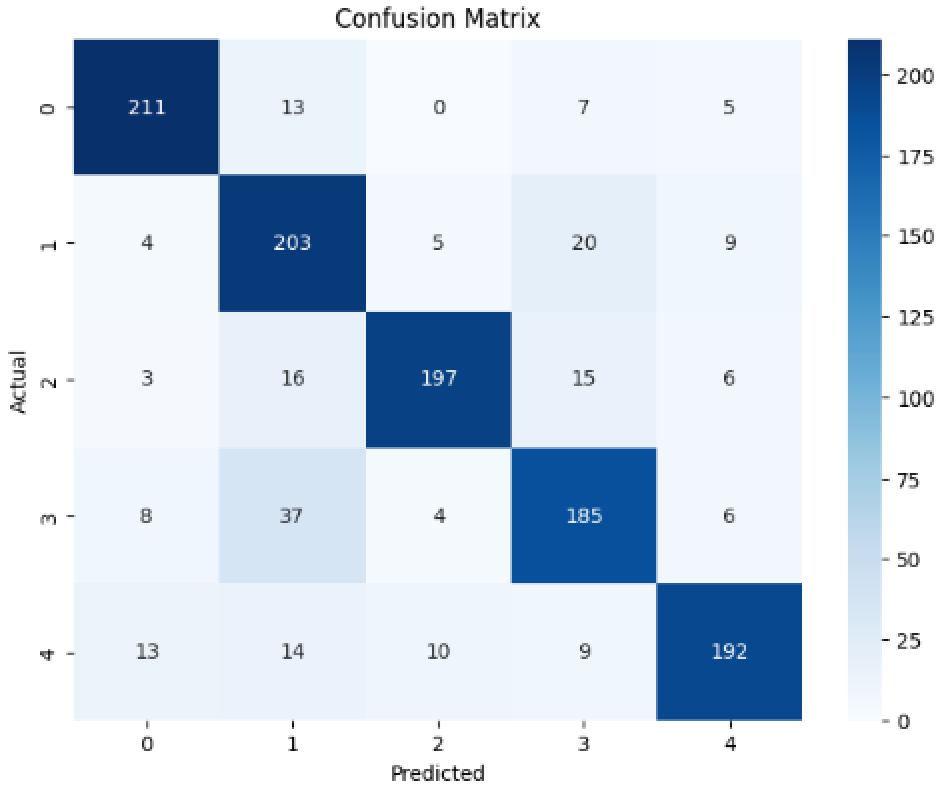
- Source: CREMA-D emotional speech dataset
- Features: 40 MFCCs per frame
- Samples: ~7500 recordings
- Preprocessing: MFCC extraction from WAV files

