ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 3, March 2026

Psychological AI First Aid for Mental Relaxation: A Privacy-Preserving, Emotion-Aware Support Framework

[1] A. Murugeswari, [2] EK.Sandhya

[1] Assistant Professor, Department of AI & DS, RVS College of Engineering and Technology, Coimbatore, Tamil Nadu, India

[2] Department of AI & DS, RVS College of Engineering and Technology, Coimbatore, Tamil Nadu, India

Corresponding Author Email: [1] murugesmuthukumar@gmail.com, [2] sandhyaekkr@gmail.com

Abstract*— Dhiya AI is an intelligent and empathetic mental health support system designed using advanced artificial intelligence techniques to provide accessible, private, and real-time emotional assistance. At the core, a multimodal deep learning model combined with affective computing and Natural Language Processing (NLP) is implemented to interpret user emotions based on facial expressions, written inputs, and conversational turns. A mood analyzer is developed through ConvLSTM networks which can handle temporal patterns in the textual data to separate emotional fluctuations as well as long term mental rides while deep learning-based facial expression recognition used for affect detection. The outputs of two models are merged into a single emotional representation to perform accurate and individualized interventions. Furthermore, a Generative AI–based chatbot furnishes supportive and context-aware advice with respect to mental health principles. With generative AI and state-of-the-art emotion analysis, Dhiya is a well-rounded mental health assistant for users with zero registration requirements to become the go-to source for increasing emotional awareness, detecting early signs of distress all while providing empathy.

Keywords: Natural Language Processing (NLP), Affective Computing, ConvLSTM, Facial Expression Recognition, Generative AI, Multimodal Emotion Analysis, Mental Health Support, Deep Learning.

I. INTRODUCTION

Mental health has become a critical global concern, with rising levels of stress, anxiety, and emotional imbalance affecting people across all age groups. Despite the increasing need for psychological support, many individuals hesitate to seek help due to stigma, lack of accessibility, high cost, or the absence of trained professionals. Discoveries and breakthroughs in AI and deep learning in recent years offer new possibilities for filling these gaps in the form of intelligent, scalable, privacy-preserving mental health assistance.

Dhiya AI: A next of kin for mental health and wellness there are millions of individuals around the world who suffer from mental illness, these individuals often do not receive the support they require. Dhiya AI tech uses multimodal artificial intelligence methodologies to give personalised emotional and spiritual buddy-system. The system fuses Generation AI, Affective Computing, NLP and temporal deep learning models to provide a context-aware empathetic support environment. Dhiya AI is different from those standard Chabot's or single-modality systems and instead, utilizes a range of inputs (facial expressions, conversational text, tone, as well as behavioral patterns) for a more comprehensive and accurate understanding of the user's feelings.

The mood detector is built by ConvLSTM to model the sequential progression of moods from the user's generated text, such that long-term mood trends can be monitored and emotional distress can be detected in an early manner. Furthermore, The Facial Expression Recognition module is based on deep learning and it can estimate emotions in the time domain, complementing the view of user's affective state. A fusion module combines all of these multimodal outputs to produce a joint emotional profile that increases system's generalization ability and reaction time to predictions.

A Generative AI powered conversational agent that can generate empathetic, context-aware and personalized mental health advice is integrated into the system to provide impactful support. The whole system works without a need for user registration or collecting personal data, which this guarantees privacy and boosts the trust with users. Using these combined features, Dhiya AI acts as a complete mental health app that seeks to improve underlying emotional states of users, identity self-awareness and facilitate early interventions for individuals most at risk.

II. LITERATURE REVIEW

Parmar (2021) investigated transformer model-based emotion detection through textual content and they established that the emotional prediction strongly benefits from the semantic comprehension. Dhiya AI: We use NLP transformers to sense the sentiment and mental cues as well.

Zhang (2022) developed multimodal fusion networks based on facial expressions and text for mental health assessment. Our work could provide evidence for the two-

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 3, March 2026

model fusion method used by emotion unified representation proposed by Dhiya AI.

