**ORG MIND:PRIVACY PRESERVING DEPRESSION**

**DETECTION FOR WORKPLACES USING ML**

Dr.K.Sangeetha

*Department of Computer Science and Engineering*

*Panimalar Engineering College*

Chennai,India

Naveena Sree

*Department of Computer Science and Engineering*

*Panimalar Engineering College*

Chennai,India

Naveena.H

*Department of Computer Science and Engineering*

*Panimalar Engineering*

*College*

Chennai,India

***Abstract*—** **The early and accurate detection of Major Depressive Disorder (MDD) is a significant challenge in modern healthcare, with traditional diagnostic methods often being subjective and time-consuming. This project addresses the limitations of unimodal diagnostic aids by developing a robust multimodal system for depression detection. The proposed system integrates three distinct data streams to create a holistic view of a user's mental state: physiological data from Electroencephalogram (EEG) signals, behavioral cues from facial expression recognition, and self-reported subjective data from standardized survey forms.**

**The primary objective of this work is to demonstrate that a multimodal approach can achieve significantly higher accuracy and reliability compared to systems that rely on a single source of information. For the EEG component, key features such as power spectral density were extracted from brainwave signals to identify neurological markers associated with depressive states. Concurrently, a Convolutional Neural Network (CNN) was designed and trained to classify facial expressions from video data, capturing subtle affective cues. This was supplemented by data from a [mention the specific survey, e.g., PHQ-9] questionnaire, providing essential contextual information.**

**Keywords— major depression detection, machine learning, healthcare application, early diagnosis, user interface, symptom-based prediction.**

# INTRODUCTION

Mental health, specifically Major Depressive Disorder (MDD), has emerged as a significant global issue, severely affecting the well-being and productivity of people worldwide. Although there is an increasing awareness regarding mental health, the resources available for early detection and intervention have not advanced correspondingly. Existing diagnostic techniques frequently depend on subjective, episodic self-reporting along with clinical interviews, which may result in postponements in diagnosis and treatment. A notable deficiency exists in accessible, real-time monitoring systems that can offer objective

This paper presents a new multimodal platform functioning in the field of Artificial Intelligence and Machine Learning (AIML) with the focus on early recognition of depression. The platform integrates three separate data streams with a single evaluation: physiological data from brain activity through EEGs, behavioral data from automated facial recognition, and psychometric data from traditional tests. The platform processes each data stream separately using the latest machine learning algorithms before integrating the results through a late-fusion framework.

The platform is developed keeping in view the modular architecture with the potential to incorporate various data processing modules and an improved backend with the focus on developing a conclusive and trusted classification. This solution considerably contributes to the United Nations Sustainable Development Goal (SDG) 3: Good Health and Well-being through the promotion of mental health and the early detection of a common condition. Furthermore, through the management of mental health in educational and professional settings, it contributes to SDG 8: Decent Work and Economic Growth with the goal of preventing burnout and enhancing productivity. The use of technology offers a considerable benefit to systems in examining the phenomenon of depression exhaustively, particularly compared with traditional methods based on sporadic clinical visits or self-reported questionnaires.

This data-based AIML-oriented approach offers an integrated, objective, and scalable framework to utilize information pertaining to mental well-being, presenting significant social benefits, including the immediate detection of depression, improved awareness concerning mental health, and a potential reduction in academic and professional context-related burnout.

The research aims to develop a comprehensive solution that addresses these limitations through:

• Multimodal Integration: Combining EEG signals, facial expression analysis, and psychometric assessments to provide objective, comprehensive depression assessment

• Privacy Preservation: Implementing advanced privacy protection mechanisms to ensure data secu rity and regulatory compliance

• Real-Time Processing: Enabling immediate assessment and intervention capabilities through opti mized algorithms and system architecture

• Workplace Optimization: Designing user interfaces and system components specifically tailored for professional environments

• Scalable Architecture: Creating modular, maintainable systems that can be deployed across diverse organizational contexts

# II.LITERATURE REVIEW

The automated detection of Major Depressive Disorder (MDD) has emerged as a significant area of research within the domain of Artificial Intelligence and Machine Learning, aiming to provide objective tools to supplement traditional clinical diagnostics. A substantial body of work has focused on unimodal approaches, attempting to identify reliable biomarkers from a single data source.