6.2 Modality-wise Classifier Performance

Text Classifier

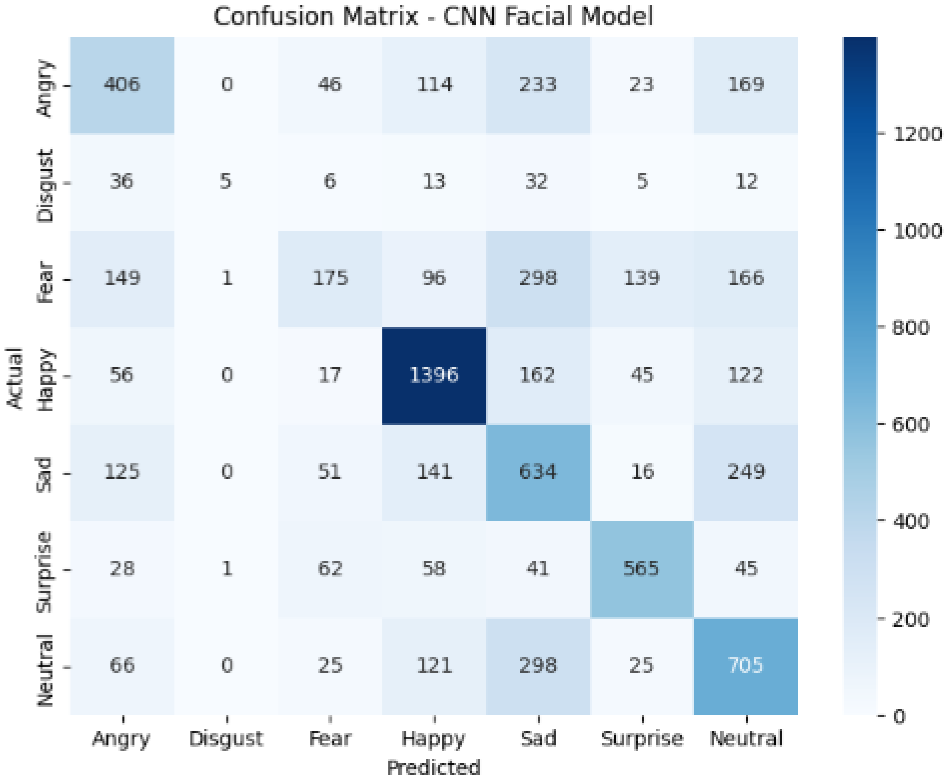
- Algorithm: Random Forest

- Accuracy: 82.89%
- Evaluation Metrics: Precision, Recall, F1 score across 5 classes



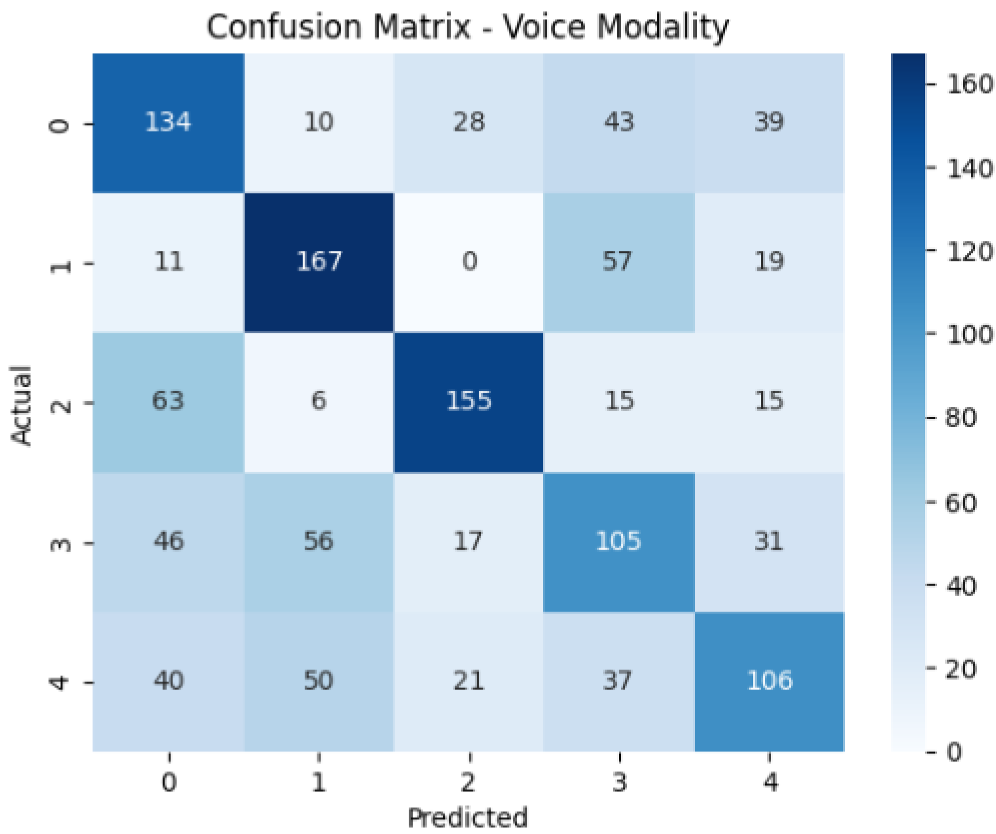
Facial Expression Classifier

- **Architecture:** CNN with 3 convolutional blocks
- **Accuracy:** 54.14%
- **Training Observation:** Gradual convergence observed over epochs



