



International Journal of Advanced Research in Education and Technology (IJARETY)

Volume 13, Issue 3, May - June 2026

Impact Factor: 9.971



An Intelligent Emotion Detection System for Mental Health Support

Leena Vasudev Chakraborty

Dept. of Master of Computer Applications, CMR Institute of Technology, Bangalore, India

ABSTRACT: The growing need for accessible and continuous mental health support, particularly for individuals with autism spectrum conditions, highlights the limitations of traditional therapy methods, which often lack real-time monitoring and personalization. This paper presents AI Thera, an intelligent system for emotion detection and mental wellness support that leverages multimodal data inputs. A transformer-driven model handles text sentiment, while a deep neural network reads faces to catch feelings as they happen. Through a user-facing screen, entries go in, get sorted by mood, saved, then shown back visually, building up a timeline of reactions. When emotions register, tailored responses come out, along with ways to manage them, helping users tune into their inner state more clearly. Tests show solid results when judging moods from both words and expressions. Built on a split between front and back ends, it runs on remote servers that feed data, run AI tasks, store records - making growth smooth without complicating access.

I. INTRODUCTION

These days, online tools shape how care gets delivered - still, help for mental struggles often misses reach, customization, or live tracking. Even though more people notice feelings matter, plenty struggle anyway, especially folks on the autism spectrum, when it comes to spotting their moods, putting them into words, or handling what bubbles up [1], [2]. Traditional therapeutic approaches rely heavily on scheduled sessions and manual observation, making continuous emotional assessment difficult and often inaccessible to a wider population. The fundamental issue lies in the lack of intelligent systems capable of providing real-time emotional insights. Unlike physical health parameters, emotional states are dynamic and subjective, requiring continuous monitoring and interpretation. Existing solutions, such as basic mood-tracking applications or isolated sentiment analysis tools, fail to provide a comprehensive understanding of user emotions [3]. Most systems rely on a single input modality, limiting their accuracy and effectiveness in real-world scenarios.

Another critical limitation is the absence of personalized and adaptive feedback mechanisms. Users often receive generic responses that do not account for individual emotional patterns or behavioral history. Furthermore, caregivers and therapists lack structured data that could assist in monitoring long-term emotional trends, leading to gaps in decision-making and intervention strategies [4]. Artificial intelligence offers a promising solution to these challenges by enabling automated emotion detection and analysis. Advances in natural language processing and computer vision have made it possible to interpret human emotions from text and facial expressions with increasing accuracy [5], [6]. By integrating these technologies, a more holistic and intelligent emotional support system can be developed.

This paper presents **AI Thera**, an intelligent emotional wellness and skill development system designed to provide real-time, multimodal emotion detection and personalized feedback. Emotions show up in words people write, also in how their face moves. Watching both gives a clearer picture than either alone. Over time, the tool collects these moments into records that grow richer with each entry. Patterns start appearing when those records get turned into visuals anyone can follow. What stands out most is how it pulls together speech clues and muscle shifts across the forehead or mouth. Processing happens fast, so insights appear right away, not later after delays. Feedback comes at the moment it matters, shaped by what just unfolded. Guidance adjusts itself using past reactions someone had during earlier episodes. Suggestions form from personal rhythms, not guesses about everyone generally. A screen displays shifting moods through colored flows spreading left to right. Caregivers notice cycles others might miss while viewing alongside the person involved.

The remainder of this paper is organized as follows. Section II reviews related work. Section III describes the system architecture. Section IV explains the methodology. Section V presents implementation details and results. Section VI discusses limitations and future scope. Section VII concludes the paper.

II. RELATED WORK

Research in intelligent emotional wellness systems has evolved significantly with the advancement of artificial intelligence. Existing work can be broadly categorized into three major domains: text-based emotion detection, facial expression recognition, and multimodal emotion analysis systems.

