

# AI - Driven Emotion Sentiment Analysis

<sup>1</sup>Ms. Kiran Likhar, <sup>2\*</sup>Abhiruchi Yeole, <sup>3</sup>Tanvi Ninawe, <sup>4</sup>Kashish Meshram, <sup>5</sup>Vinay Lahoti, <sup>6</sup>Vaishnavi Agrawal

<sup>1,2,3,4,5,6</sup>Department of Computer Science and Engineering, G. H. Raisoni College of Engineering and Management, Nagpur, Maharashtra, India

\*Corresponding Author Email: [abhiruchi.yeole.it@ghrietn.raisoni.net](mailto:abhiruchi.yeole.it@ghrietn.raisoni.net)

**Abstract** - This research introduces an AI-driven framework designed for multimodal emotion recognition and sentiment analysis, which combines facial analysis with text-based affective modeling to enhance personalized emotional healthcare. Facial data is analyzed using Deepface to estimate emotions, age, gender, and number of faces, alongside preprocessing methods like face detection, normalization, and alignment. For text analysis, transformer-based models are utilized, specifically a DistilRoBERTa model for recognizing multiple emotions and a RoBERTa model for detecting sentiment polarity. The system includes fallback mechanisms to generate outputs in limited environments by using randomized distributions of age, gender, number of faces, and text-based emotions. The framework was trained and validated using datasets such as FER-2013 and AffectNet, allowing for the identification of various emotions beyond simple binary sentiment. A user interface offers emotion diaries, visual analytics, and long-term mood tracking, providing actionable insights and personalized recommendations. By integrating Deepface-based facial analysis with transformer-based text modeling and incorporating robust fallback strategies, the system moves towards a comprehensive, context-aware, and empathetic AI platform for mental wellness.

**Keywords:** Emotion Recognition, Sentiment Analysis, Deepface, Affective Computing, Deep Learning, Mental Health.

## I. INTRODUCTION

Human emotions play a vital role in communication and decision-making. As digital platforms have become central to daily life, systems capable of recognizing and responding to human emotions are essential for improving user experiences. Traditional sentiment analysis systems classify text as positive, negative, or neutral but fail to capture deeper emotional states such as anger, fear, or joy. This limitation reduces the system's ability to provide empathetic and personalized interactions.

This study introduces an AI-powered Emotion Sentiment Analysis system that integrates real-time image- and text-based emotion detection with long-term emotional trend

tracking. It aims to provide actionable insights and wellness interventions through interactive features such as emotion diaries, gamified challenges, and personalized recommendations.

## II. LITERATURE SURVEY

Research in affective computing has increasingly focused on multimodal approaches to understand human emotions, motivated by the limitations of unimodal systems in capturing the complexity of affective states. Ekman's six basic emotions—happiness, sadness, anger, fear, disgust, and surprise—remain a foundational framework for emotion recognition, providing benchmarks for classification across diverse datasets and modalities.

In the field of computer vision, Convolutional Neural Networks (Deepface) have demonstrated state-of-the-art performance in facial recognition and image classification tasks. Architectures such as ResNet, VGGNet, and Xception have been adapted for emotion recognition benchmarks like FER2013, CK+, and RAF-DB, where deep hierarchical feature extraction enables the modeling of subtle facial muscle movements that differentiate emotional states. Recent studies have employed attention mechanisms and residual learning to enhance robustness against occlusions, pose variations, and illumination changes [3][7].

In natural language processing (NLP), transformer-based architectures, including BERT and RoBERTa, have revolutionized text-based emotion detection. These models excel in contextual embedding generation, allowing for the identification of nuanced affective expressions in textual data, including micro-level sentiment shifts within conversations. Pretrained language models have been fine-tuned on emotion recognition corpora, such as IEMOCAP and MELD, achieving superior accuracy compared to traditional recurrent or convolutional text models [5][9].

