

Personalized Mental Health Assistant Using Emotion Recognition from Text and Facial Expressions: A Multimodal Framework for Emotion-Aware Wellness Support

Paturi Sai Spandana*(2303031460149@paruluniversity.ac.in), Ponnaganti Dharani Naga

Priya(2303031460252@paruluniversity.ac.in), Konakalla OSNVJBS

Harini(2303031460088@paruluniversity.ac.in), Gowrni Sai Sandhya(2303031460051@paruluniversity.ac.in),

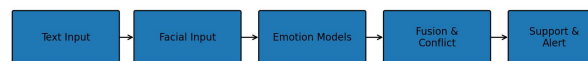
Bandaru Siva Nagendra(2303031460017@paruluniversity.ac.in)

Department of Artificial Intelligence and Machine Learning, Parul University, Vadodara, India

Corresponding author: Paturi Sai Spandana — 2303031460149@paruluniversity.ac.in

Received: _____ Revised: _____ Accepted: _____ Published: _____

Graphical Abstract



Graphical Abstract: Text and facial cues are analyzed and combined to support emotion-aware assistance.

Abstract -

Mental and emotional well-being can be difficult to support consistently because professional care may be limited by cost, availability, geographic distance, or social stigma. This paper presents the design of a Smart Mental Health Assistant that combines text-based and facial-expression-based emotion recognition to estimate a user's emotional state. The framework uses text processing and facial analysis as complementary sources, followed by emotion fusion and downstream modules for mood trend analysis, personalized supportive responses, explainability, privacy-aware processing, and safety-aware escalation. Existing digital wellness systems often rely on a single modality or isolated support mechanisms, which can provide an incomplete representation of the user's emotional state. The proposed framework addresses this limitation by integrating multiple emotional cues within a single pipeline and by explicitly separating emotional-wellness support from clinical diagnosis. The system is intended to provide supportive conversations, recommendations, and appropriate crisis-support resources when risk indicators are detected. At the current development stage, the implementation is being evaluated on a local machine and full empirical testing and deployment are not yet complete; therefore, this manuscript does not claim unverified accuracy or clinical effectiveness.

Keywords -

Multimodal Emotion Recognition; Text Emotion Recognition; Facial Emotion Recognition; Emotion Fusion; Mental Wellness Assistant; Conversational AI; Mood Tracking; Crisis Detection.

1. Introduction

Mental and emotional well-being underpins many dimensions of daily functioning, while chronic stress, unmanaged anxiety, or sustained low mood can affect personal and social life. Despite growing awareness of mental health, meaningful support remains difficult for many people to access consistently because of cost, availability, geographic distance, or social stigma. This gap has motivated research into AI-based systems capable of offering scalable emotional support as a complement, not a replacement, to professional care.

A large share of digital mental-wellness tools rely on self-report mechanisms such as structured questionnaires, mood-rating scales, or free-text journaling prompts. Although useful, self-report approaches require the user to accurately introspect and articulate an emotional state. Text-based emotion recognition can infer affect directly from written language, while facial-expression analysis can provide complementary visual cues. Combining these modalities can reduce dependence on a single source when one signal is ambiguous or incomplete.

Any system that processes emotional and behavioural information also carries obligations around privacy, transparency, and safety. The proposed assistant is therefore

positioned as an emotional-wellness support tool rather than a diagnostic instrument for depression, anxiety, or other psychiatric conditions.

The main objectives are to (i) recognize emotions from text and facial expressions, (ii) combine the two outputs to obtain a more robust emotional-state estimate, (iii) identify inconsistencies between modalities, (iv) track recent emotional trends, (v) provide personalized supportive suggestions and conversations, and (vi) identify high-risk situations and encourage appropriate human or professional support. The remainder of this paper reviews related work, presents the proposed methodology, summarizes the reviewed literature, discusses the current evaluation status and limitations, and concludes with future scope.

2. Related Work

2.1 Text-Based Emotion Recognition

Text-based emotion recognition infers a user's affective state directly from written language such as chat messages, journal entries, or free-text responses. Transformer-based language models have substantially improved contextual language understanding. Devlin et al. introduced BERT, demonstrating the value of bidirectional contextual representations for language understanding [1]. For the proposed assistant, text is treated as one of the primary emotional signals and can be processed using conventional machine-learning methods or contextual encoders depending on the implementation stage.