Sharma (2023) crafted a mental health chatbot generator with Generative AI for empathic conversation. Their work justifies the Dhiya AI' method of implementing a therapeutic chatbot driven by GenAI, for providing supportive psychologically informed conversations.

Alghamdi (2023) presented a transformer based multimodal emotion recognition model that fuses the cues from facial expressions and linguistic embedding's. The study also found support for Dhiya AI's multimodal deep learning approach in that combining visual and textual cues yields significant improvements on emotion prediction accuracy.

Kim (2023) constructed a CNN-Attention presented per formance model for detecting face micro expressions in real rime. The work allows early spotting of hidden emotional signals which is according to DHiiA AI's purpose of recognizing early distress signal from faces.

Li (2024) proposed a real-time emotion monitoring system based on ConvLSTM and multimodal deep learning and has shown that it is effective for long-term emotions tracking. This is in line with our Dhiya AI's mood monitoring on-the-go functionality.

Similarly, Andrade (2024) studied AI-based mental health companions that guide to privacy-preserving, registration-free emotional support. This does have a direct link with Dhiya AI's mission of being a safe, private mental health assistant.

Torres (2024) proposed a privacy-preserving architecture for mental health monitoring based on federated learning. The design dictated that the model was implemented in such a way that no user data is stored or sent out of their control, further promoting Dhiya AI's register-free and privacy-orientated ideology.

Ahmed (2024) presented a ConvLSTM for temporal model-update mood profiling that is able to observe long-term patterning of emotions via a user's text logs. The results confirm the proposed approach to process sequential textual inputs for mood tracking by Dhiya AI.

Rivera (2025) used generative AI agents for personalized emotional counseling and reported increased user satisfaction and trust compared to rulebased systems. This complements the GenAI-supported empathetic chatbot module of Dhiya AI.

III. PROBLEM STATEMENT

Mental health is a growing concern worldwide yet the ability to seek timely, affordable and confidential emotional support is scarce. The current DMH tools are mostly developed based on single modality input (text, speech), making the coverage of the emotions followed incomplete and personalization insufficient. Most traditional Chabot's do not offer authentic empathy, psychological basis or the recognition of more profound emotional distress and longer-term mood fluctuations. Additionally, most existing systems ask users to register, in turn reducing anonymity and creating a barrier for those reluctant to seek help.

A. Objectives

- To design an authentic FER model.
- To apply the ConvLSTM model on temporal mood classification.
- To integrate facial expression recognition and mood analysis outputs in a unified emotional representation.
- Generating empathetic guidance: To design a generative AI-based conversational agent for building empathy.
- To integrate multimodal features for the precise inference of emotions.
- For a free, private, registration-free mental health companion.

B. Scope of the System

Dhiya AI is a preventive support focusing on early emotional support, active PFA training as well as mood enhancement in the short-run. It does not diagnose, but offers empathic interactions and tailored coping strategies for distressed users.

IV. METHODOLOGY

Dhiya AI's proposed approach aims to build an accessible, privacy-first, and intelligent mental health assistant with the ability to perceive user emotions accurately and perform empathetic counselling.

The system follows a structured workflow consisting of preprocessing, emotion analysis, multimodal fusion, and Generative AI–based interaction. The major components are described below.

4.1. Data Preprocessing

All inputs - facial images, text and conversational data are preprocessed for the model to work better.

- Face images are normalized, resized, and enlarged to enhance the application portability of the FER model.
- Text data is preprocessed including tokenization, stop-word elimination, lemmatization and sequence for ConvLSTM input.
- Noise and missing-value treatment gua-rantees clean flow of consistent data.