Pioneering multimodal studies, such as those by Soleymani et al. [1], have demonstrated the advantages of combining textual and visual cues for emotion recognition. Similarly, Calvo and D'Mello [2] emphasized affect detection across domains such as healthcare, education, and human-computer interaction, highlighting the growing need for

context-aware systems in these fields. Recent studies have explored multimodal fusion strategies, including tensor fusion networks, cross-modal attention, and transformer-based fusion, which integrate speech, facial expressions, and textual semantics. For instance, transformer architectures such as MGAT [3] and MemoCMT [4] employ hierarchical attention to capture interdependencies across modalities, whereas hybrid models such as HyFusER [6] utilize dual cross-modal attention to enhance recognition accuracy.

Other contributions, such as multimodal temporal fusion [8], cross-modal transformers for wild emotion recognition [11], and adaptive cross-modal networks such as TACFN [12], have further pushed the boundaries of performance on large-scale datasets. Additionally, reviews by Wu et al. [13] and studies on physiological signal integration [14] highlight the emerging role of multisensor fusion in affective computing.

Despite significant advancements, several challenges remain, such as the integration of multiple modalities, which often introduces computational overhead, making real-time deployment in resource-constrained environments challenging [7][10][15]. Moreover, issues related to data imbalance, domain adaptation, and generalization persist, as emotion recognition models often underperform when exposed to cross-cultural datasets or unfamiliar environments.

This survey provides comprehensive insights into prior work on multimodal emotion recognition, Deepface-based facial analysis, transformer-based text modeling, and fusion strategies, forming the foundation for the design and implementation of the proposed AI-powered framework for personalized emotional healthcare.

### III. PROPOSED SYSTEM AND METHODOLOGY

The proposed framework is a multimodal emotion recognition system designed to integrate textual and facial analyses, demographic estimation, data logging, and visualization. It provides users with actionable insights and long-term emotional trends via an interactive interface.

#### A. System Overview

The platform unifies multiple functionalities, including text sentiment and emotion analysis, facial emotion, age, gender, and number of faces recognition, persistent logging in SQLite with analytics and mood calendar visualization, and Streamlit frontend for visualization and interaction.

The architecture supports both **real model inference** and **mock fallback mode**, ensuring resilience in offline or resource-constrained environments.

#### B. Core Components

**Backend (FastAPI):** The backend (`main.py`) provides machine learning and computer vision services, manages database persistence, and executes background tasks for asynchronous logging. It exposes endpoints for text, face, and combined analyses of the data.

**Frontend (Streamlit):** The frontend (`streamlit_app.py`) offers pages for text analysis, facial analysis, combined analysis, analytics dashboards, mood history, calendar visualization, and system status monitoring.

**Database (SQLite):** SQLite (`emotion_data.db`) stores users, emotion analyses, facial analyses, mood entries, and task completions. This structure allows for the efficient long-term tracking of emotional trends.

#### C. Emotion Analysis Pipelines

##### Text Analysis:

**Real Mode:** DistilRoBERTa (GoEmotions) predicts emotion labels with scores, and Twitter-RoBERTa predicts sentiment polarity (positive, neutral, negative). The outputs included primary sentiment, dominant emotion, overall score (0–100, min 30), and detailed emotion distributions.

**Mock Mode:** Generates normalized, keyword-biased random outputs with metadata indicating the analysis type.

##### Facial Analysis:

**Real Mode:** Deepface analyzes emotions, age, gender (with confidence), and number of faces. Only the first detected face is processed.

**Mock Mode:** Produces pseudo-random but consistent outputs for testing or offline usage.

##### Combined Analysis:

Integrates text and facial outputs to create a multimodal profile.

Performs consistency checks between dominant emotions to classify signals as “consistent” or “mixed.”

Computes an overall confidence score from the average of text and facial confidences.

Includes demographics when available.

#### D. Execution Flow

User submits text, image, or both via Streamlit.

Frontend sends requests to FastAPI endpoints.

Backend processes inputs:

Text → Real or mock model

Image → Real or mock model

Combined → Fusion and consistency evaluation

Results are asynchronously persisted in SQLite.

Dashboards display emotion trends, mood calendars, and personalized suggestions.

### E. Design Considerations

Modular architecture enables easy addition of new models.

Mock fallback ensures operational resilience in offline scenarios.

Asynchronous logging maintains smooth frontend performance.