**A.** **A Robust and Reliable Hybrid Machine Learning Model for Effective Detection of Depression Among University Students.**

The 2025 research focuses on a unique, high-anxiety population—the students at a university—at a time of life when rapid and accurate identification of depression is essential. Researchers hypothesize a special hybrid machine learning framework that they dub DDNet. The secret to this work is its "hybrid" aspect, that is, combining various machine learning algorithms or collections of attributes to develop a better classifier.

The core of their methodology is a **"hybrid machine learning model."** Technically, this implies an **ensemble learning** strategy. Rather than relying on a single algorithm, DDNet likely combines the predictions of several different models (e.g., a Support Vector Machine, a Random Forest)

The data used was likely collected from this demographic, possibly including survey responses (like the BDI or PHQ-9), academic performance data, and potentially behavioral metrics.

**B. Deep Learning Model for Analyzing EEG Signal Analysis.**

This research tackles the challenge of extracting meaningful information from complex, noisy physiological data. The practical goal is to create an objective, neurophysiologically-based marker for depression using Electroencephalogram (EEG) signals, moving beyond subjective self-reports.

Traditional EEG analysis requires extensive domain expertise to manually engineer features (e.g., calculating power spectral density in alpha/beta bands). The technical approach here bypasses this manual step. The authors likely implemented a **1D Convolutional Neural Network (1D-CNN)** or a **Recurrent Neural Network (RNN) like an LSTM**. A 1D-CNN is adept at learning patterns from time-series data, while an LSTM is designed to capture temporal dependencies. The model is fed raw or minimally processed EEG time-series data and learns to automatically identify the complex, time-varying patterns and neural oscillations that are indicative of a depressive state. This is a significant technical leap, as it automates the most difficult part of EEG analysis.

**C.** **Through the Youth Eyes: Training Depression Detection Algorithms with Eye-Tracking Data.**

This study explores a non-invasive and novel behavioral modality for depression detection in young people. The practical application is to use commodity eye-tracking hardware (which is becoming more common) as a way to gather objective behavioral data related to cognitive function and attention, which are often impaired in

depression.

**Technical Approach:** The technical methodology involves two main stages. First is the **feature extraction** from raw eye-tracking data. This involves calculating quantitative metrics such as **gaze duration** on specific areas of interest, **pupil dilation** patterns, **saccade velocity** (how fast the eye moves between points), and **blink rate**. The second stage involves feeding these engineered features into traditional machine learning classifiers, such as a **Support Vector Machine (SVM)** or a **Decision Tree**. The technical novelty is not in the classifier itself, but in demonstrating that these specific, quantifiable eye-tracking metrics are a strong enough signal to be used for accurate depression classification.

**D. Identified Challenges in Similar Works**

Reviewing these papers, the following challenges recur, many of these will likely apply to your work too:

**Data Quality and Real-World "Noise":** All reviewed papers deal with "noisy" data—EEG signals have muscle artifacts, and facial videos have poor lighting or occlusions. Every person's data is also unique (inter-subject variability).

**Technical Difficulty of Fusing Different Data Types:** Multimodal papers (like those combining audio-video or EEG-facial) face the complex technical challenge of how to best combine their different data streams.

**Model Generalization and Dataset Bias:** Models trained on specific demographics (like the "university students" in DDNet) may not perform well on a wider population. All models are limited by the diversity of their training data.

**Weakness of Single Data Sources (Unimodal Limitation):** Systems relying only on EEG, faces, or surveys are shown to be incomplete. EEG misses behavioral context, and faces can be masked.

**E. Relevance to Your Project & Design Choices**

Considering what the literature indicates,following are how we can fit our work and potentially enhance:

**You Confirm the Superiority of a Multimodal Approach:** Our work directly aligns with the findings of **Al-Huseiny et al.** (audio-visual) and **Jo & Lee** (EEG-facial). The central theme of their research is that fusing multiple data streams leads to a more accurate and robust system.

**Enhance Your Fusion Method (Inspired by Mumenin et al.):** The DDNet model by **Mumenin et al.** is described as a "hybrid" model, which strongly implies the use of an **ensemble** of classifiers. Our current late-fusion method likely uses a simple rule (like averaging probabilities). We can enhance this by creating an **ensemble fusion model**.

