International Journal of Leading Research Publication (IJLRP)

E-ISSN: 2582-8010 ● Website: www.ijlrp.com ● Email: editor@ijlrp.com

Multimodal Depression Detection

An Integrated Multimodal Framework for Automated Depression Severity Assessment

Anuj Sahu[1], Chandra Prakash Singar[2], Puja Gupta[3], Divyansh Ganote[4], Sanjay Patil[5]

[1,2,3,4,5]Department of Information Technology, Shri G.S. Institute of Technology & Science, Indore.

Abstract:
Depression remains one of the most prevalent mental health disorders globally, yet early detection remains challenging due to its multifaceted and heterogeneous clinical presentation [1]. This study presents a novel multimodal depression detection system that integrates five distinct analytical modalities: facial expression analysis, vocal tone assessment, video-based emotion recognition, text sentiment analysis, and clinical questionnaire evaluation. Using a probability-weighted scoring mechanism combined with an unequal weighting scheme that prioritizes explicit user input, the system generates a comprehensive depression severity score on a 0–100 scale with clinically relevant classification thresholds. The architecture leverages convolutional neural networks (CNNs) [4] for image and video processing, long short-term memory (LSTM) networks
[5] for audio analysis, a trained sequence model for text classification, and large language model (LLM) integration via the Groq API for open-ended questionnaire assessment. By combining noisy ambient snapshots with explicit clinical input, the system achieves a holistic psychological profile that extends beyond single-modality approaches [8]. This paper describes the technical implementation, mathematical framework, and clinical classification methodology underlying this integrated system.

Keywords: depression detection, multimodal analysis, machine learning, emotion recognition, clinical assessment, CNN, LSTM, NLP, LLM.

1. Introduction

Depression is a leading cause of disability worldwide, affecting over 280 million people globally [1] and representing a significant public health burden. Despite its prevalence, many cases remain undetected or under-diagnosed, particularly in the early stages when intervention is most effective. Traditional diagnostic approaches rely primarily on clinical interviews and standardized questionnaires such as the Beck Depression Inventory (BDI) [2] and the Patient Health Questionnaire (PHQ-9) [3], which are subject to recall bias, social desirability bias, and the limited temporal resolution of periodic assessments.

Recent advances in machine learning and multimodal signal processing have created opportunities to develop automated screening tools that capture emotional state through multiple channels simultaneously [8]. Each modality—facial expressions, vocal tone, video behavior, linguistic content, and explicit self-report—provides complementary information about depressive symptomatology. However, previous systems have typically focused on individual modalities in isolation, losing the rich contextual information available when multiple signals are integrated [8].

The central innovation of this work is a probability-weighted aggregation framework that combines five distinct modalities using unequal weighting that reflects the reliability and clinical relevance of each

International Journal of Leading Research Publication (IJLRP)
E-ISSN: 2582-8010 ● Website: www.ijlrp.com ● Email: editor@ijlrp.com

input. By assigning greater weight to explicit, conscious communication (questionnaires and text) while incorporating ambient behavioral signals (facial expressions, vocal patterns, and video), the system achieves a more robust and interpretable depression severity assessment.

This paper describes the complete technical architecture, the mathematical framework for score calculation, and the clinical classification thresholds that translate the integrated score into actionable recommendations for users. The system was developed and is publicly available at the GitHub repository referenced above.