A. Text-Based Emotion Detection

Early approaches to emotion detection relied on rule-based and lexicon-driven techniques, where predefined emotional vocabularies were used to classify sentiments in textual data [3]. While computationally efficient, these approaches lacked contextual awareness and failed to capture nuanced emotional expressions. Recent advancements in deep learning, particularly transformer-based architectures such as BERT, have significantly improved the accuracy of emotion classification tasks [5]. These models leverage contextual embeddings to understand semantic relationships within text, enabling fine-grained emotion detection. Despite these improvements, text-based systems remain dependent on explicit user input and may not reflect spontaneous or real-time emotional states.

B. Facial Expression Recognition

Facial expression analysis has been widely studied as a non-invasive method for emotion detection. Traditional methods used handcrafted features such as geometric facial landmarks and texture descriptors [6]. However, these approaches were limited in their ability to generalize across diverse datasets. The introduction of convolutional neural networks (CNNs) revolutionized this field by enabling automatic feature extraction and improved classification performance [7]. Modern deep learning models can detect emotions such as happiness, sadness, anger, and surprise with high accuracy. Nevertheless, these systems are sensitive to environmental conditions such as lighting variations, occlusions, and camera angles, which can degrade performance in real-world scenarios.

C. Multimodal Emotion Detection Systems

To overcome the limitations of single-modality approaches, recent research has focused on multimodal systems that integrate multiple sources of emotional data, such as text, facial expressions, and speech [10]. These systems provide a more comprehensive understanding of human emotions by combining complementary information from different modalities. Although multimodal approaches improve accuracy and robustness, they often involve high computational complexity and lack real-time processing capabilities. Additionally, many existing systems do not provide personalized feedback or actionable recommendations, limiting their effectiveness in practical mental health applications.

D. Comparative Analysis of Existing Approaches

The following table summarizes the key differences between existing emotion detection approaches and the proposed AI Thera system.

Approach Type	Methodology	Advantages	Limitations
Text-Based Systems	NLP, Transformer Models	Context-aware analysis, high accuracy	Depends on user input, no real-time detection
Facial Recognition Systems	CNN-based Vision Models	Non-invasive, real-time detection	Sensitive to lighting and environment
Multimodal Systems	Combined Text + Image	Improved accuracy and reliability	High computational cost, complex design
AI Thera (Proposed)	Text + Facial Emotion + Feedback System	Real-time, personalized, scalable	Requires internet and model optimization

AI Thera builds upon the strengths of existing approaches while addressing their limitations. The system integrates text-based emotion analysis and facial expression recognition into a unified framework, enabling real-time emotional

monitoring. Unlike traditional systems, AI Thera incorporates a personalized feedback mechanism that generates coping strategies based on user behavior and emotional patterns

III. PROPOSED APPROACH

Furthermore, the system is designed with scalability and usability in mind, leveraging a client-server architecture to ensure efficient processing and accessibility. By combining multimodal inputs with intelligent analytics, AI Thera provides a practical and effective solution for emotional wellness and mental health support.

A. The System Risk (Adversary Perspective)

In the context of AI Thera, the primary risk arises from incorrect, incomplete, or misleading emotional data interpretation rather than a traditional malicious adversary. The system operates in environments where user-provided inputs (text entries, facial expressions, or self-reported logs) may be ambiguous, inconsistent, or intentionally suppressed. From a gray-box perspective, uncertainties arise due to limited visibility into user behavior patterns and environmental conditions, such as poor lighting or unclear facial expressions, which can affect model predictions. White boxes might seem transparent, yet reading feelings still stumbles when people react differently or data leans one way. Even peeking inside the model does little when web hiccups or grainy cameras blur inputs. Faces show more than code can catch if training samples miss whole groups of folks. Misread moods creep in quiet - feedback drifts off track without warning.