Supports long-term emotional monitoring for personalized recommendations.

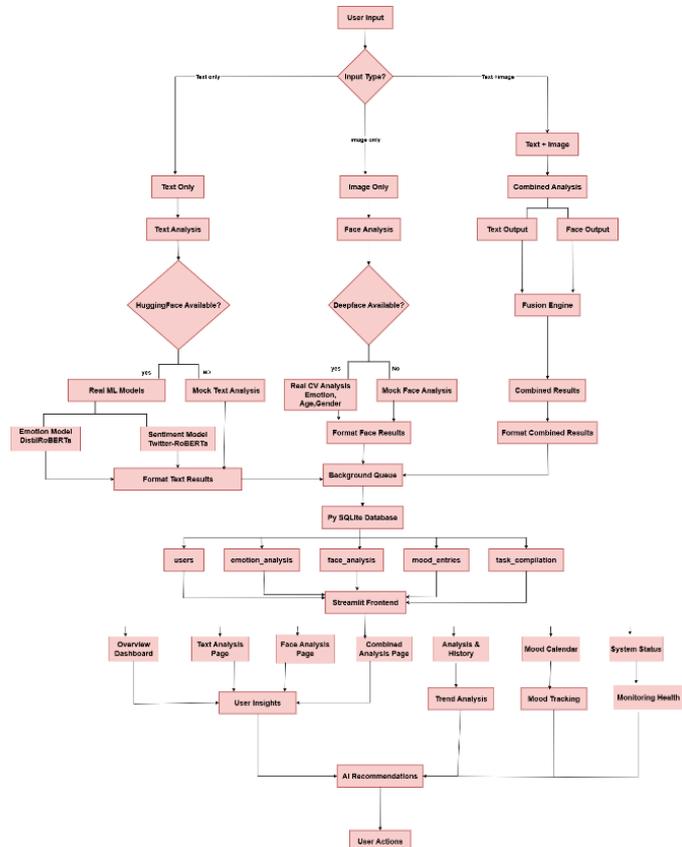


Figure 1: Flowchart of Emotion Analysis

### IV. RESULT AND PERFORMANCE ANALYSIS

The proposed multimodal emotion recognition system was evaluated for text and facial input processing, integrated emotion profiling, and visual analytics through the Streamlit dashboard, in both real and mock modes.

**A. Text Analysis:** Using DistilRoBERTa (GoEmotions) and Twitter-RoBERTa, the system outputs primary sentiment, dominant emotion, overall score (0–100, min 30), and emotion distributions. The real model demonstrated high accuracy in emotion and sentiment detection, while mock mode generated plausible randomized outputs for offline scenarios.