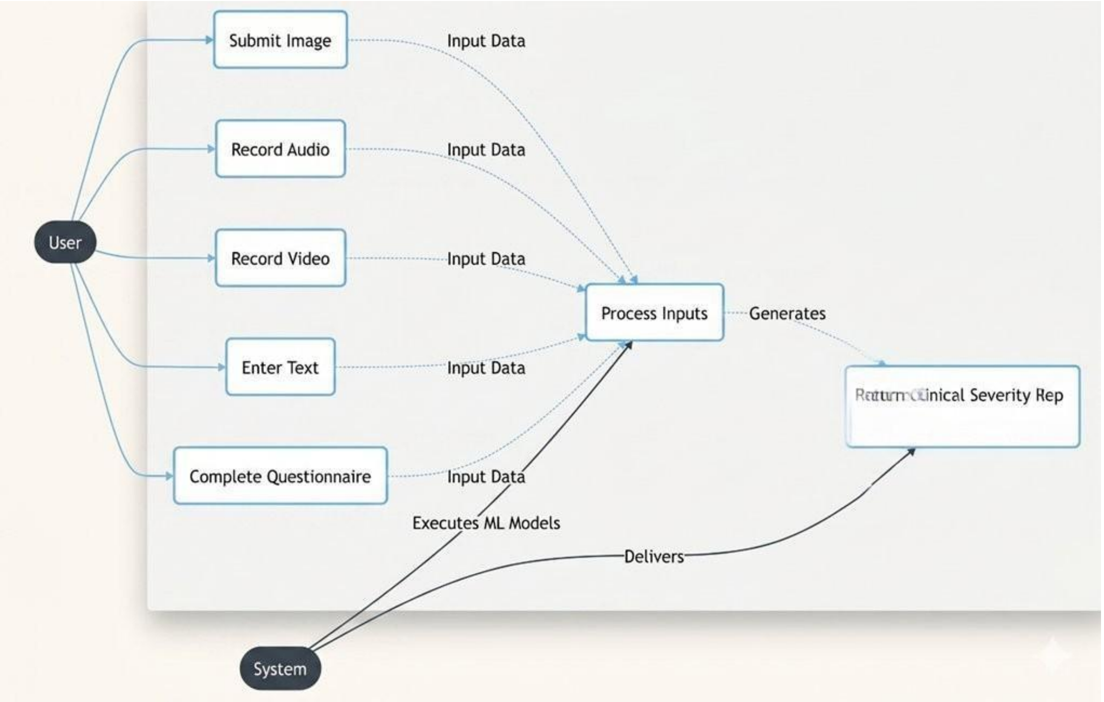

Figure 1:User interaction flows and system processing pipeline showing how each input modality feeds into the central processing engine and generates the clinical severity report.

2. Literature Review

2.1 Depression Detection and Assessment

Clinical assessment of depression has traditionally relied on structured diagnostic criteria and validated questionnaire instruments. The Beck Depression Inventory (BDI) [2] and the Patient Health Questionnaire (PHQ-9) [3] represent gold-standard approaches for quantifying depressive symptom severity. However, these instruments are subject to several limitations: they require explicit patient cooperation, are vulnerable to reporting bias, and provide only periodic snapshots rather than continuous monitoring.

Recent work has explored automated screening approaches using machine learning. Text-based approaches have shown promise in identifying depressive language patterns through sentiment analysis and neural language models. Audio-based systems have leveraged prosodic features (pitch, intensity, duration) to detect depression-related vocal changes [8]. Facial expression recognition systems have demonstrated correlations between emotion predictions and depressive mood states [6]. However, these modalities have largely been explored independently.

2.2 Multimodal Emotion Recognition

Multimodal emotion recognition represents an established field within affective computing, as pioneered by Picard [7]. Research has consistently demonstrated that integrating multiple modalities yields superior performance compared to single-modality approaches [8]. The complementary nature of facial, vocal, and linguistic signals means that failures in one modality can be compensated by reliable signals in another. Baltrušaitis et al. [8] provide a comprehensive survey and taxonomy of multimodal machine learning, identifying key challenges including modality alignment, fusion strategies, and cross-modal translation.

Key challenges in multimodal systems include temporal alignment of asynchronous data streams, fusion of heterogeneous feature spaces, handling missing or noisy modality data gracefully, and designing interpretable scoring mechanisms for clinical use.

2.3 Deep Learning for Emotion Recognition

Convolutional neural networks (CNNs) [4] have become the standard approach for facial expression and video-based emotion recognition, effectively learning spatially-localized features that correspond to facial muscle movements. The seminal work of LeCun, Bengio, and Hinton [4] established the foundation for applying deep learning to perceptual tasks. Long short-term memory (LSTM) networks [5] have proven effective for audio analysis by capturing temporal dependencies in acoustic features such as mel-frequency cepstral coefficients (MFCCs). The RAVDESS dataset [6] has been widely used to train and validate audio-based emotion recognition models.

For text-based emotion analysis, both rule-based approaches (e.g., VADER sentiment analysis) and neural sequence models have demonstrated effectiveness. Recently, large language models have shown remarkable capability in zero-shot and few-shot emotion classification tasks when provided with appropriate prompts and domain context, making them suitable for interpreting open-ended clinical responses [8].