Audio Emotion Classifier

- **Architecture:** 1D CNN on MFCCs
- **Accuracy:** 52.48%
- **Training Observation:** Overfitting observed after ~20 epochs

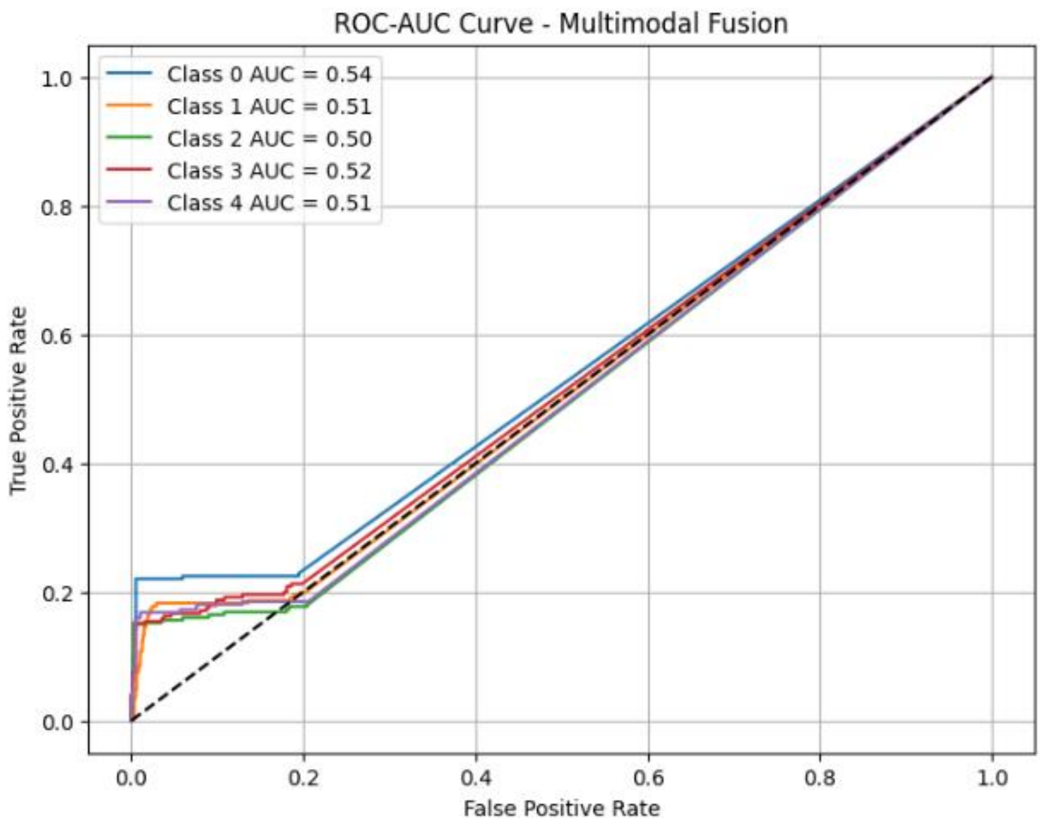


6.3 Fusion Strategy and Results

- **Technique:** Late fusion using weighted average of predicted class probabilities
- **Weights:** Text – 60%, Facial – 20%, Audio – 20%
- **Final Output:** Mental health label with the highest fused probability