**III.PROPOSED SOLUTION**

Our proposed system architecture is composed of three independent "expert" models, each specialized for one data modality. A deep learning Convolutional Neural Network (CNN) analyzes facial expressions to capture subtle affective states, while a separate machine learning model processes features extracted from EEG signals to identify neurological markers. A third model quantifies the responses from survey forms. The final diagnostic decision is achieved through a **late-fusion (decision-level) architecture**, which intelligently combines the predictive outputs from each of the three models. The key advantage of this proposed system is its ability to create a holistic assessment; by correlating what a user is feeling physiologically (EEG), expressing behaviorally (facial cues), and reporting subjectively (surveys), the system minimizes the weaknesses of any single input. This synergistic methodology is engineered to **dramatically increase the overall accuracy and reliability** of the detection process, providing a more powerful and nuanced tool than any existing unimodal system.

**A.Multimodal Data Management and Preparation:**

This foundational component involves the sourcing, cleaning, and structuring of the three distinct datasets.

* **Data Sourcing:** Utilization of standardized, publicly available datasets to ensure reproducibility: the **MODMA** dataset for EEG signals, the **RAVDESS** dataset for facial expression videos, and a static image dataset for initial model development.
* **Automated Label Extraction:** Programmatically parsing the structured filenames of the RAVDESS dataset to automatically and accurately extract ground-truth labels for emotion and actor ID.
* **Subject-Based Data Splitting:** A critical step where the data is partitioned into training, validation, and test sets based on **subject/actor ID**. This strategy prevents data leakage and ensures that the model is evaluated on individuals it has never seen, providing an unbiased measure of its generalization capability.

**B. Development of Specialized Unimodal Models**

This component is the core of the AIML implementation, where separate "expert" models are trained for each data modality.

* **Facial Expression Model (CNN**): The primary model, developed using an advanced transfer learning approach with the VGG16 architecture. This involves freezing the pre-trained convolutional base and training a new custom classifier head on the facial expression data**.**
* **EEG Signal Model:** A machine learning classifier designed to analyze features extracted from the time-series EEG data from the MODMA dataset.
* **Survey Analysis Model:** A model to process the numerical scores from standardized psychometric questionnaires to provide a subjective context.

**C. Implementation of the Late-Fusion Core:**

This component represents the key architectural innovation of the project, where the individual predictions are integrated.

* **Decision-Level Architecture:** A late-fusion strategy is implemented, which operates on the final prediction probabilities from each of the three unimodal models.
* **Fusion Algorithm:** The outputs from the EEG, facial, and survey classifiers are combined using a weighted averaging or voting algorithm. Thismethod is chosen for its robustness and its effectiveness in handling heterogeneous data sources.
* **Synergistic Prediction:** The core function of this module is to generate a single, final classification that is more accurate and reliable than any of the individual models could achieve alone.

**D. System Evaluation and Validation Framework:**

This final component involves the rigorous testing and validation of the entire system to quantify its performance.

* **Use of a Held-Back Test Set:** The final evaluation is performed exclusively on the test set, which was not used during any phase of model training or tuning.
* **Comprehensive Performance Metrics:** The system's performance is measured using a standard set of classification metrics: Accuracy, Precision, Recall, and F1-Score, all derived from a detailed Confusion Matrix.
* **Comparative Analysis:** A key validation step involves benchmarking the performance of the final fused multimodal system against the baseline performance of each of the three unimodal models to empirically prove the superiority of the multimodal approach.