3. System Architecture and Methodology

3.1 Overview

The multimodal depression detection system consists of five parallel processing pipelines, each designed to extract emotional information from a distinct data modality. These pipelines operate independently and generate continuous probability distributions over emotional categories. The outputs are then aggregated using a mathematically principled probability-weighted scoring approach followed by weighting-based normalization to produce a unified depression severity score on a 0–100 scale.

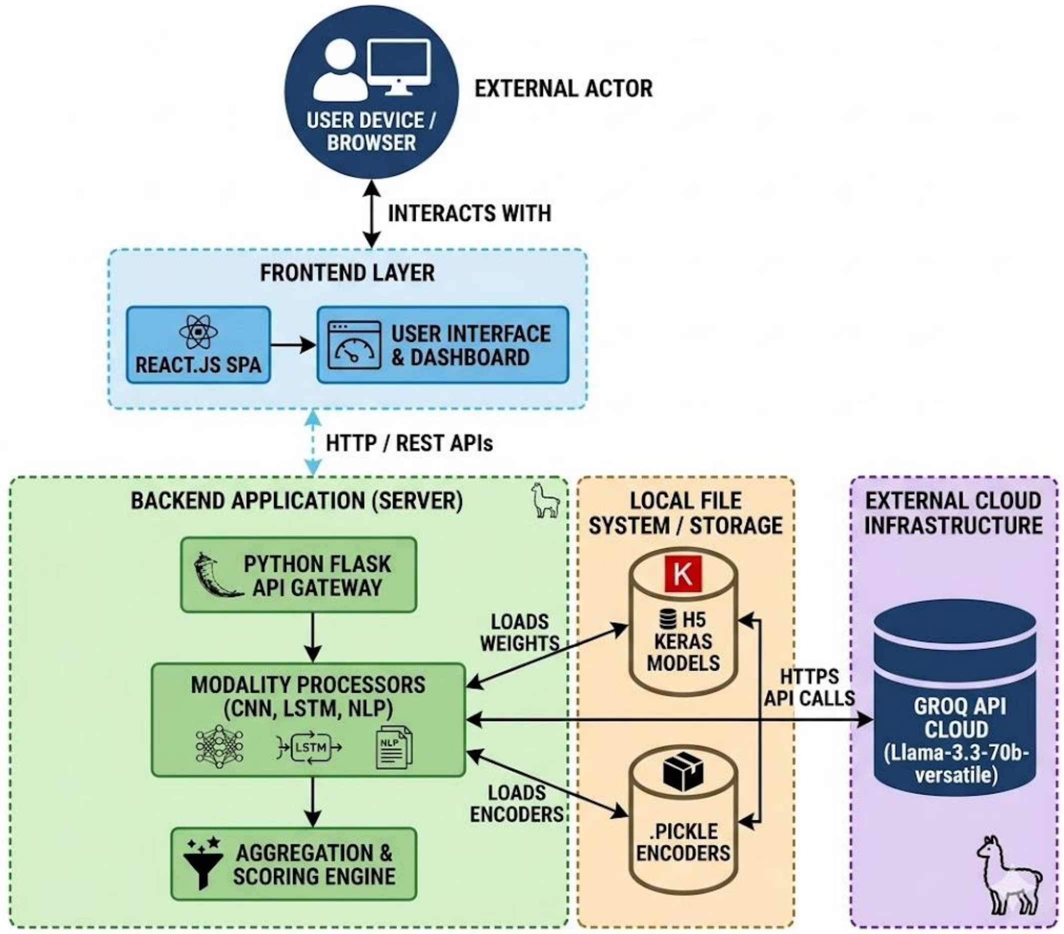

Figure 2: Shows the three-tier architecture comprising the React.js frontend layer, Python Flask backend with modality processors and aggregation engine, and external Groq Cloud infrastructure running Llama-3.3-70b-versatile.

The architecture, depicted in Figure 2, comprises three principal layers. The **Frontend Layer** is implemented as a React.js Single Page Application (SPA) providing the user interface and dashboard. The **Backend Application Server** is a Python Flask API Gateway that routes inputs to modality-specific processors (CNN, LSTM, NLP), performs aggregation and scoring, and manages communication with external services. The **External Cloud Infrastructure** hosts the Groq API running the Llama-3.3-70b-versatile large language model for semantic analysis of open-ended questionnaire responses.