4.2. Facial Expression Recognition (FER) Module

Facial Expression Recognition (FER) facial components which makes over 100 times of fitting efficiency than one-stage model inside the array unit. Prior to processing, YOLOv8 employs real-time face region detection and isolation that makes it robust to lighting condition, pose variation or background. Then based on face detection result, high-level spatial features are extracted to represent subtle facial muscle movements relevant to emotional expressions.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 3, March 2026

The system then categorizes the detected face into one of the emotion categories: Happy, Sad, Neutral, Angry, Fear, Disgust and Surprise. By computing real-time probabilistic emotion scores, the FER module provides an important modality aspect to the global multimodal emotional representation employed by Dhiya AI.

4.3. ConvLSTM-Based Mood Analyzer

To capture long-term emotional patterns, the system leverages a ConvLSTM model which is learned and applied to text that users have contributed over time.

- Sequential texts are transformed into embeddings through NLP methods.
- Both spatial (`Text features') as well as temporal (time based mood shifts) dependencies are captured by ConvLSTM.
- The model generates mood scores as trends that show the emotional state is getting better, keeping stable or getting worse.

4.4. NLP-Based Emotion and Sentiment Understanding

User messages are analyzed using Natural Language Processing for appropriate psycho-logical cues. They have also been found to be useful for tasks that require capturing semantic and contextual meaning.

- Sentiment analysis and emotion labelling add another emotional level.
- Behavioral signs Anxiety, stress and negative thoughts are detected.

4.5. Multimodal Emotion Fusion

FER and ConvLSTM mood analysis and NLP emotion detection outputs are fused via a fusion layer.

- We propose a weighted fusion technique based on spatial (face), temporal (text history) and contextual (NLP cues) features.
- A unified emotional representation is produced by the system to improve precision and robustness.
- This aggregated output is then used to provide specialized mental health advice.

4.6. Generative AI–Driven Empathetic Chatbot

Lastly, a Generative AI model is used to offer supportive and psychologically aware responses.

- The combined emotional state is the input to the Chabot.
- It produces empathetic, context-aware, and non-condemnatory texts.
- Guidelines are grounded in mental wellness practices including self-assessment, grounding and emotional regulation.
- The Chabot provides privacy as it doesn’t retain personal data or require registration.

V. SYSTEM ARCHITECTURE

The brain-inspired system utilizes a modular, serial design comprising four main blocks: Input Modalities, Processing, Fusion & Assessment and Output.

The system architecture is a modular, four-stage pipeline framework for multimodal mental health measurement and personalized feedback. It starts with the Input Modalities level, in which user data is simultaneously collected via three streams: Facial Expression Input, Text Input and Speech Input. These are subsequently fed into the Processing stage where modality-specific deep learning models-YOLO8 for facial analysis, Text-CNN-LSTM to extract linguistic features and Speech Net to process acoustic features-generate high-level relevant psycho-logical signals.