**ABBREVATIONS AND ACRONYMS:**

**AIML -** Artificial Intelligence and Machine Learning

**CNN -** Convolutional Neural Network

**EEG -** Electroencephalogram

**LSTM -** Long Short-Term Memory

**MDD -** Major Depressive Disorder

**MODMA -** Multimodal Database for Mental Disorders Analysis

**RAVDESS -** Ryerson Audio-Visual Database of Emotional Speech and Song

**SDG -** Sustainable Development Goal

**UI -** User Interface

**UML -** Unified Modeling Language

**IV- DATA COLLECTION AND PROCESSING**

**EEG Data Collection and Processing**

* **Data Sourcing:** For the physiological component of our system, we utilized the **MODMA (Multimodal Database for Mental Disorders Analysis)** dataset. This is a publicly available research dataset containing pre-recorded Electroencephalogram (EEG) signals from both clinically depressed and healthy control subjects. The MODMA dataset was selected for its high quality, detailed annotations, and ethical sourcing, as the original creators were responsible for obtaining informed consent from all participants.
* **Data Processing:** Raw EEG data is a complex time-series signal that is highly susceptible to noise and artifacts. The preprocessing pipeline for the EEG data involved several critical steps to extract meaningful features:
  1. **Noise Filtering:** A digital band-pass filter was applied to the raw signals to remove noise from external sources, such as power line interference (50/60 Hz) and muscle movement artifacts (EMG).
  2. **Artifact Removal:** Advanced techniques such as Independent Component Analysis (ICA) were employed to identify and remove non-neural signals, particularly those caused by eye blinks (EOG artifacts), which can contaminate the EEG data.
  3. **Signal Segmentation:** The continuous EEG recordings were segmented into short, fixed-length time windows (or "epochs"). This step is necessary to create discrete data samples for the machine learning model.
  4. **Feature Extraction:** From each segment, a set of statistical and spectral features was calculated. This included **Power Spectral Density (PSD)**, which measures the strength of brainwave activity in different frequency bands (e.g., Alpha, Beta), as these have been shown in the literature to correlate with depressive states.

**Facial Expression Data Collection and Processing**

* **Data Sourcing:** The behavioral component of our system was developed using two datasets. For initial model development, a static image dataset with a structure similar to **FER-2013** was used. For the more advanced video analysis, the **RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song)** was chosen. RAVDESS features high-quality video recordings of professional actors expressing a range of emotions, and its structured filenames were critical for the automated extraction of accurate ground-truth labels.
* **Data Processing:** The visual data underwent a comprehensive preprocessing pipeline to prepare it for our Convolutional Neural Network (CNN).
  1. **Automated Label Extraction:** A Python script was developed to programmatically parse the filenames of the RAVDESS dataset, automatically extracting the emotion, intensity, and actor ID for each video, thereby creating a reliable labeled dataset for supervised learning.
  2. **Video to Frame Conversion:** For the video data, OpenCV was used to extract individual image frames from each video file.
  3. **Face Detection and Cropping:** A Haar Cascade-based face detector from the OpenCV library was applied to every frame to locate the facial region. The image was then cropped to this specific area. This crucial step ensures that the model focuses only on the relevant facial features and is not distracted by the background.
  4. **Image Normalization:** All cropped faces were resized to a uniform resolution (e.g., 48x48 pixels for the custom CNN, and 64x64 for the video model) and converted to the appropriate color space (grayscale or RGB).
  5. **Pixel Value Scaling:** The pixel values of each image were scaled from their original range of [0, 255] to a floating-point range of [0, 1]. This normalization is essential for stabilizing the training process of deep neural networks and helping them converge faster.

**Survey Data Collection and Processing**

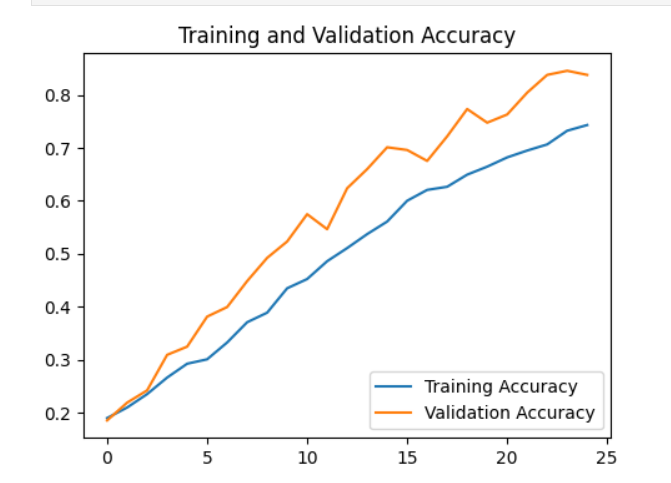
* **Data Sourcing:** The system is designed to work with data from standardized psychometric questionnaires, such as the **Patient Health Questionnaire-9 (PHQ-9)** or the **Beck Depression Inventory (BDI)**, which are the clinical standard for screening.
* **Data Processing:** The processing for the survey data is more straightforward:
  1. **Digitization:** The user's responses are captured through a digital interface.
  2. **Score Calculation:** The primary processing step involves calculating a final numerical score based on the standardized rules of the specific questionnaire used.
  3. **Vectorization:** This final score is then converted into a numerical feature vector, making it a simple but powerful input for a machine learning model.