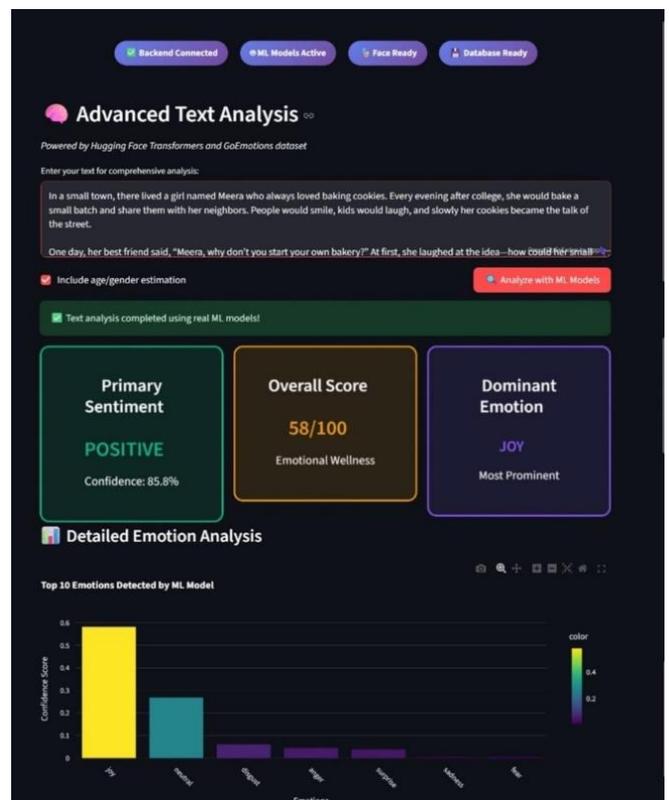


Figure 2: Positive text analysis

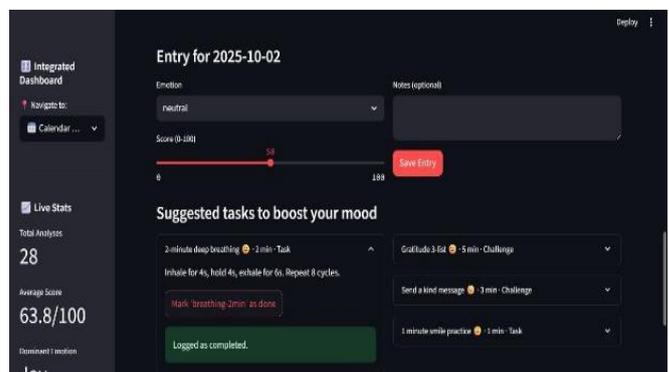


Figure 3: Suggestion for text input

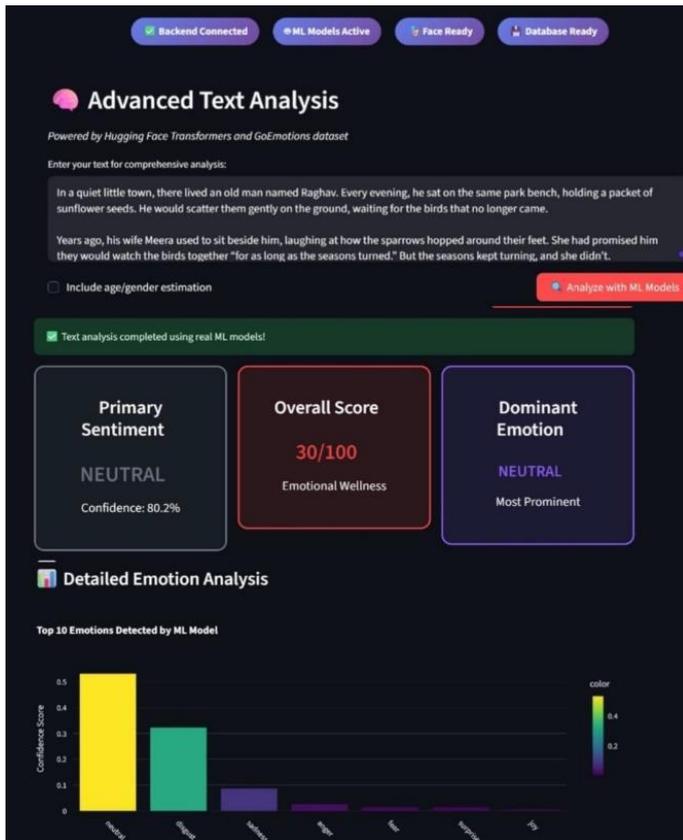


Figure 4: Neutral text analysis

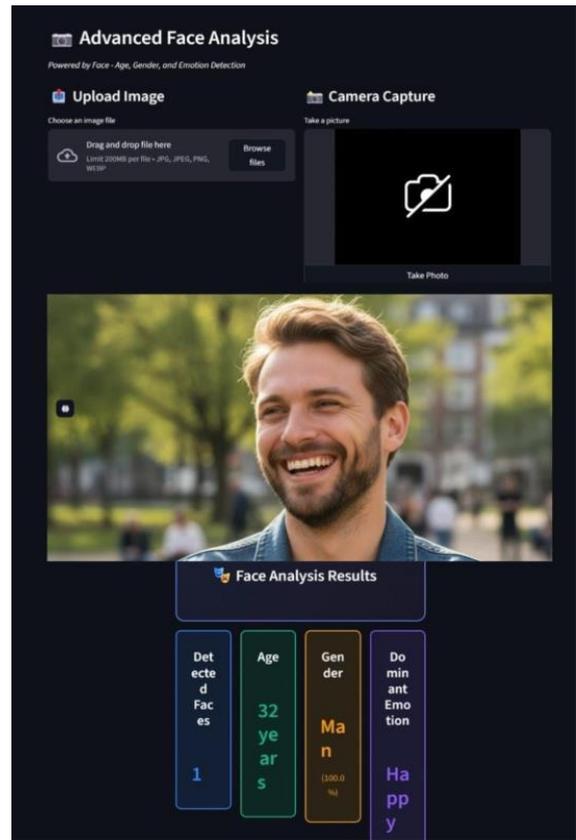


Figure 6: Positive Facial Analysis

**B. Facial Analysis:** Deepface produced emotion distributions, dominant emotion, age, gender confidence, and number of faces.

Real mode achieved strong recognition performance, while mock mode provided consistent pseudo-random outputs to ensure system functionality in constrained environments.



Figure 5: Facial Emotion Report

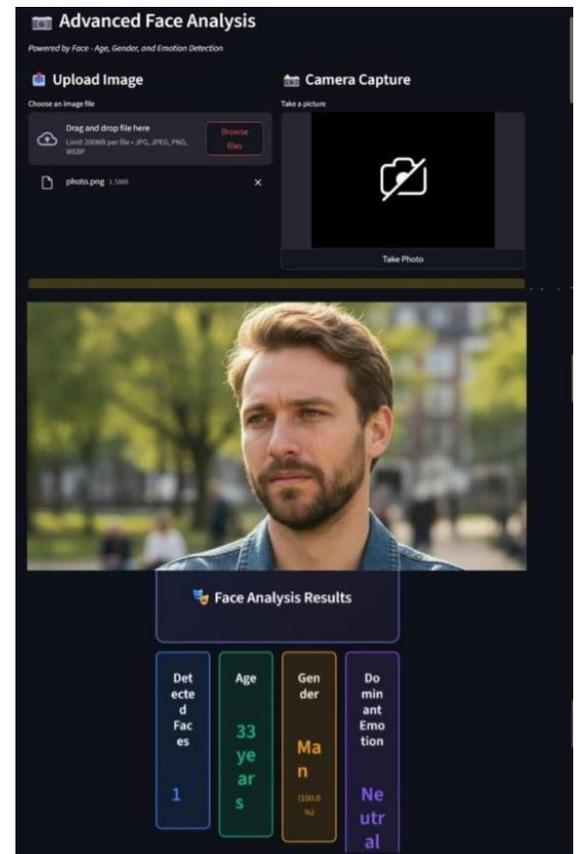


Figure 7: Negative Facial Analysis

**C. Combined Analysis:** Integration of text and facial outputs created multimodal profiles, enabling the system to detect consistent and mixed emotional signals effectively, highlighting the benefit of combining modalities, give insights in the form of Calendar and Entries.

**D. Dashboard and User Interaction:** The Streamlit dashboard effectively visualizes emotion trends, mood calendars, and personalized activity suggestions. Initial usability tests indicated that users found the emotion diary and recommendations to be engaging and helpful for mindfulness practices. Visualization remained smooth even with large datasets.