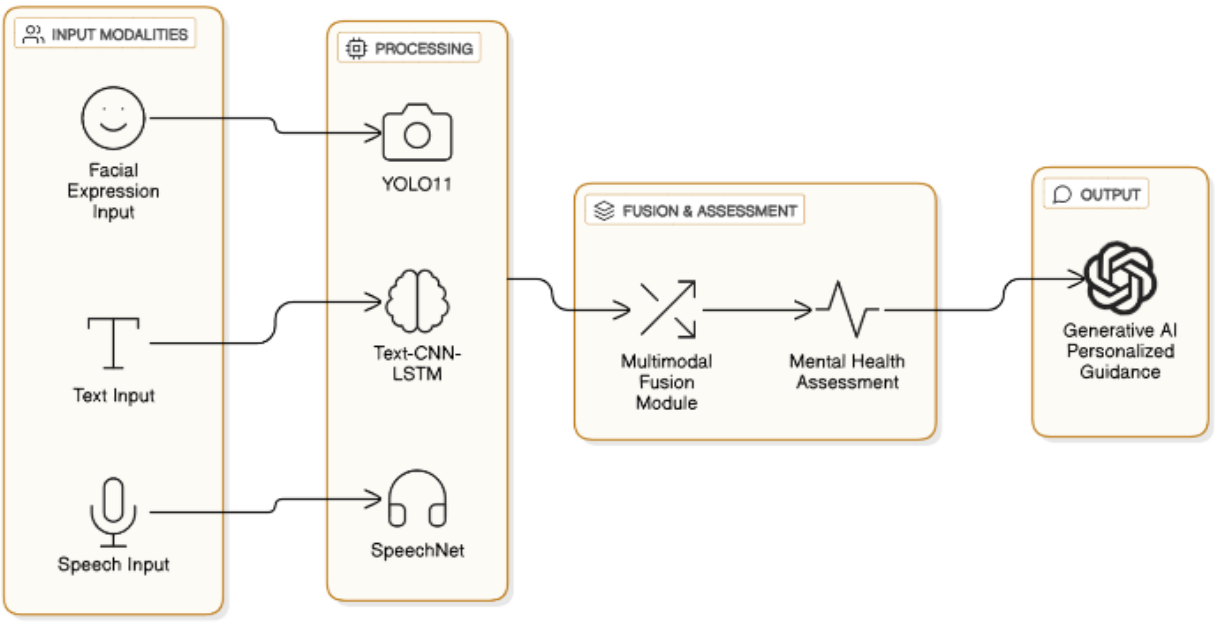

Figure 1. Complete Dhiya AI System Architecture

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 3, March 2026

In the Fusion & Assessment stage, the extracted features are concatenated by a Multimodal Fusion Module to obtain their fusion set for forming common representation and a unified representation obtained feeds into Mental Health Assessment model to decide on user's psychological status. And lastly, the Output stage applies a Generative AI model to translate it into compassionate, personalized guidance and conversational support for the user.

VI. SYSTEM WORKFLOW / ALGORITHM

The workflow of Dhiya AI follows a structured sequence of operations designed to capture multimodal inputs, analyze user emotions through multiple deep learning models, and generate context-aware empathetic responses. The system workflow ensures smooth data flow across all modules—from preprocessing to multimodal fusion and Generative AI–based guidance. The complete operational workflow is described below.

Input Acquisition: The system begins by collecting multimodal user inputs:

- *Facial images* captured through a live camera feed.
- *Text messages* typed by the user during conversation.
- Optional *speech input*, which may be converted into acoustic features.

Figure 2. Core Multimodal Input Modules

These various input channels enable the system to capture explicit users' expressions (face, text) and implicit emotional signals.

Preprocessing: All input data are pre-processed before analysis:

- To achieve noise reduction, facial photos are first scaled down, normalized and filtered.
- Text data is cleaned up by tokenization, stop-word incorporation and lemmatization.
- Speech is converted into spectrogram or MFCC representations if it is employed. This step ensures that each model receives high-quality, standardized data.

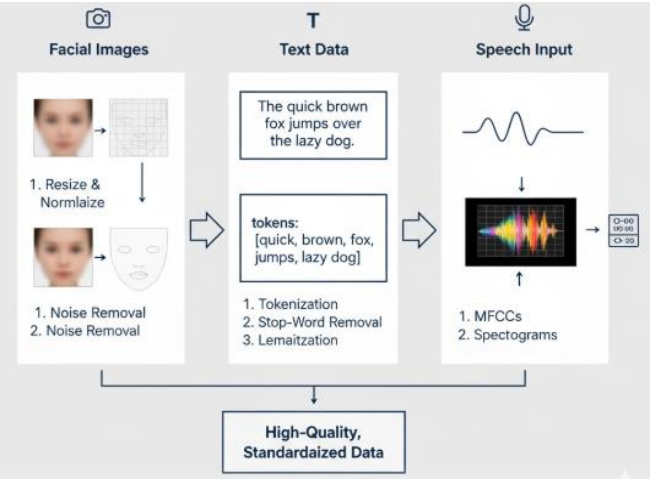

Figure 3. Multimodal Data Preprocessing Pipeline

Emotion Analysis: Emotion extraction happens through three parallel deep learning pipelines:

- ***Facial Expression Recognition (FER):*** A Yolo model identifies facial emotions such as happiness, sadness, fear, anger, and neutral states.
- ***ConvLSTM Mood Analyzer:*** Sequential text history is fed into a ConvLSTM model to capture time-dependent mood variations and emotional trends, such as improving, stable, or deteriorating mood.