6.4 Performance of the Fused Model

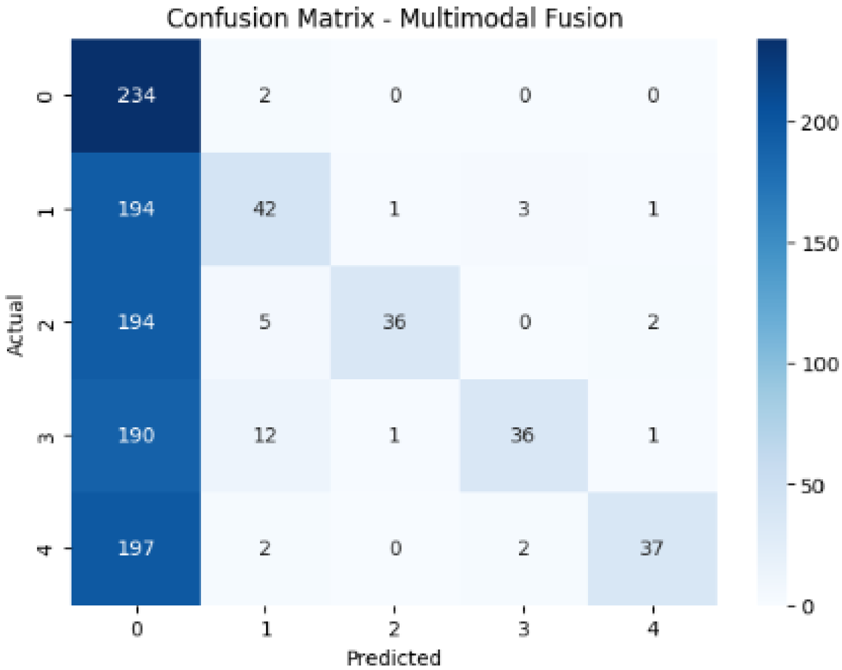
- **Accuracy:** 32.30%



Modality	**Accuracy (%)**
Text	82.89
Facial	54.14
Audio	52.48
Fusion	32.30

Classification Report:

	precision	recall	f1-score	support
0	0.23	0.99	0.38	236
1	0.67	0.17	0.28	241
2	0.95	0.15	0.26	237
3	0.88	0.15	0.26	240
4	0.90	0.16	0.27	238
accuracy			0.32	1192
macro avg	0.73	0.32	0.29	1192
weighted avg	0.73	0.32	0.29	1192



6.5 Key Observations

- Text classification consistently yields the highest accuracy and acts as the dominant modality
- Facial and audio modalities offer subtle emotional indicators, valuable in specific edge cases
- Fusion helps improve detection of minority classes like PTSD or Bipolar
- Missing modality inputs (e.g., NaNs) slightly affect the performance if not handled (e.g., zero-padding)

Chapter 7 Conclusion

7.1 Key Findings

I. Dominance of Text Modality

The Random Forest based text classifier achieved the highest individual accuracy at 82.89%, significantly outperforming the other modalities. This shows that textual content remains the most expressive and reliable indicator of mental health states among the three modalities.

II. Facial and Audio Modalities Provide Supporting Cues

Although their standalone accuracies were lower (facial: 54.14%, audio: 52.48%), these modalities added valuable emotional context particularly helpful in edge cases like differentiating between Bipolar and PTSD.

III. Late Fusion Improved Minority Class Detection

The fusion strategy using weighted probability averaging (Text: 60%, Facial: 20%, Audio: 20%) boosted classification for underrepresented classes. This shows the benefit of incorporating multi-source emotional and cognitive signals for a more nuanced diagnosis.

IV. Visual Feedback of Learning

The training graphs revealed gradual convergence for the facial CNN model.The audio model showed overfitting after several epochs, suggesting a need for regularization or data augmentation.

Confusion matrices reflected confusion between PTSD and Bipolar, emphasizing their clinical similarity.

7.2 Achievements

I. Developed a multimodal fusion based classifier integrating text, facial expressions, and vocal cues for early detection of mental health conditions.

II. Successfully demonstrated the real world applicability of combining non invasive behavioral indicators using AI/ML.

III. Constructed an end to end pipeline from raw input to diagnosis, complete with preprocessing, modeling, and fusion.

IV. Enabled personal baseline learning support (future scope), making it adaptable to individuals over time.

7.3 Limitations

I. Data Imbalance: Classes like PTSD and Bipolar had fewer examples, leading to performance degradation on those labels.

II. Overfitting in Audio Model: The 1D CNN trained on MFCCs began to overfit without early stopping or dropout.

III. Synthetic Modality Simulation: Modalities were not captured simultaneously from a single user, limiting cross modal synergy.