2.2 Facial Emotion Recognition

Facial expressions can reveal transient emotional states that a person may not articulate in words. Convolutional neural networks (CNNs) have become a common architecture for facial-expression recognition because they can learn spatial patterns from face images. Hong et al. investigated CNN-based facial classification and demonstrated that even comparatively simple deep-learning architectures can extract useful emotion-related visual patterns [2]. In the proposed system, facial analysis provides a complementary signal to text-based emotion recognition.

2.3 Multimodal Emotion Fusion

A system that relies on only one emotional signal can be affected when that signal is ambiguous, sparse, or noisy. Multimodal machine-learning research therefore studies ways of combining complementary sources. Baltrušaitis, Ahuja, and Morency provide a foundational taxonomy of multimodal machine learning and fusion strategies [3]. The proposed assistant applies this general principle at the decision level by combining text and facial emotion outputs rather than treating either modality as universally authoritative.

2.4 Conversational AI for Mental Health Support

Conversational agents are increasingly used to provide scalable mental-health support outside formal clinical

settings. Gaffney, Mansell, and Tai reviewed conversational agents for mental-health problems and highlighted the diversity of intervention and evaluation approaches [4]. A later systematic review and meta-analysis also examined AI-based conversational agents for mental health and well-being [5]. These findings motivate the use of supportive conversation in the proposed system while maintaining clear boundaries around clinical diagnosis.

2.5 Longitudinal Mood Tracking

A single emotional-state estimate captured at one moment cannot fully represent a user's mood trajectory over days or weeks. Digital-phenotyping research has explored longitudinal behavioural and emotional indicators for understanding changes over time. Aledavood et al. investigated multimodal digital phenotyping for mood-related assessment [6]. The proposed assistant similarly maintains recent emotion estimates so that sustained changes can be considered rather than relying only on a single interaction.

2.6 Personalization and Adaptive Intervention

The same estimated emotional state may call for different supportive actions depending on user preferences and previous responses. Reinforcement-learning-based personalization has been explored for adaptive digital interventions, including the StayWell and DIAMANTE studies [7], [8]. The proposed framework uses user feedback and recent emotional trends as inputs for selecting supportive actions; however, advanced reinforcement-learning implementation remains a future extension rather than a completed component.

2.7 Explainable AI in Mental Health

Emotional and mental-health inferences can influence how users are subsequently engaged or supported, making transparency important. Research on explainable AI has examined interpretable approaches for clinical risk prediction and observable-cue-grounded explanations [9], [10]. LIME and SHAP provide widely used model-agnostic and attribution-based approaches [11], [12]. Explainability is therefore considered a design objective of the framework, while a complete explainability implementation remains subject to future validation.

2.8 Privacy-Preserving AI and Data Security

Emotional and behavioural data are sensitive, so systems handling such information should minimize unnecessary collection and protect user control. Federated-learning research has investigated model training without centralizing raw user data, including work such as FedMentalCare and broader ethics reviews [13], [14]. These approaches inform the privacy-aware design direction of the proposed framework; a full federated deployment is not claimed as completed in the current local implementation.

2.9 Safety, Crisis Detection, and Risk-Aware Design

A wellness assistant may encounter signs of significant distress and should respond conservatively. Ophir et al. demonstrated machine-learning-based detection of suicide risk from textual social-media posts [15]. The proposed assistant includes a safety-aware concept that identifies high-risk signals and encourages appropriate human or professional support rather than presenting a diagnosis or treatment recommendation.

2.10 Research Gap

Existing work has advanced individual components such as text emotion recognition, facial emotion recognition, multimodal fusion, conversational support, longitudinal mood analysis, explainability, privacy, and safety. However, these capabilities are often studied separately. The proposed Smart Mental Health Assistant focuses on integrating text and facial emotion recognition with fusion, emotional conflict detection, trend analysis, supportive recommendations, adaptive conversation, and safety-aware escalation within a single workflow.

3. Methodology

The proposed Smart Mental Health Assistant is organized around interconnected modules for text emotion recognition, facial emotion recognition, decision-level fusion, emotional conflict detection, emotional stability analysis, support recommendation, adaptive conversation, and safety-aware response.