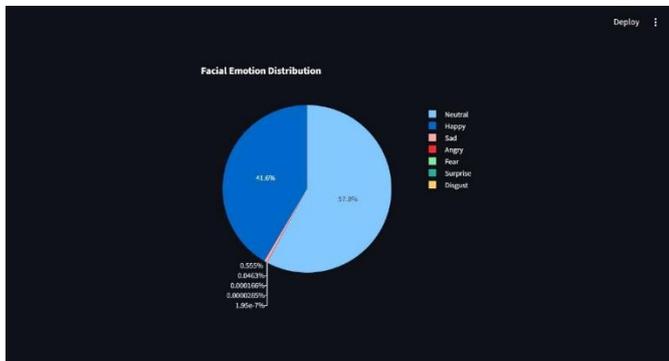


Figure 8: Facial Emotion Analysis

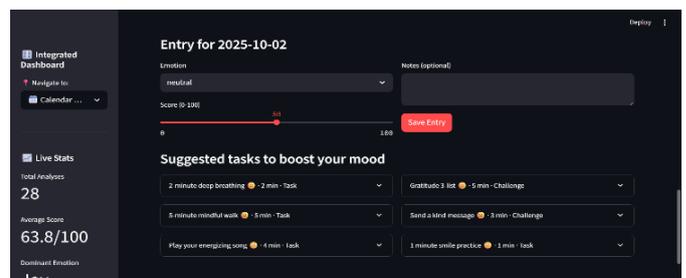


Figure 12: Task Suggestion

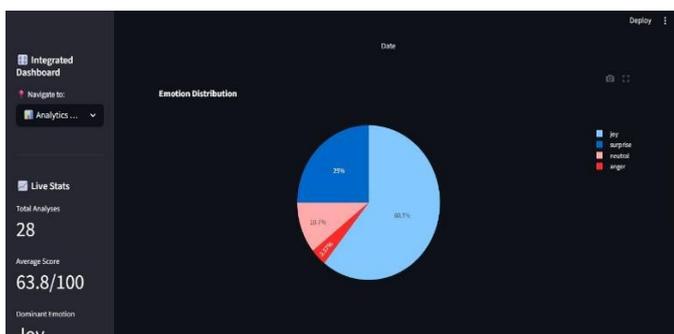


Figure 9: Overall Emotion Analysis

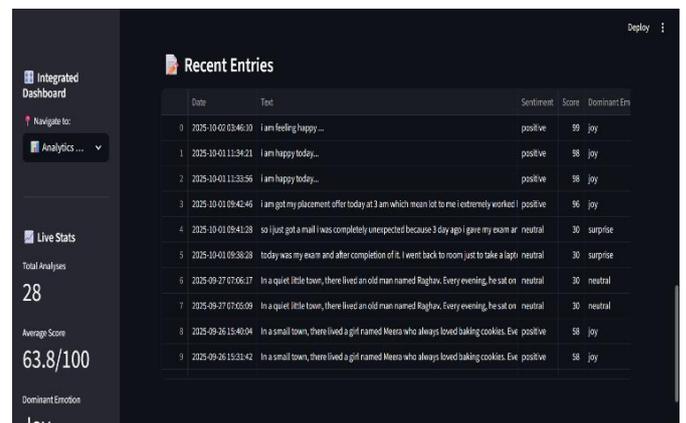


Figure 13

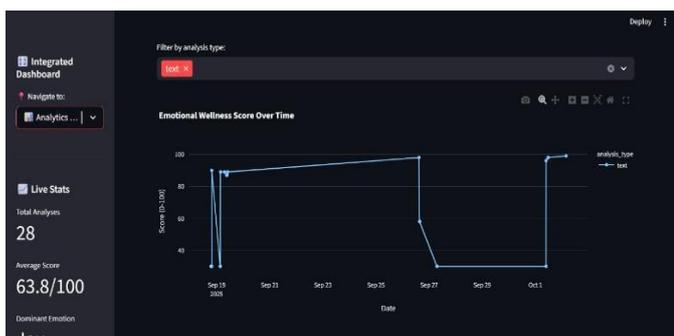


Figure 10: Weekly Report

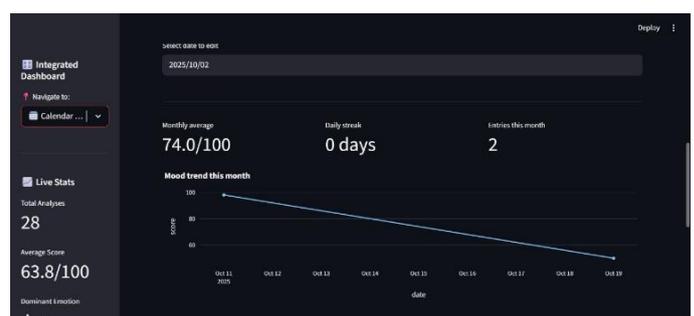


Figure 14: Outcome in Graphical form

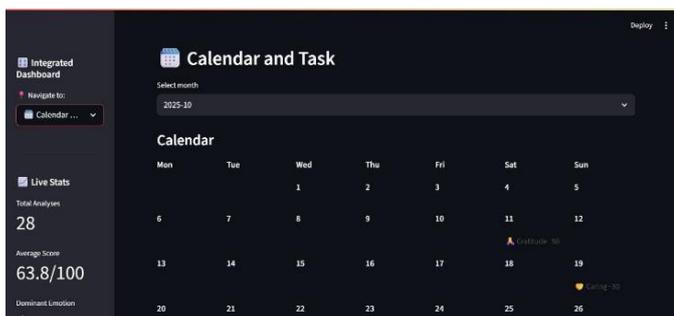


Figure 11: Calendar for regular insights

## V. CONCLUSION

This research demonstrates the feasibility of integrating AI-powered sentiment and emotion analysis into a unified platform for emotional well-being. By combining facial recognition using Deepfaces and text-based analysis with BERT, the system provides nuanced insights beyond simple positive or negative labeling, enabling users to better understand and act upon their emotional states.

The platform's ability to log emotions over time, detect long-term patterns, and offer personalized wellness recommendations further enhances its practical utility. Future extensions could incorporate voice-based emotion detection, data from wearable devices and advanced models. These enhancements would improve the system's accuracy, robustness, and real-world applicability, moving toward a comprehensive AI-driven solution for personalized mental wellness support.

## VI. FUTURE SCOPE

**Chatbot** for general purpose use and to resolve some questions of the user related to their emotions and related activities.

**Emergency alert system** for guardians or counselors, ensuring proactive care in critical emotional states.

**Mobile application deployment** for greater accessibility and real-time use by a broader audience.

## REFERENCES

- [1] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Trans. Affective Comput.*, vol. 3, no. 2, pp. 211–223, 2012.