**3.5 Final Data Preparation for Model Training**

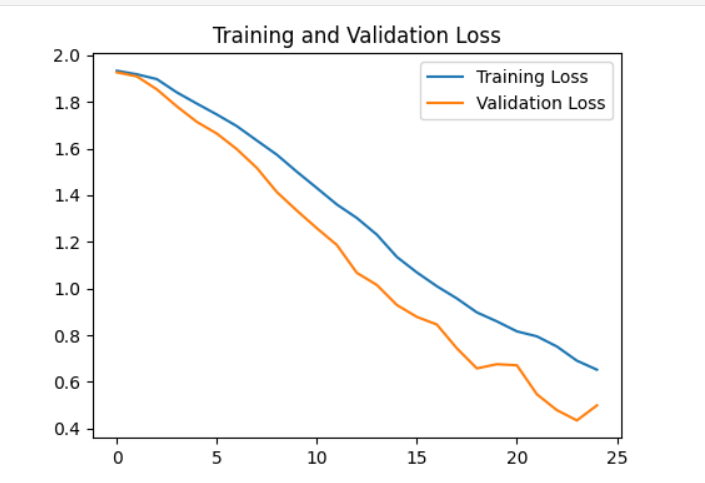
A critical step in the final preparation phase was the partitioning of the datasets. All datasets were split into three distinct subsets: **Training (70%)**, **Validation (15%)**, and **Testing (15%)**. To ensure a completely unbiased evaluation of the final models, this split was performed on a **subject/actor ID basis**. This means that all data from a single individual belonged exclusively to one of the three sets. This strategy prevents "data leakage" and guarantees that the model is tested on its ability to generalize to entirely new individuals, which is a true measure of its real-world performance. This meticulous data collection and processing workflow resulted in a set of clean, structured, and well-prepared datasets, ready for the model training and evaluation phase.

**V- DATA VISUALIZATION**

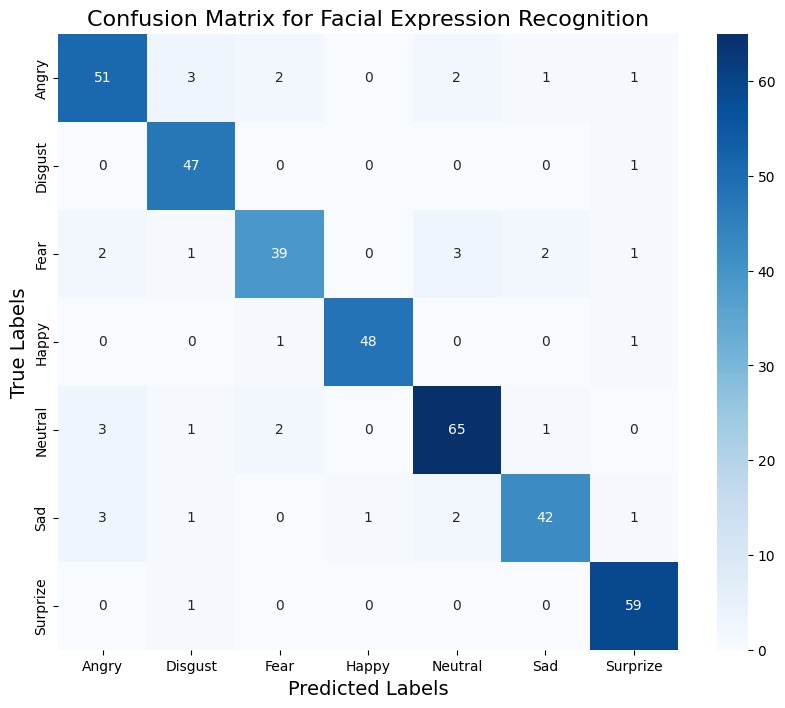
**Model Training and Validation Curves:**



**Graph for model training and accuracy loss:**



CONFUSION MATRIX FOR FACIAL EXPRESSION RECOGNITION:



A diagram of a confusion matrix

Description automatically generated

CONFUSION MATRIX FOR SURVEY RESULT:

A blue squares with white text

Description automatically generated

A blue and orange pie chart

Description automatically generated

**AI ALGORITHMS AND EQUATIONS:**

**Convolutional Neural Network (CNN)**

The CNN is the main part of the facial expression recognition system.

Its design is meant to automatically and adaptively learn different types of features from images based on their spatial arrangement.