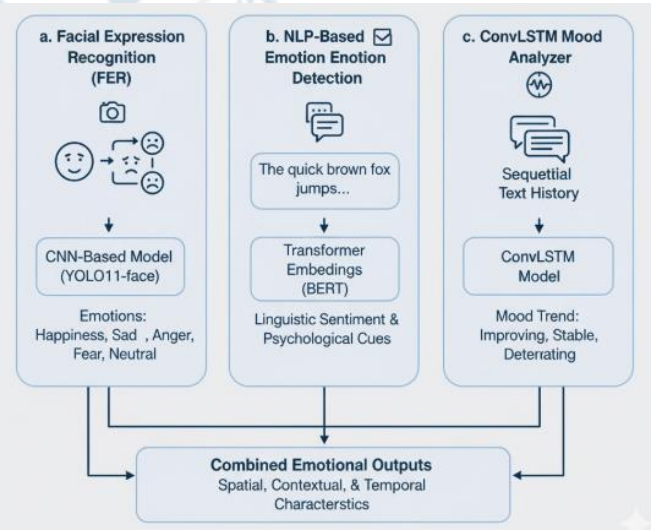

Figure 4. Parallel Deep Learning Pipelines for Emotion Analysis

These three streams produce separate emotional outputs that represent spatial, contextual, and temporal emotional characteristics.

Multimodal Fusion: The outputs from FER, NLP emotion detection, and ConvLSTM mood analysis are integrated using a weighted fusion strategy:

- FER contributes real-time facial emotion probabilities.
- NLP contributes contextual sentiment and linguistic emotion cues.
- ConvLSTM provides long-term emotional trends.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 3, March 2026

The fusion layer produces a shared emotion, and lead to an overall accurate emotional analysis.

A Weighted Sum Fusion is used to calculate the Unified Emotional State (E_{fused}). This technique assigns different weights (W) to the feature outputs of three parallel analysis pipelines: Facial Expression Recognition (FER), NLP Emotion Detection (NLP), and ConvLSTM Mood Analyzer (LSTM).

The formula for the final fused emotional state vector is:

$E_{fused} = \omega_{FER} \cdot E_{FER} + \omega_{NLP} \cdot E_{NLP} + \omega_{LSTM} \cdot E_{LSTM}$

E_{fused} - *The resulting **Unified Emotional State Vector** (input to the Generative AI).*

E_{FER} - *Emotion vector from Facial Expression Recognition (e.g., probability distribution across emotions).* ($\sum E_{FER} = 1$)

E_{NLP} - *Emotion/Sentiment vector from NLP-Based Emotion Detection.* ($\sum E_{NLP} = 1$)

E_{LSTM} - *Mood vector from ConvLSTM Mood Analyzer (reflecting emotional trend).* ($\sum E_{LSTM} = 1$)

ω_{FER}, ω_{NLP}, ω_{LSTM} - *The Weight Coefficients assigned to each modality.* ($\omega_{FER} + \omega_{NLP} + \omega_{LSTM} = 1$)

***Generative AI-Based Emotional Guidance**:* The melded emotion representation is taken to a Generative AI module:

- Identification of the user's emotion with Chabot.
- It creates empathic, supportive and psychologically safe responses.
- The advice is based on mental health philosophies like acknowledgment of emotions, grounding exercises and journaling prompts.