3.1 Text Emotion Recognition

The text module accepts a user's written input and converts it into a numerical representation using TF-IDF features. A Support Vector Machine (SVM) classifier is used to predict the emotion category from the resulting feature vector. The output is an emotion label and associated model confidence where available.

3.2 Facial Emotion Recognition

The facial module uses OpenCV for face detection and a CNN-based classifier for facial emotion recognition. Video frames are processed to identify the face region, after which the cropped facial representation is passed to the trained CNN to obtain an emotion prediction.

3.3 Decision-Level Emotion Fusion

The system combines the text and facial predictions at the decision level. When both modalities are available, their predicted emotion labels and confidence information are compared to derive a combined emotional-state estimate. If only one modality is available, the system can continue using the available signal rather than requiring both inputs.

3.4 Emotional Conflict Detection

Text and facial predictions may disagree. The conflict-detection module explicitly identifies such inconsistencies

instead of silently treating one modality as correct. A conflict flag can be used to lower confidence in the final interpretation and to guide the assistant toward a neutral or supportive response.

3.5 Emotional Stability and Trend Analysis

The assistant can maintain recent emotion estimates and examine their trend over multiple interactions. The objective is to distinguish an isolated emotional change from a repeated pattern. A stability score can summarize recent consistency or fluctuation, while sustained negative trends can be passed to the recommendation and safety modules.

3.6 Smart Recommendation and Adaptive Support

Based on the estimated emotional state, the assistant can provide supportive actions such as breathing exercises, music or relaxation suggestions, guided reflection, or a supportive conversational prompt. User feedback and recent emotional context can be considered when selecting subsequent support. The system is designed for wellness support and does not provide clinical diagnosis.

3.7 Safety-Aware Alert Mechanism

When the system identifies high-risk or strongly concerning signals, it can surface appropriate crisis-support resources and encourage the user to seek human or professional help. The alert mechanism is intended as a safety feature rather than a diagnostic classifier.

3.8 Current Implementation Status

The current system runs on a local machine. Full deployment, systematic testing, and complete empirical benchmarking have not yet been completed. Accordingly, this manuscript does not report fabricated accuracy, precision, recall, F1-score, latency, or clinical-effectiveness values.

Algorithm 1 — Text and Facial Emotion Fusion

Input: Text T, facial frame V

Step 1: Preprocess T and extract TF-IDF features.

Step 2: Predict text emotion using the SVM classifier.

Step 3: Detect the face in V using OpenCV.

Step 4: Classify the facial region using the CNN emotion model.

Step 5: Compare the two modality predictions and confidence values.

Step 6: Produce the combined emotional-state estimate and conflict flag.

Output: Fused emotion estimate, confidence, and conflict status.

Table 1. Comparative Summary of Reviewed Literature

Technique Used	Application Area	
BERT (Transformer)	Text Emotion Recognition	Deep context understanding
CNN Facial Classifier	Facial Emotion Recognition	Useful visual cues
Multimodal ML Survey	Emotion Recognition	Foundational research
Conversational Agent Review	Mental Health Chatbots	Mixed-methods
Meta-Analysis of AI CAs	Mental Health & Well-Being	Evidence synthesis
MoMo-Mood Digital Phenotyping	Longitudinal Mood Tracking	Multimodal markers
RL-Personalized Texting	Adaptive Intervention	Personalized support

4	RL-Personalized Messaging	Digital Health Personalization
023	Interpretability Survey	Explainable AI
w, 2024	XAI Review	Explainable AI in Mental Health
	FedMentalCare	Privacy-Preserving AI
	FL Ethics Review	Privacy-Preserving AI
	ANN Suicide-Risk Detection	Safety / Crisis Detection
	LIME	Explainable AI
	SHAP	Explainable AI

5. Results and Discussion

5.1 Evaluation Status

The current implementation is operational on a local machine, but systematic testing has not yet been completed. Therefore, this section intentionally does not report numerical performance claims. Before submission as an experimental research article, the system should be evaluated using a held-out dataset and documented metrics such as accuracy, precision, recall, F1-score, confusion matrices, and, where relevant, inference time.