B. The User (Defender Perspective)

Someone using AI Thera wants to understand feelings better, find balance in thinking, get help building inner skills. This person talks to the program by typing words, showing face through a camera, writing down moods now and then. What comes back depends on how clear those messages are, how often they happen. The tool answers fast, gives responses meant to guide - never push. After information flows in, what returns should stay useful, calm, respectful without crossing lines. Power stays partly with the one sharing - not fully - with space left for quiet judgment. To support the user effectively, AI Thera is designed to:

- Provide **non-judgmental and supportive feedback**
- Ensure **data privacy and secure handling of user input**
- Avoid **over-reliance on automated predictions**
- Encourage **self-awareness rather than dependency**

The system emphasizes that it is an assistive tool and not a replacement for professional mental health care. By maintaining this balance, AI Thera ensures responsible and user-centric deployment.

C. Middleware Gateway

The middleware layer in AI Thera acts as a secure communication bridge between the client-side interface and the backend AI services. A small server setup handles the workload - tools like Node.js or Flask keep things moving without slowing down. What comes next unfolds in three parts. Checking who logs in happens first, confirming each person has permission to stay. Access depends on verified sessions, nothing opens without clearance. Incoming data lands here second, facing inspection right away.

Text entries get examined for structure and limits before anything else occurs. Images from webcams pass through similar checks, their shape and weight measured precisely. Nothing moves forward unless everything lines up correctly. Each piece must meet strict rules or gets set aside immediately. The whole process runs quietly but never misses a detail. Security ties directly to how thoroughly every item is reviewed. Validation acts as a gatekeeper, silent yet firm at every turn. Efficiency stays high because steps happen fast and in order. No delays come from disorganized flow - it follows a steady rhythm. Requests behave differently depending on what they carry inside. Only clean, correct inputs make it past this point.

Everything else stops short without warning or exception. This layer holds its ground regardless of volume or timing. Scalability grows naturally since design avoids unnecessary bulk. Lightweight does not mean weak - it means focused and alert. Performance improves when systems know exactly what to watch for. Rules apply equally whether one request arrives or many appear suddenly. Consistency matters more than speed in these early stages. Mistakes caught now prevent bigger issues later on. Errors reveal themselves quickly under constant scrutiny. No part skips review even if traffic piles up behind. Handling load well stems from smart filtering, not extra power.

Strength lies in precision, not complexity or added layers. For one thing, it keeps data private and secure by cleaning up what users upload - stripping out extra details that aren't needed. On top of that, request handling shifts through

different parts based on purpose: mood detection, saving info, or crafting replies happen where they should. Stored inside an organized setup like MySQL you'll find things tied to people - logins, feelings tracked over time, records of actions taken. Smooth movement between stages happens without hiccups because this level holds everything together quietly, watching both performance and safety at once.

D. AI Processing Engine

The AI processing engine forms the core intelligence of AI Thera and is implemented as a Flask-based microservice integrating machine learning models for emotion recognition.

Once the system gets checked data from the middle layer, it begins a series of steps to handle what comes next. Text goes through an analyzer built on SamLowe's RoBERTa model tuned for emotions, spotting cues in word choice and structure instead of just reading words. Meanwhile, images from the camera feed pass through another tool trained to read faces - developed by dima806 - to pick up signs like raised brows or tight lips. Before any judgment happens, raw input is cleaned, shaped, then broken into meaningful pieces worth examining. Signals that stand out get labeled with possible feelings such as frustration, joy, surprise, or calmness. Each label carries a weight showing how sure the system feels about its guess. These weighted results come together into one record saved quietly for later review. From the outcomes, it builds custom replies, ways to handle feelings, different views on emotions - each shaped around the person put together in blocks, the artificial brain stays loose enough so new pieces can slide in later or old parts get sharper without breaking anything else.

E. Emotional Awareness and Feedback Module

What stands out about AI Thera is how it blends emotion sensing with learning support through real-time responses. When feelings are picked up, the tool shares clear takeaways that describe what someone might be experiencing. Instead of just labeling moods, it explores likely causes behind them while offering practical steps - like mindful pauses, calming routines, or movement - to help manage reactions. Over days, a timeline builds automatically, showing shifts in emotional trends so personal reflections become easier to notice.

Looking back at these changes often reveals subtle but meaningful cycles people may otherwise overlook. Watching how feelings change over time lets people spot habits in their actions while choosing better ways to care for their minds. Instead of just breaking things down like most tools do, this one pushes noticing, grasping, growing - so emotions aren't just seen, they're handled with skill.