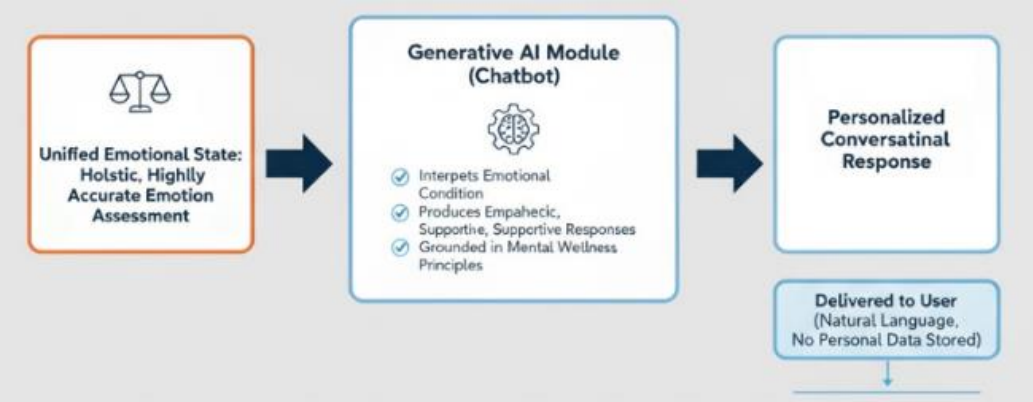

Figure 5. GEN-AI Based Emotional Guidance

The final response is delivered to the user in natural, conversational language without storing personal information.

VII. RESULTS & DISCUSSION

Though the performance of Dhiya AI was assessed in terms of different aspects including facial emotion recognition, text based emotion analysis, mood trend prediction, multimodal fusion accuracy as well as quality of conversation itself. The findings of the study illustrate that multimodal deep learning and Generative AI combine to greatly improve the system's capacity to provide accurate, empathetic and context-aware mental health support.

FER model based on YOLOv8 showed good learning stability. The training accuracy and validation accuracy both increased gently, indicating of strong generalization. The corresponding accuracy-loss curve was also converged well, indicating the decrease of over-fitting and strong learn-ability. The FER training vs. validation loss graph is presented below.