3.2 Image Analysis Module

The image analysis module captures a static facial expression photograph. The captured image is converted to grayscale to reduce computational complexity and improve model generalization. The image is resized to 48×48 pixels—a standard resolution for facial emotion recognition [4]—and normalized to the range [0, 1] by dividing pixel values by 255.

A custom-trained convolutional neural network (*modelfiltercollab.keras*) processes the preprocessed facial image. The CNN learns to identify facial action units and their combinations, mapping visual features to discrete emotion categories. The model produces a probability distribution across seven basic emotions derived from Ekman's taxonomy of universal facial expressions [6]: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral.

3.3 Audio Analysis Module

The audio module records a brief voice sample from the user. The raw audio signal is processed using the librosa library to extract 40 mel-frequency cepstral coefficients (MFCCs)—a standard representation in audio processing that approximates human auditory perception by applying mel-scale frequency warping to the power spectrum. MFCCs have been extensively validated as effective features for emotion recognition in speech, as demonstrated by the RAVDESS corpus [6].An LSTM neural network [5] (*audioemotionlstm_model.h5*) processes the sequence of MFCC frames. The LSTM architecture is particularly well-suited to this task because it can learn long-range temporal dependencies in acoustic patterns that correlate with emotional states. The model generates probabilities across the same seven emotion categories as the facial expression module.

3.4 Video Analysis Module

The video analysis module captures a continuous webcam feed. Frame-by-frame face detection is performed using OpenCV's Haar Cascade classifier, a computationally efficient approach using a cascade of weak classifiers trained on positive and negative face exemplars. Each detected face frame is passed to a trained CNN (*videoemotionmodel.keras*) that outputs probability distributions over seven emotion categories. Temporal aggregation is performed by computing the mean of per-frame probability vectors, producing a temporally-smoothed estimate robust to transient expressions and detection errors.

3.5 Text Analysis Module

The text analysis module prompts the user to provide a free-text response describing their current emotional state or recent experiences. Text is preprocessed by stripping special characters and tokenizing into individual words. The resulting token sequence is padded to a fixed length using a trained tokenizer. A trained NLP sequence model (*EmotionRecognitionModel.keras*) processes the padded token sequences, learning distributed representations of emotional language and generating probability distributions over the seven emotion categories.

3.6 Questionnaire Analysis Module

The questionnaire represents the most explicit form of clinical input and is therefore weighted most heavily in the final aggregation, consistent with best practices in multimodal clinical assessment [8].

Multiple-Choice Questions (MCQ)

Fifteen standardized clinical questions assess depressive symptomatology across key domains including sleep quality, energy levels, social engagement, concentration, and hopelessness—domains well-established in the PHQ-9 and BDI instruments [2, 3]. Responses are rated on a four-point scale: "Rarely" (0), "Sometimes" (1), "Often" (2), "Almost Always" (3). The raw MCQ score ranges from 0 to 45 and is normalized as follows:

Normalized MCQ Score = (Raw MCQ Score / 45) × 50

Open-Ended Questions and LLM Integration

Seven open-ended questions prompt the user to describe their emotional state, recent stressors, sleep quality, and motivation levels in their own words. Each open-ended response is sent to the Groq API running the *llama-3.3-70b-versatile* large language model, which acts as a mental health expert and evaluates each response on a 0–50 depression severity scale. The LLM is prompted to consider emotional language use, described functional impairments, reported duration of symptoms, and indicators of hopelessness or anhedonia. The average of the seven LLM evaluations constitutes the Open-Ended Score.

Fallback Mechanism

If the Groq API is unavailable, the system falls back to the VADER Sentiment Intensity Analyzer—a lexicon-based approach validated for social media and general text sentiment analysis. The negative sentiment intensity is scaled to the 0–50 range to provide a depression severity estimate.

Combined Questionnaire Score

Questionnaire Score = Normalized MCQ Score + Open-Ended Score

This yields a combined score ranging from 0 to 100, representing the explicit clinical assessment of depression severity.

3.7 Probability-Weighted Scoring for Modality Aggregation

The core innovation of this system is the probability-weighted scoring mechanism that converts emotion probability distributions into continuous depression severity estimates while preserving nuance and uncertainty. This approach draws on the expected-value formalism used in Bayesian decision theory [8].