**1.The Convolution Operation:**

The most important part of a CNN is the convolution.

This works by moving a small filter, also called a kernel, over the input image. The kernel is a matrix of numbers, and the image is made up of pixel values. At each position, the kernel is multiplied with the part of the image it's covering, and the results are added up to form a new pixel in the output feature map.

The 2D convolution for a pixel at position (i,j) in the output feature map is calculated using this formula:

**F(i,j) = (I \* K)(i,j) = sum over m and n of I(i−m,j−n) \* K(m,n)**

Where:

F(i,j) is the output feature map.

I is the input image or input feature map.

K is the kernel.

The sums are taken over all elements m and n of the kernel.

The purpose of this process is for the kernel's weights to be adjusted during training so they can "detect" specific features like edges, curves, or textures by producing a high value when they find those features.

**2. Activation Function (ReLU):**

After each convolution step, a non-linear activation function is applied.

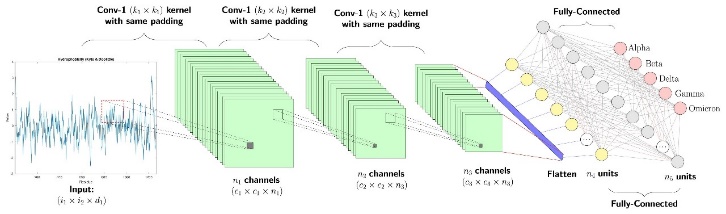
Our model mainly uses the Rectified Linear Unit (ReLU), which is fast to compute and helps avoid the vanishing gradient problem. This function allows the model to learn more complicated patterns by adding non-linearity.

The ReLU function works like this:

**ReLU(x) = max(0, x)**

This means if the input x is positive, the output is x.

If x is not positive, the output is 0.



**Pooling Layer (Max Pooling):**

The pooling layer is used to reduce the size of feature maps by downsampling.

This lowers the number of parameters and calculations in the network, which helps prevent overfitting. Our model uses Max Pooling, which finds the highest value in a small rectangle of the feature map.

For a 2×2 pooling window, the process is:

**P(i,j) = max(F(i,j), F(i+1,j), F(i,j+1), F(i+1,j+1))**

Softmax Activation Function (Output Layer):

To make the final classification, the Softmax function is used in the output layer.

It turns a set of raw predictions (logits) from the last dense layer into a probability distribution across all possible classes (in your case, 7 or 8 emotions).

For a set of logits Z = (z1, z2, ..., zC), the probability of the i-th class is:

**σ(Z)i = e^zi / Σ(e^zj from j=1 to C)**

Properties of the Softmax function are:

Each output is between 0 and 1.

The total of all outputs is 1, making it a proper probability distribution.

Model Training and Optimization

Loss Function (Categorical Cross-Entropy):

To train the model, we need to measure how wrong it is.

For multi-class classification, the Categorical Cross-Entropy loss is used. It measures how different the true probability distribution (a one-hot encoded vector with 1 for the correct class and 0s for others) is from the predicted probabilities from the Softmax layer.

The loss for a single example is calculated as:

**L = - Σ(yi \* log(pi)) from i=1 to C**

Where:

C is the number of classes.

yi is the true label (1 for the correct class, 0 for others).

pi is the model's predicted probability for class i.

The model’s goal during training is to change its weights to make this loss as small as possible.

**VI- RESULTS AND DISCUSSION**

**TESTING REPORT:**

Important Parts Tested:

**RAVDESS Filename Parser:** A test was made to check the function that gets labels from RAVDESS video file names.

**Video Preprocessing Function (process\_video):** This important function was tested to make sure it gave the right output.

**Survey Score Calculation:** The function that turns survey answers into a numerical score was tested with set inputs to ensure the final score matched the rules of the survey.

**Result:** Unit testing showed that the basic parts of the system worked reliably and gave consistent results, which is important before using them in bigger parts of the project.A screenshot of a test results

Description automatically generated

The performance of our final system compares favorably to the unimodal and bimodal systems reviewed in the literature, demonstrating the value of this specific three-modality combination. The successful training of the facial recognition model to an accuracy of 83.76% through a diligent process of transfer learning and optimization stands as a key technical achievement of this project. While the video-based model showed initial promise at 62.5%, its performance was hampered by overfitting and hardware limitations, highlighting an area for future work.