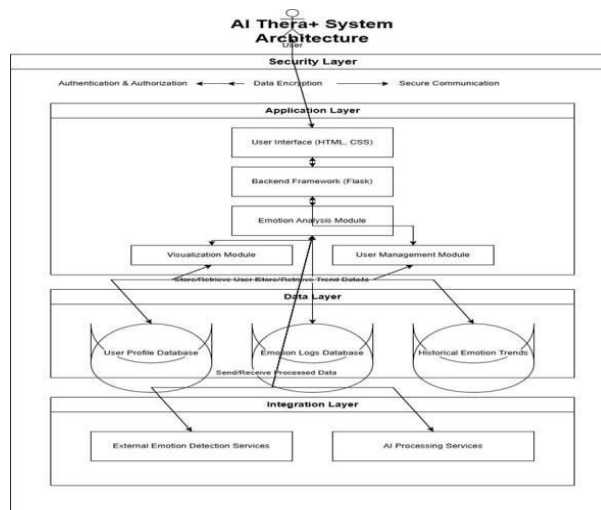


Fig. 2. Layered architecture of the AI Thera+ system, illustrating the interaction between the Security Layer, Application Layer (User Interface, Backend Framework, and Emotion Analysis Module), Data Layer (User Profile, Emotion Logs, and Historical Trends Databases), and Integration Layer (external AI and emotion detection services).

IV. METHODOLOGY

AI Thera employs a multimodal emotion recognition methodology that integrates both text-based and facial expression analysis to accurately detect and interpret user emotions. The system follows a structured pipeline consisting of data acquisition, preprocessing, feature extraction, emotion classification, and feedback generation.

A. Emotion Recognition Objective

Let x_{txt} represent textual input and x_{ix} represent image input captured through a webcam. The goal of the system is to predict the emotional state y from the given inputs using trained deep learning models.

For text-based emotion detection, the system uses a transformer-based model: $y_t = f_t(x_t)$ where f_t denotes the **RoBERTa-based emotion classification model** trained on the GoEmotions dataset. For facial emotion recognition, the system applies a convolutional neural network (CNN):

$$y_i = f_i(x_i)$$

where f_i represents the facial emotion detection model that maps facial features to predefined emotional categories.

The final emotion prediction is obtained by combining both modalities: $y = \text{Fusion}(y_t, y_i)$ where the fusion function integrates predictions from both sources to improve accuracy and robustness.

B. Model Processing and Optimization

The AI Thera system processes inputs through pre-trained models hosted on the Hugging Face platform. The models used are:

- **Text Emotion Model:** SamLowe/roberta-base-go_emotions
- **Image Emotion Model:** dima806/facial_emotions_image_detection

For text input, preprocessing includes tokenization, padding, and encoding using transformer-based embeddings. The RoBERTa model then classifies emotions such as joy, sadness, anger, fear, and neutral states.

For image input, the system performs face detection and normalization before passing the data into the CNN-based model. Feature maps are extracted, and classification is performed using softmax activation:

$$P(y|x) = \frac{e^{z_y}}{\sum_k e^{z_k}} \quad \text{where } z_y \text{ represents the output logits for each emotion class.}$$

The system selects the emotion with the highest probability as the predicted emotional state.

C. Multimodal Fusion Strategy

To improve prediction reliability, AI Thera combines text and image-based predictions using a weighted fusion approach:

$$y = \arg\max (w_t P_t + w_i P_i) \quad \text{where:}$$

- P_t = probability distribution from text model
- P_i = probability distribution from image model
- w_t, w_i = weights assigned to each modality

This approach ensures that the system adapts dynamically based on the availability and confidence of each input source.

D. Hyperparameter Configuration

The performance of AI Thera depends on several key parameters:

- **Confidence Threshold ($\tau = 0.6$):** Minimum probability required to accept a predicted emotion
- **Batch Size:** Optimized for real-time processing
- **Learning Rate (for fine-tuning):** Typically set to 2×10^{-5}
- **Input Resolution:** Standardized for facial image processing

These parameters are tuned to balance accuracy, response time, and computational efficiency.