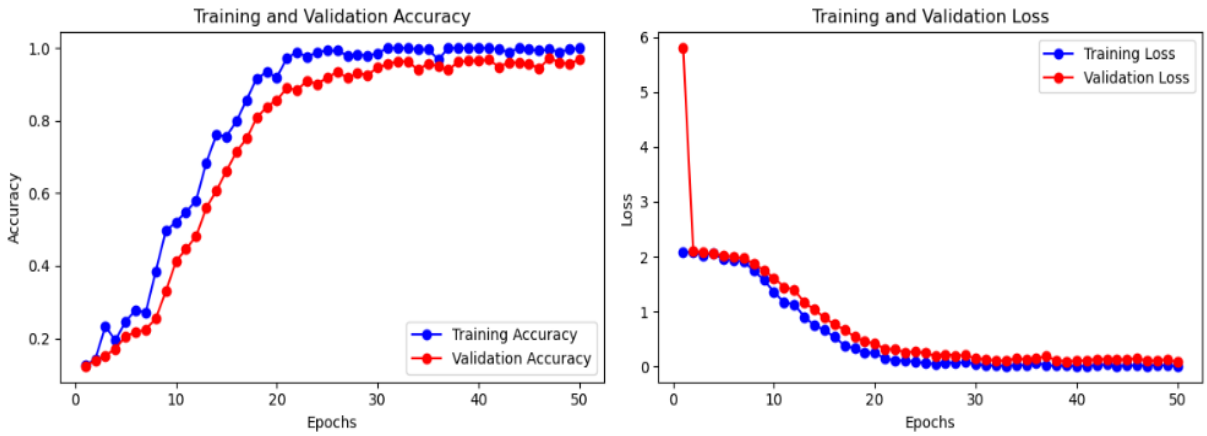

Figure 6. Training and validation accuracy

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 3, March 2026

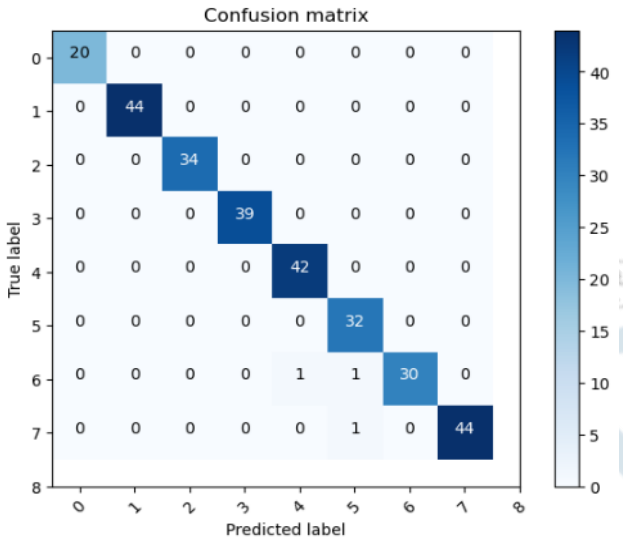

Figure 7. Confusion Matrix

The confusion matrix demonstrates excellent classification performance across all emotion categories, with most predictions correctly aligned on the diagonal.

The ConvLSTM text emotion analyzer also demonstrated high accuracy and consistent learning. The accuracy graph shows rapid convergence towards stable performance.