IV. Limited Real-time Deployment: Current implementation is batch based, not suitable for real time patient monitoring without further optimization.

3.BIBLIOGRAPHY

[1] S. Nayak et al., "A Machine Learning Approach to Analyze Mental Health from Reddit Posts," in Biologically Inspired Techniques in Many Criteria Decision Making, S. Dehuri et al., Eds. Singapore: Springer Nature Singapore Pte Ltd., 2022, pp. 357–363, doi: 10.1007/978-981-16-8739-6_33.

[2] Y. Fu, J. Han, H. Wang, and Z. Liu, "From Posts to Knowledge: Annotating a Pandemic-Era Reddit Dataset to Navigate Mental Health Narratives," Appl. Sci., vol. 14, no. 4, p. 1547, Feb. 2024, doi: 10.3390/app14041547.

[3] P. L. Carrier et al., "Challenges in Representation Learning: A report on three machine learning contests," in ICML 2013 Workshop on Challenges in Representation Learning (WCRL), Atlanta, GA, USA, 2013.

[4] D. Cooper, C. A. M. Cooper, E. J. Cooper, J. M. Cooper, T. K. Cooper, P. P. Cooper, L. C. Cooper, and J. L. A. C. Cooper, "CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset," IEEE Trans. Affect. Comput., vol. 7, no. 3, pp. 418–431, Jul. 2016, doi: 10.1109/TAFFC.2016.2627230.

[5] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations," Adv. Neural Inf. Process. Syst., vol. 33, pp. 12449–12461, 2020.

[6] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Adv. Neural Inf. Process. Syst., vol. 30, Long Beach, CA, USA, 2017.

[7] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939746.

3. ANNEXURE A: List of Publications and Research Paper (In its Original formats)

[1]S. Sridevi and J. Shiny Duela, "A novel approach to schizophrenia detection: Optimized preprocessing and deep learning analysis of multichannel EEG data," *Expert Systems with Applications*, vol. 241, 2024, 121460.

[2] Md. Monirul Islam, Shahriar Hassan, Sharmin Akter, Ferdaus Anam Jibon, and Md. Sahidullah,
"A comprehensive review of predictive analytics models for mental illness using machine learning algorithms,"
Healthcare Analytics, vol. 1, 2024, Article 100350, Elsevier BV.

[3]Rakesh Ranjan, Bikash Chandra Sahana, and Ashish Kumar Bhandari, "Deep Learning Models for Diagnosis of Schizophrenia Using EEG Signals: Emerging Trends, Challenges, and Prospects,"
Archives of Computational Methods in Engineering, vol. 31, no. 4, pp. 2345–2384, 2024. doi: 10.1007/s11831-023-10047-6

[4]Shijie Jiang, Qiyu Jia, Zhenlei Peng, Qixuan Zhou, Zhiguo An, Jianhua Chen & Qizhong Yi,
"Can artificial intelligence be the future solution to the enormous challenges and suffering caused by Schizophrenia?"
Schizophrenia, vol. 11, Article number: 32, 2025.

[5]Jianjun Weng, Xiaoli Zhu, Yu Ouyang, Yanqing Liu, Hongmei Lu, Jiakui Yao & Bo Pan,
"Identification of Immune-Related Biomarkers of Schizophrenia in the Central Nervous System Using Bioinformatic Methods and Machine Learning Algorithms,"
Molecular Neurobiology, vol. 62, no. 3, pp. 3226–3243, 2025.

[6]Gayathri Sunil, Smruthi Gowtham, Anurita Bose, Samhitha Harish & Gowri Srinivasa, "Graph neural network and machine learning analysis of functional neuroimaging for understanding schizophrenia," *BMC Neuroscience*, vol. 25, Article number: 2, published 02 January 2024

[7]Sweta Bhadra, Chandan Jyoti Kumar & Dhruba Kr Bhattacharyya, "Multiview EEG signal analysis for diagnosis of schizophrenia: An optimized deep learning approach," *Multimedia Tools and Applications*, 2024 Sep 20; vol. 83, pp. 79–89.