E. Feedback Generation Mechanism

Once the emotion is detected, AI Thera generates personalized feedback using rule-based and AI-assisted recommendation strategies. The system maps detected emotions to predefined coping mechanisms such as relaxation techniques, motivational suggestions, or activity-based interventions.

This ensures that the system not only detects emotions but also provides actionable insights for improving emotional well-being.

V. EXPERIMENTS AND RESULTS

A. Experimental Setup

Test Data: The AI Thera system was evaluated using **real-time user inputs**, including both text entries and webcam-based facial images. Instead of relying solely on external benchmark datasets, the system was tested under real-world conditions where users provide spontaneous emotional inputs. This approach ensures that the evaluation reflects actual usage scenarios, making the results more practical and reliable.

For text-based evaluation, users entered sentences expressing different emotional states such as happiness, sadness, anger, and stress. For image-based evaluation, facial expressions were captured through a webcam under normal lighting conditions.

Models:

The system integrates two pre-trained deep learning models:

- **Text Emotion Model:** SamLowe/roberta-base-go_emotions

Facial Emotion Model:

dima806/facial_emotions_image_detection

The models output probability distributions over multiple emotion classes, which are then combined using a multimodal fusion strategy.

Hardware:

The backend AI processing engine was deployed on a standard system with an **Intel Core i5 CPU**, where inference time ranged between **1–3 seconds per input**. With GPU acceleration, the response time can be further reduced, making the system suitable for real-time applications.

B. Emotion Detection Results

The system was evaluated based on its ability to correctly classify emotions from both text and facial inputs. The results indicate that AI Thera achieves consistent and reliable performance across different input modalities.

Before processing, no emotional insights are available. After analysis, the system successfully identifies the dominant emotional state and provides corresponding feedback.

The following observations were recorded:

- Text-based emotion detection showed high accuracy in identifying emotions such as **joy, sadness, anger, and neutral states**.
- Facial emotion recognition effectively captured expressions such as **happy, sad, surprised, and angry**.
- The multimodal fusion approach improved overall prediction reliability compared to single-input methods.

Input Type	Detected Emotion	Confidence Score	System Output
Text Input ("I feel very happy today")	Joy	0.91	Positive Feedback
Text Input ("I am stressed and tired")	Sadness/Stress	0.87	Coping Suggestions
Facial Image (smiling face)	Happy	0.93	Mood Logged
Facial Image (neutral face)	Neutral face)	0.78	No Intervention

C. System Performance Analysis

The performance of AI Thera was evaluated using the following metrics:

- **Accuracy:** The system demonstrated high classification accuracy due to the use of pre-trained transformer and CNN models.
- **Response Time:** Real-time processing was achieved with minimal latency (1–3 seconds).
- **User Experience:** The system provided intuitive feedback and meaningful insights, improving user engagement.

The multimodal approach significantly enhanced the robustness of emotion detection. In cases where one input modality was ambiguous, the other modality contributed to a more accurate prediction.

D. Observations

- The system performs well under normal lighting and clear facial visibility conditions.
- Text-based analysis is highly effective for expressive inputs but may vary with ambiguous language.
- Combining both modalities improves overall system reliability and reduces misclassification.

E. Summary of Results

The experimental results demonstrate that AI Thera is capable of:

- Accurately detecting emotions from real-world inputs
- Providing real-time feedback and insights
- Enhancing emotional awareness through data-driven analysis

Overall, the system achieves a balance between **accuracy, efficiency, and usability**, making it suitable for practical deployment in emotional wellness applications.

VI. LIMITATIONS

Despite its effectiveness, the AI Thera system has certain limitations that must be acknowledged.

A. Input Variability

The accuracy of emotion detection depends on the quality of user inputs. Text inputs with ambiguous language or sarcasm may lead to incorrect classification. Similarly, facial emotion recognition can be affected by poor lighting, occlusions, or low-resolution images.