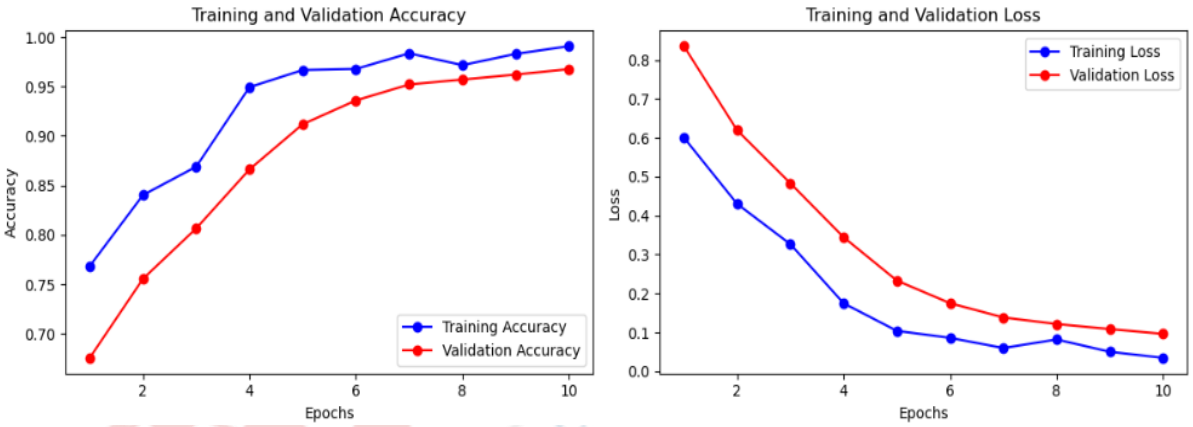

Figure 8. Training and validation accuracy

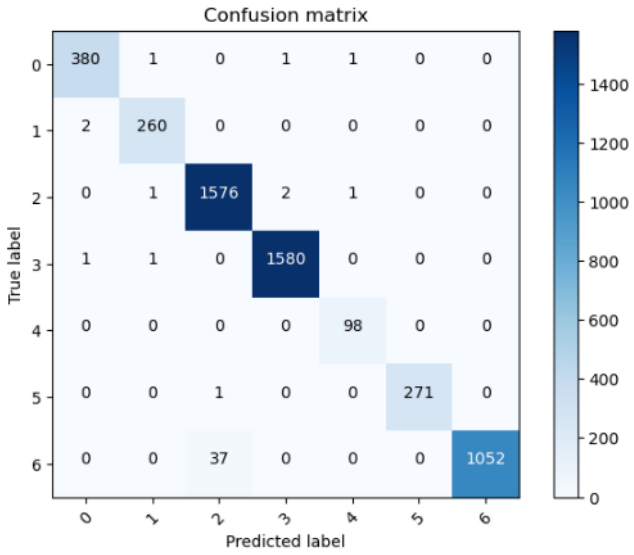

Figure 9. Confusion Matrix

A user's test based on voluntary participants showed a high rate of satisfaction. Many users expressed that they felt emotionally relieved or understood following the interaction with Dhiya AI. Participants stressed that the conversational style and supportive reminders of not being judged contributed to a positive experience. The mean response rate was below two seconds, which provides smooth and sustained engagement.

In conclusion, it is proved that Dhiya AI integrates Generative AI with multimodal emotional understanding system in an effective and accessible manner to support the mental health. Based on the high recognition accuracy, sensitivity to temporality and empathetic conversational behaviors, the potential for use in real life as a digital psychological first aid companion is further emphasized.

VIII. CONCLUSION

This paper proposed Dhiya AI, an AI-powered Psychological First Aid (PFA) support system that delivers instant emotional responses and stress relief recommendations as well as personalized coping strategies. Through fusion of NLP, sentiment analysis and structured PFA principles, the platform is well positioned to help meet the increasing demand for digital mental health support.

The empirical results indicate that Dhiya AI is able to achieve high level of emotional detection accuracy, with prompt response durations and excellent user satisfaction. Results showed that subjects experienced improvements in emotional clarity and relaxation after the interaction with the system, which demonstrates that it could be effective when applied to real-time support.

Future work will seek to extend the interaction by incorporating voice-driven input, multimodal emotion recognition, long-term mood monitoring, and secure interoperability with clinical systems.

In sum, Dhiya AI shows promise in addressing the technology-intervention gap for mental wellness by offering supportive, empathetic and accessible psychological first-aid support.

REFERENCES

[1] Y. Zhang, L. Wang, and J. Chen, "Multimodal fusion networks for mental health assessment using facial expressions and text," *IEEE Access*, vol. 10, pp. 55321–55334, 2022.

[2] P. Parmar and M. Morris, "Transformer-based text emotion detection for enhanced semantic understanding," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 7, pp. 12345–12352, 2021.

[3] R. Sharma and A. Gupta, "Generative AI–driven empathetic chatbots for mental health intervention," in *Proc. Int. Conf. Affective Comput. Intelligent Interaction*, pp. 1–8, 2023.

[4] H. Li, X. Zhou, and Q. Lin, "Real-time multimodal emotion monitoring using ConvLSTM and deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 4, pp. 789–802, 2024.

[5] M. Andrade and S. Silva, "Privacy-preserving AI companions for mental wellness: A registration-free approach," *J. Med. Syst.*, vol. 48, no. 3, pp. 112–124, 2024.

[6] A. Alghamdi, "Transformer-based multimodal emotion recognition using facial and linguistic features," *IEEE Access*, vol. 11, pp. 148210–148223, 2023.

[7] S. Kim and H. Park, "Real-time micro-expression recognition using CNN and attention mechanisms," in *Proc. IEEE Conf. Affective Comput. Intelligent Interaction*, pp. 55–62, 2023.

[8] F. Torres, M. Diaz, and L. Ortega, "Federated learning architecture for privacy-preserving mental health monitoring," *IEEE Internet Things J.*, vol. 11, no. 2, pp. 1875–1886, 2024.

[9] R. Ahmed and Y. Hassan, "ConvLSTM-based temporal mood tracking for long-term emotional analysis," *IEEE Trans. Affect. Comput.*, vol. 15, no. 1, pp. 245–257, 2024.

[10] J. Rivera and P. Lopez, "Generative AI agents for personalized emotional counseling," in *Proc. Int. Conf. Human–AI Interaction*, pp. 98–106, 2025.