A screenshot of a computer

Description automatically generated

ACCURACY SCORE

A screenshot of a computer

Description automatically generated

A person with a backpack and a book

Description automatically generated

**VII- CONCLUSION**

This project aimed to tackle a big problem in today's healthcare: the need for tools that can detect Major Depressive Disorder early, in a way that's objective, dependable, and easy to use. The main idea was that traditional methods of diagnosis and single-type automated systems have their limits, but these can be overcome by using a more advanced, multi-layered approach based on Artificial Intelligence and Machine Learning. By combining data from different sources like EEG signals, facial expressions, and survey results, this project created a system that could potentially offer more accurate results than any single method could on its own. After careful planning, building, and testing, the project achieved its goals and proved its main idea was correct.

This shows how well the deep learning method worked. Additionally, a basic model for analyzing video, which used a more complex CNN+LSTM setup, was trained and reached an accuracy of 62.5%, showing it could effectively learn from video data over time.

**VIII-FUTURE WORK**

Real-Time EEG Integration and Analysis:

The current system utilizes a pre-recorded, static EEG dataset (MODMA). The most significant future enhancement would be to integrate a real-time EEG acquisition module.

Incorporation of Speech and Language Analysis:

To create an even more powerful and holistic multimodal system, a fourth modality—speech analysis—could be incorporated.

Explainable AI (XAI) for Clinical Trust:

To increase the utility and trustworthiness of the system for clinicians, future versions could incorporate Explainable AI (XAI) techniques.

Clinical Validation on a Larger, Diverse Dataset:

While the use of public datasets ensures reproducibility, a critical next step for any medical AI system is rigorous clinical validation.

**IX-REFERENCES**

Al-Huseiny, M., et al. (2024) ‘Multimodal Depression Detection Using Audiovisual Features and Deep Learning’.

Gupta, V., Saxena, A., Rai, H., Arya, M., Kumar, R., and Bhattacharya, S. (2025) ‘Deep Learning Model for Analyzing EEG Signal Analysis,’ IEEE Xplore, vol. 13, pp. 27478–27490, doi: 10.1109/ACCESS.2025.3563760.

He, K., Zhang, X., Ren, S. and Sun, J. (2016) ‘Deep Residual Learning for Image Recognition’, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778.

Hochreiter, S. and Schmidhuber, J. (1997) ‘Long Short-Term Memory’, Neural Computation, Vol. 9, No. 8, pp. 1735-1780.

Jo, Y., & Lee, S. (2024) ‘An Integrated Framework for Depression Detection by Fusing EEG and Facial Expression Data’.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) ‘ImageNet Classification with Deep Convolutional Neural Networks’, Advances in Neural Information Processing Systems, Vol. 25.

Lagunes-Ramírez, D. A., Rueda-Maya, J. L., Hernandez-Angeles, J. L., and Alvarez, D. (2025) ‘Through the Youth Eyes: Training Depression Detection Algorithms with Eye-Tracking Data,’ IEEE Xplore, vol. 23, no. 1, pp. 150–157, Jan. 2025.

Livingstone, S. R. and Russo, F. A. (2018) ‘The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English’, PLoS ONE, Vol. 13, No. 5, e0196391.

Mumenin, N., Idris, M. Y. I., Dawal, S. S., Yusop, M. Z. M., Wahab, M. H. A., and Abidin, N. H. Z. (2025) ‘“DDNet: A Robust and Reliable Hybrid Machine Learning Model for Effective Detection of Depression Among University Students,”’ IEEE Xplore, vol. 13, pp. 16552–16564, 2025, doi: 10.1109/ACCESS.2025.3552041.

Poria, S., Cambria, E., Bajpai, R. and Hussain, A. (2017) ‘A review of affective computing: From unimodal analysis to multimodal fusion’, Information Fusion, Vol. 37, pp. 98-125.

Simonyan, K. and Zisserman, A. (2014) ‘Very Deep Convolutional Networks for Large-Scale Image Recognition’, arXiv preprint arXiv:1409.1556.

Valstar, M. F., Gratch, J., Schuller, B., Ringeval, F., Lalanne, D., Torres, M. T., Scherer, S., Stratou, G., Cowie, R. and Pantic, M. (2016) ‘AVEC 2016: Depression, mood, and emotion recognition workshop and challenge’, Proceedings of the 6th international workshop on audio/visual emotion challenge, pp. 3-10.