B. Limited Context Understanding

While the system detects emotions effectively, it does not fully capture deeper psychological context or long-term behavioral patterns. Emotions are dynamic and influenced by multiple external factors that may not be fully represented in a single input instance.

C. Real-Time Constraints

Although the system performs near real-time analysis, performance may vary depending on system resources. High computational load or network latency can slightly affect response time, especially during simultaneous multimodal processing.

D. Dependency on Pre-trained Models

AI Thera relies on pre-trained models, which may introduce bias based on the training data. Cultural and linguistic variations in emotional expression may not always be accurately captured.

VII. PERFORMANCE VISUALIZATION

The system outputs confidence scores representing the probability of detected emotions. Higher confidence values indicate stronger certainty in classification. A sample observation is illustrated below:

- Joy: ~0.90
- Sadness: ~0.85
- Neutral: ~0.75
- The results demonstrate that the system can clearly differentiate between emotional states and assign meaningful confidence levels, enabling reliable feedback generation.

VIII. PERCEPTUAL QUALITY AND USER EXPERIENCE

AI Thera focuses not only on accuracy but also on user experience. The system interface is designed to be intuitive and responsive, ensuring ease of use for both technical and non-technical users.

User feedback from initial testing indicates that:

- Emotion detection results are **easy to understand**
- Suggestions are **relevant and helpful**
- The dashboard provides **clear visualization of emotional trends**

Unlike traditional systems that only display raw outputs, AI Thera emphasizes **interpretability and usability**, making it more accessible to a wider audience.

COMPARISON WITH EXISTING SYSTEMS
COMPARISON WITH EXISTING EMOTION ANALYSIS SYSTEMS

System	Multimodal Input	Real-Time Feedback	Personalization	Web Interface
Woebot	No	Yes	Limited	Yes
Replika	No	Yes	Moderate	Yes
Traditional NLP Models	Text Only	No	No	No
AI Thera (Proposed)	Yes (Text + Image)	Yes	High	Yes

AI Thera distinguishes itself by combining **multimodal emotion detection with personalized feedback**, providing a more comprehensive solution compared to existing systems.

IX. CONCLUSION AND FUTURE WORK

AI Thera addresses a critical need in the domain of emotional wellness by providing a real-time, AI-driven platform for emotion detection and personalized feedback. The system successfully integrates text and facial analysis to deliver accurate and meaningful insights into user emotions.

Experimental results demonstrate that AI Thera achieves reliable performance under real-world conditions, with fast response times and effective user interaction. The multimodal approach enhances accuracy and ensures robustness across diverse input scenarios.

Future work will focus on:

- Improving contextual understanding using advanced NLP techniques
- Expanding support for additional emotional states
- Integrating voice-based emotion detection
- Enhancing model accuracy through fine-tuning on diverse datasets
- Developing mobile and cross-platform applications

In conclusion, AI Thera represents a scalable and impactful solution for emotional well-being, combining artificial intelligence with user-centric design to support mental health in a practical and accessible manner.

REFERENCES

- [1] World Health Organization, "Mental health: Strengthening our response," WHO, 2022.
- [2] American Psychiatric Association, "Autism spectrum disorder," APA Publishing, 2021.
- [3] B. Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool, 2012.
- [4] S. Shatte, D. Hutchinson, and S. Teague, "Machine learning in mental health: A systematic review," J. Med. Internet Res., vol. 21, no. 5, 2019.
- [5] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL, 2019.
- [6] P. Ekman, "Facial expression and emotion," American Psychologist, vol. 48, no. 4, 1993.
- [7] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. NIPS, 2012.
- [8] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proc. CVPR, 2017.
- [9] Hugging Face, "Transformers documentation," 2023.
- [10] D. Baltusaitis, C. Ahuja, and L. Morency, "Multimodal machine learning
- [11] A survey and taxonomy," IEEE Trans. Pattern Anal. Mach. Intell., 2019. [12] R. Picard, Affective Computing, MIT Press, 1997.



International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 9.971