Step 1: Depression Severity Mapping

The seven basic emotions [6] are mapped to a continuous depression severity scale as follows. This mapping reflects the clinical intuition that happiness indicates minimal depression, while sadness is the prototypical depressive emotion [1]:

Emotion	Severity Score
Happy / Joy	1
Surprised	2
Neutral	3
Angry	5
Disgust	6
Fearful	7
Sad	9

Table 1: Emotion-to-Depression Severity Mapping

Step 2: Expected Value Calculation

For each of the four behavioral modalities (Image, Audio, Video, Text), the model outputs a probability distribution $P(\text{emotion}_i)$ across the seven emotions. The depression severity score for that modality is calculated as the expected value [8]:

Modality Score = $\sum P(\text{emotion}_i) \times \text{severity\_score}_i$

For example, if the image analysis produces the following probabilities: Happy=0.05, Surprised=0.03, Neutral=0.10, Angry=0.15, Disgust=0.12, Fearful=0.20, Sad=0.35, the image modality score would be: (0.05×1) + (0.03×2) + (0.10×3) + (0.15×5) + (0.12×6) + (0.20×7) + (0.35×9) = **5.88**. This approach preserves information about the full distribution rather than selecting the single most likely emotion.

Step 3: Final Aggregation with Unequal Weights

The individual modality scores and questionnaire score are aggregated using a weighted combination. The weighting scheme prioritizes explicit, conscious input while incorporating ambient behavioral signals, consistent with the principle that self-reported clinical data carries higher reliability for depression assessment [3]:

Modality	Weight (%)	Raw Points	Rationale
Questionnaire	~44%	40	Explicit clinical self-report [2, 3]

Text Analysis	~22%	20	Conscious linguistic expression
Video Analysis	~17%	15	Temporal behavioral signal
Audio Analysis	~11%	10	Prosodic emotional signal [6]
Image Analysis	~6%	5	Static facial expression [4]
Total	**100%**	**90**	

Table 2: Modality Weights for Final Aggregation

Step 4: Normalization to 0–100 Scale
The raw aggregated score is computed as:
Raw Total = (Questionnaire Score / 100 × 40) + (Text Score / 9 × 20) + (Video Score / 9 × 15) + (Audio Score / 9 × 10) + (Image Score / 9 × 5)
The raw total (range 0–90) is then scaled to the final 0–100 scale:
Final Depression Severity Score = (Raw Total / 90) × 100

3.8 Clinical Classification Thresholds

The final 0–100 depression severity score is classified into four clinical categories. Thresholds were selected to align with established depression screening instruments [2, 3] while accounting for the multimodal nature of the assessment:

Score Range	Classification	Recommendation
0–39	Minimal or No Depression	Maintain healthy habits: regular exercise, sleep, social engagement
40–49	Mild Depression	Engage in positive activities; consider professional therapy
50–59	Moderate Depression	Consult a mental health professional for formal assessment and treatment planning
60–100	Severe Depression	Immediate consultation with a doctor is strongly recommended

Table 3: Clinical Classification Thresholds and Recommendations

4. Technical Implementation

4.1 Backend Architecture

The backend is implemented in Python using the Flask web framework, which provides lightweight routing and request handling suitable for serving predictions to the frontend. Model training and inference are performed using TensorFlow/Keras [4], providing a flexible and well-documented framework for defining and executing CNN and LSTM architectures. The scikit-learn library is used for auxiliary machine learning tasks such as feature scaling and normalization.
Computer vision operations—video frame extraction and face detection using Haar Cascades—are implemented with OpenCV. Audio signal processing for MFCC feature extraction uses the librosa library. Natural language processing leverages the NLTK toolkit for VADER fallback sentiment analysis and the Groq API for LLM-based evaluation using the llama-3.3-70b-versatile model.

4.2 Frontend Architecture

The frontend is implemented in React.js, a modern JavaScript library for building interactive user interfaces. React's component-based architecture facilitates modular development of the distinct assessment modules. The Axios library manages asynchronous HTTP requests between the frontend and backend. CSS and Tailwind CSS provide responsive design and visual consistency across different screen sizes. Assessment results are visualized as radar charts providing an intuitive multimodal view of the depression profile.