5.2 Expected Behaviour Based on the Design

The text and facial modules are intended to provide complementary emotional cues. Decision-level fusion should be particularly useful when one modality is ambiguous, while conflict detection should make disagreements explicit. Trend analysis can provide context across interactions, and the recommendation module can map detected emotional states to supportive actions. These are design expectations and should not be presented as measured findings until experimental validation is completed.

5.3 Limitations

The current work is limited by the incomplete testing and deployment stage. Model performance depends on the quality and representativeness of the emotion datasets, facial conditions such as lighting and pose, and the reliability of text classification. Emotional inference is inherently uncertain, and the system should not be treated as a substitute for professional mental-health assessment.

6. Conclusion and Future Scope

This paper presents a Smart Mental Health Assistant framework that combines emotion recognition from text and facial expressions with decision-level fusion, emotional conflict detection, trend analysis, supportive recommendations, adaptive conversation, and safety-aware escalation. The framework is designed as an emotional-wellness support tool rather than a clinical diagnostic system. The current implementation runs locally and remains under development; therefore, the paper deliberately avoids unsupported performance or deployment claims.

Future work will focus on completing systematic testing, benchmarking the text and facial models, validating the fusion strategy, improving recommendation quality,

completing deployment, evaluating usability, and strengthening privacy and safety mechanisms. Experimental results should be added after these evaluations are completed.

7. Author Declarations

Data Availability

The current implementation and evaluation materials are maintained locally. Public release of project-specific data and code is not yet available.

Study Limitations

Systematic testing and deployment are not yet complete, so no unverified quantitative performance claims are reported.

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding Source

No external funding was reported for this work.

Authors' Contributions

All authors contributed to the conception, literature review, development, documentation, discussion, and revision of the project and manuscript.

Acknowledgements

The authors acknowledge the guidance of Ms. Ritu Agrawal, Professor.

8. References

- [1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.
- [2] A. J. Hong, D. DiStefano, and S. Dua, "Can CNNs Accurately Classify Human Emotions? A Deep-Learning Facial Expression Recognition Study," arXiv preprint arXiv:2310.09473, 2023.
- [3] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 2, pp. 423–443, 2019.
- [4] H. Gaffney, W. Mansell, and S. Tai, "Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review," JMIR Mental Health, vol. 6, no. 10, e14166, 2019.
- [5] "Systematic Review and Meta-Analysis of AI-Based Conversational Agents for Promoting Mental Health and Well-Being," PMC10730549, 2023.
- [6] T. Aledavood et al., "Multimodal Digital Phenotyping Study in Patients With Major Depressive Episodes and Healthy Controls (Mobile Monitoring of Mood): Observational Study," 2025.
- [7] "Personalizing a Mental Health Texting Intervention Using Reinforcement Learning" (StayWell), PMC12673119, 2025.
- [8] "Effectiveness of a Digital Health Intervention Leveraging Reinforcement Learning: Results From the Diabetes and Mental Health Adaptive Notification Tracking and Evaluation (DIAMANTE) Randomized Clinical Trial," PMC11496924, 2024.
- [9] "Explainable AI for Clinical Risk Prediction: A Survey of Concepts, Methods, and Modalities," arXiv preprint arXiv:2308.08407, 2023.

- [10] “Multimodal Observable Cues in Mood, Anxiety, and Borderline Personality Disorders: A Review of Reviews to Inform Explainable AI in Mental Health,” PMC12723009, 2024.
- [11] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why Should I Trust You?: Explaining the Predictions of Any Classifier,” in Proc. ACM SIGKDD, 2016.
- [12] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” in Proc. NeurIPS, 2017.
- [13] N. Sarwar, “FedMentalCare: Towards Privacy-Preserving Fine-Tuned LLMs to Analyze Mental Health Status Using Federated Learning Framework,” arXiv preprint arXiv:2503.05786, 2025.
- [14] “Federated Learning in Healthcare Ethics: A Systematic Review of Privacy-Preserving and Equitable Medical AI,” PMC12896918, 2025.
- [15] Y. Ophir, R. Tikochinski, C. S. C. Asterhan, I. Sisso, and R. Reichart, “Deep Neural Networks Detect Suicide Risk from Textual Facebook Posts,” Scientific Reports, vol. 10, p. 16685, 2020.

9. Author Profiles

Author profile photographs were not available in the source manuscript and therefore have not been fabricated. Add the authors' profile photographs and short biographies here before final submission if required by the journal template.