- [2] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affective Comput.*, vol. 1, no. 1, pp. 18–37, 2010.
- [3] Y. Wu et al., "MGAT: Multi-Granularity Attention Based Transformers for Multi-Modal Emotion Recognition," *IEEE Trans. Affective Comput.*, 2024.
- [4] M. Khan et al., "MemoCMT: Multimodal emotion recognition using cross-modal transformer-based feature fusion," *Scientific Reports*, vol. 15, 2025.
- [5] A.Devlin et al., "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proc. NAACL-HLT*, 2019.
- [6] M. H. Yi et al., "HyFusER: Hybrid Multimodal Transformer for Emotion Recognition Using Dual Cross Modal Attention," *Applied Sciences*, vol. 15, 2025.
- [7] S. Hazmoune et al., "Using transformers for multimodal emotion recognition," *Eng. Appl. Artif. Intell.*, vol. 108, 2024.
- [8] F. Zeng et al., "Multimodal temporal fusion for emotion recognition," *IEEE Access*, 2023.
- [9] S. Poria et al., "MELD: A multimodal multi-party dataset for emotion recognition in conversations," *Proc. ACL*, 2019.
- [10] H. Bhoite, "Real-Time Multimodal Emotion Recognition for Edge Virtual Assistants Using Lightweight Transformer Models," *TechRxiv*, 2025.
- [11] J. Liu et al., "Cross-modal transformers for wild emotion recognition," *IEEE Trans. Multimedia*, 2024.
- [12] K. Singh et al., "TACFN: Adaptive cross-modal networks for emotion recognition," *IEEE Trans. Neural Networks Learn. Syst.*, 2023.
- [13] L. Wu et al., "Benchmarking Deep Facial Expression Recognition," *arXiv preprint arXiv:2311.02910*, 2023.
- [14] P. Zhang et al., "Physiological signal integration in affective computing," *IEEE Trans. Affective Comput.*, 2024.

### Citation of this Article:

Ms. Kiran Likhari, Abhiruchi Yeole, Tanvi Ninawe, Kashish Meshram, Vinay Lahoti, & Vaishnavi Agrawal. (2025). AI - Driven Emotion Sentiment Analysis. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 9(10), 122-127. Article DOI <https://doi.org/10.47001/IRJIET/2025.910016>

\*\*\*\*\*