4.3 Data Flow

The system operates through the following sequence: (1) The user submits multimodal data via the React SPA; (2) Axios sends an HTTP POST request to the Flask API Gateway; (3) The API Gateway routes the payload to the appropriate modality processor; (4) Each processor extracts features and runs model inference (CNN for image/video, LSTM for audio, sequence model for text, LLM via Groq API for questionnaire); (5) Probability distributions are returned to the Aggregation Engine; (6) The Aggregation Engine applies probability-weighted scoring and the unequal weighting scheme; (7) The final 0–100 severity score and clinical classification are returned as a JSON response; (8) The React UI renders radar charts, the severity score, and actionable clinical recommendations.

4.4 Model Storage and Loading

Pre-trained model weights are stored in the local file system as .h5 Keras model files and .pickle encoder files, as shown in Figure 2. The backend loads these weights at startup to minimize inference latency. The questionnaire module routes open-ended responses to the Groq Cloud API via HTTPS, ensuring that the computationally intensive LLM inference is offloaded to cloud infrastructure.

5. Discussion

5.1 Advantages of Multimodal Integration

The multimodal approach offers several significant advantages over single-modality systems [8]:

Complementarity: Each modality captures different aspects of emotional expression. Facial expressions reflect automatic emotional reactions [6], vocal prosody reflects emotional tone independent of linguistic content [6], text reflects conscious expression, and questionnaires provide explicit self-assessment [2, 3]. Integration of these complementary signals produces a more complete picture than any single modality alone.

Robustness: Single-modality systems are vulnerable to modality-specific failures. For example, a user with Parkinson's disease may have reduced facial expressivity, leading to false negatives in face-based systems. Multimodal integration ensures that depression indicators detected in other modalities compensate for modality-specific limitations, a well-documented benefit of multimodal fusion [8].

Temporal Coverage: The system captures emotional information across multiple timescales—static snapshots (image), brief samples (audio), extended periods (video), and explicit self-reflection (questionnaire). This temporal diversity reduces vulnerability to transient mood fluctuations.

Interpretability: The probability-weighted approach maintains interpretability by explicitly modeling each emotion's contribution to the final score. Clinical practitioners can understand which modalities contributed most to the final assessment, supporting clinical decision-making.

5.2 Limitations and Future Work

Clinical Validation: The system has not been validated against a clinical reference standard or compared to gold-standard diagnostic interviews. Future work should include prospective validation studies comparing system outputs to clinician diagnoses using instruments such as the SCID or HDRS [1, 2, 3].

Model Generalization: The individual CNN and LSTM models were trained on specific emotion recognition datasets and may not generalize optimally to diverse populations with varying demographics, cultural backgrounds, or depressive presentations [8].

Fairness and Bias: Facial recognition systems are known to exhibit performance disparities across demographic groups. Future work should analyze system performance across different populations and implement bias mitigation strategies [8].

Temporal Dynamics: The current system treats each assessment as independent. Integrating longitudinal information from previous assessments could improve detection of changes in depressive symptoms over time—an important clinical use case for continuous monitoring [1].

Threshold Validation: The clinical classification thresholds were assigned based on clinical judgment. Data-driven optimization against clinical outcomes should inform threshold refinement [2, 3].

6. Conclusion

This work presents a comprehensive multimodal depression detection system that integrates facial expression, vocal tone, video analysis, text sentiment, and clinical questionnaires using a mathematically principled probability-weighted aggregation framework [8]. By assigning greater weight to explicit clinical input while incorporating ambient behavioral signals, the system achieves a holistic assessment that extends beyond single-modality approaches.

The technical implementation leverages established deep learning architectures—CNNs [4] for image and video processing, LSTMs [5] for audio analysis, sequence models for text classification—and modern large language models via the Groq API for semantic interpretation of open-ended questionnaire responses. The probability-weighted scoring approach preserves information about predictive uncertainty rather than collapsing to single maximum-likelihood predictions, and the unequal weighting scheme reflects the differential clinical reliability and significance of various inputs [2, 3].

Future work should focus on prospective clinical validation, comparison against gold-standard diagnostic instruments such as the PHQ-9 [3] and BDI [2], analysis of performance across demographic groups, and refinement of clinical classification thresholds based on empirical outcomes. With appropriate validation and refinement, multimodal depression detection systems such as this have the potential to significantly improve access to early identification and intervention for this prevalent and debilitating condition [1].

REFERENCES:

1. World Health Organization. (2023). *Depressive disorder (depression)*. WHO Fact Sheet. Retrieved from https://www.who.int/news-room/fact-sheets/detail/depression
2. Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.
3. Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure.
4. *Journal of General Internal Medicine*, 16(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x
5. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553),436–444. https://doi.org/10.1038/nature14539
6. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
7. Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3–4), 169–200. https://doi.org/10.1080/02699939208411068 Picard, R. W. (1997). *Affective Computing*. MIT Press.
8. Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607
9. Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). *PLOS ONE*, 13(5), e0196391. https://doi.org/10.1371/journal.pone.0196391
10. Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment

analysis of social media text.

11. *Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM-14).*
12. GitHub Repository: Shudhanshu771. (2024). *MULTIMODAL-DEPRESSION-DETECTION*. Retrieved from https://github.com/Shudhanshu771/MULTIMODAL-DEPRESSION-DETECTION
13. Gupta, P. and Kulkarni, N., 2013. An introduction of soft computing approach over hard computing. International Journal of Latest Trends in Engineering and Technology (IJLTET), 3(1), pp.254-258
14. Gupta, P., Sharma, V. and Varma, S., 2021. People detection and counting using YOLOv3 and SSD models. Materials Today: Proceedings.
15. Gupta, P., Kulkarni, A. and Sarda, A., 2013. An embedded health care supervisory systems. International Journal of Latest Trends in Engineering and Technology (IJLTET), 3, pp.379-386
16. Gupta, P., Kumar, S., Singh, Y.B., Singh, P., Sharm, S.K. and Rathore, N.K., 2022. The impact of artificial intelligence on renewable energy systems. NeuroQuantology, 20(16), p.5012
17. Gupta, P., Varma, S., Arya, N. and Bhagel, R., 2023. Intelligent Security System Based on the Internet of Things (IoT). In Intelligent Sensor Node-Based Systems (pp. 177-191). Apple Academic Press
18. Gupta, P., Sharma, V. and Varma, S., 2022, September. An Algorithm for Counting People using Dense Nets and Feature Fusion. In 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 1248-1253). IEEE.
19. Rathore, P., Gupta, P., Jain, S. and Shrivastava, Y., 2022. A Study of the Automated Vehicle Number Plate Recognition System. i-manager's Journal on Pattern Recognition, 9(2), p.30.
20. Jain, R., Gupta, P., Bansal, S. and Singh, V., 2022. An AI-Based Online Exam Proctoring Framework. I-Manager's Journal on Computer Science, 10(3)
21. Rajput, A., Gupta, P., Ghodeshwar, P., Varma, S., Sharma, K.K. and Singh, U., 2023, June. Study of Cloud Providers (Azure, Amazon, and Oracle) According To Service Availability and Price. In 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN) (pp. 1177-1188). IEEE.
22. Manurkar, V., Gupta, P., Sharma, G., Singh, U. and Manurkar, N., 2022. A Face Mask Identification System based on the Internet of Things and Machine Learning for Detecting Covid-19. NeuroQuantology, 20(16), p.3930.
23. Gupta, P., Arya, N., Singar, C.P., Chaudhari, A., Singh, U. and Gupta, S., 2025. Safety of Pedestrians in AI-Optimized VANETs for Autonomous Vehicles via Real-Time Vehicle-to-Vehicle Communication. In AI-Driven Transportation Systems: Real-Time Applications and Related Technologies (pp. 169-181). Cham: Springer Nature Switzerland.
24. Gupta, P. and Singh, U., 2025. Evaluation of several yolo architecture versions for person detection and counting. Multimedia Tools and Applications, pp.1-24.
25. Singh, U, Gupta, P., Shukla, M., Sharma, V., Varma, S. and Sharma, S.K., 2023. Acknowledgment of patient in sense behaviors using bidirectional ConvLSTM. Concurrency and Computation: Practice and Experience, 35(28), p.e7819.
26. Gupta, P., Shukla, M., Arya, N. and Singh, U., 2023. 5 Internet Smart and of Intelligent Things in Healthcare Systems. IoT in Healthcare Systems: Applications, Benefits, Challenges, and Case Studies, p.77.
27. Gupta, P. and Singh, U., 2022. Traffic Management for Smart City Using Deep Learning. Autonomous Vehicles Volume 1: Using Machine Intelligence, pp.149-159