A Privacy-preserving Mobile Framework for On-device Emotion Recognition and Personalized Emotional Support

Daniel S. Tang [1], Ling S. Lew [2,*], and Jia L. Tan [3]

[1] Faculty of Information Science and Technology (FIST), Multimedia University (MMU), Melaka, Malaysia
[2] Centre for Advanced Analytics, CoE for Artificial Intelligence, FIST, MMU, Melaka, Malaysia
[3] A-Level Department, Sunway College, Bandar Sunway, Selangor, Malaysia
Email: siqi123987qaz@gmail.com (D.S.T.); sllew@mmu.edu.my (L.S.L.); avatanlj@gmail.com (J. L.T)
*Corresponding author

***Abstract*—Recently, there has been a surge in the number of mental health support apps available to combat stress and improve emotional health. But most mobile applications are based on the processing of sensitive biometric information in the cloud and are self-reported, which compromises privacy, requires users to rely on network connectivity, and reduces user engagement. Moreover, few mobile development frameworks offer real-time emotion recognition on-device with multiple user inputs and personalized emotional support. The study proposed a novel mobile framework called the Feeling Diagnostic Application (FDA) for performing facial emotion recognition and sentiment analysis on text without sharing biometric information to third-party servers. The proposed framework combines a custom-trained Convolutional Neural Network (CNN), which is deployed using TensorFlow Lite for face emotion recognition; a rule-based sentiment analysis engine for text input; a high-risk emotion detector for facial expressions; and an adaptive recommendation engine for providing personalized multimedia content to the user based on the detected emotion. All the inferences are executed locally, thereby enabling offline operation and providing user privacy while minimizing security concerns with cloud-based inference. Experimental evaluation using 250 facial expression samples collected from 50 participants and 50 manually labeled text samples achieved 71.6% facial emotion recognition accuracy, 74.0% text sentiment analysis accuracy, 88.5% high-risk expression detection accuracy, and an average inference latency below 100 ms on a consumer-grade Android smartphone.**

***Keywords*—mental health, emotion recognition, privacy-preserving mobile application, on-device artificial intelligence, sentiment analysis, edge computing**

I. Introduction

Uncontrolled stress is now considered a significant potential risk factor for emotional unbalance and consequently associated with mental health problems like anxiety and depression and low productivity in learning and working situations. Agarwal *et al.* [1] emphasize the need for urgent intervention tools that can bridge the gap between rising mental health demands and limited clinical resources to address the escalation of these issues to a high burden on the population. While digital mental health solutions are typically less expensive than conventional counseling, many solutions on the market do not have the ability to maintain long-term use. In addition, financial restrictions and social stigma against help-seeking are still significant barriers to the uptake and effectiveness of these tools.

To illustrate this practical trade-off, consider a user experiencing a gradual build-up of stress during a workday. The user is either expected to report their emotions by journaling or using a facial analysis tool that analyzes sensitive biometric data via cloud services.

These practical constraints are in line with previous digital mental health studies. Koh *et al*. [2] have found that users are displeased with the lack of personalization of recommendations based on their emotional state. In addition, many applications rely on repetitive logging of moods or journaling, which can cause fatigue over time. Another point of concern is privacy, as cloud-based processing architectures have to rely on the use of sensitive biometric data that has to be transmitted and stored on external servers, and this will hinder adoption by highly privacy-sensitive users [3].

All these restrictions point to a larger gap in research. Current methods are mostly aimed at a single goal, like privacy, accessibility, or emotion recognition, and not at embedding them in a cohesive mobile infrastructure that enables flexible interaction, complete processing onboard, and real-time personalized emotional support. There are limited solutions available that integrate all these functions into a light-weight architecture. The above gap has been the motivation for the present study.

To address these challenges, this paper proposes a privacy-centric emotional diagnostic framework developed to be accessible, on-device, and real-time effective. Guided by this motivation, the study is structured around the following research question:

Manuscript received May 20, 2026; revised June 16, 2026; accepted August 4, 2026; published November 10, 2026.

Is it possible to have a holistic, fully on-device mobile framework for facial emotion recognition and text-based sentiment analysis that provides real-time, private, and tailored emotional assistance without sending sensitive biometric data to external servers?

In order to answer this question, the proposed system offers dual-input interaction with the possibility for the user to communicate with the system using both facial and textual input [4] to cater for the different preferences in communication and thus to minimize interaction barriers. The two input modalities are designed as alternative user interaction methods rather than simultaneous fusion mechanisms, allowing users to select the input mode that best suits their context and preference. All emotion analysis can be performed on-device by exploiting edge-based processing [3] to ensure the privacy and confidentiality of facial data, along with improving user trust, security, and usability. These design elements create a mobile structure that is both private, accessible, and user-centered and can provide real-time emotional self-support without any reliance on cloud processing of biometric data.

To clearly demonstrate how these issues are dealt with in the present study, Table I shows a mapping of the major problems, the associated objectives, and the resulting contributions of the proposed framework.

TABLE I. MAPPING OF PROBLEMS, OBJECTIVES, AND RESULTS

Problem	Objective	Result
Privacy concerns over cloud storage of sensitive biometric data	Implement on-device emotion analysis to ensure privacy	Privacy-Preserving Edge AI Architecture
Low engagement and user fatigue due to manual mood tracking	Provide dual-input interaction (facial and textual) for more flexible interaction	Flexible and Accessible Interaction
Lack of real-time, personalized support	Integrate a real-time logic engine for dynamic emotion-based support	Closed-Loop Adaptive Support

Table I summarizes how identified issues in the existing systems for the delivery of digital mental health interventions relate to the decisions made in the design of the proposed framework. Each problem is mapped to a specific objective, which in turn leads to a concrete contribution: on-device processing for improved user privacy, dual-modality input for reduced interaction fatigue, and a real-time logic engine that minimizes the time lag between emotion detection and personalized support delivery.

The major contributions of this study are threefold:

(1) Privacy-Preserving Edge AI Architecture: A mobile architecture that uses on-device facial expression recognition via TensorFlow Lite and rule-based textual sentiment analysis that does not send sensitive biometric data to external servers.

(2) Flexible and Accessible Interaction Design: A dual-input flexible system aimed at alleviating the user fatigue induced by manual logging and ensuring accessibility to users with different levels of connection to the internet.

(3) Closed-Loop Adaptive Support: The integration of a real-time logic engine not only able to detect crises but also to dynamically map emotions recognition to personalized multimedia recommendations to bridge the gap between emotion recognition and personalized support.

II. LITERATURE REVIEW

This section provides a review of existing studies associated with mental health and emotion recognition systems in order to provide the basis for the proposed framework. Prior studies have shown various technology-based solutions to emotion detection and mental health assistance, such as facial expression detection, multimodal sensing, text-based sentiment analysis, physiology monitoring, and general well-being applications.

A. Facial Expression Recognition Systems

In recent years, with the development of computer vision and deep learning, Facial Expression Recognition (FER) has been one of the most widely investigated methods for automated emotion measurement. Early systems relied primarily on convolutional Neural Network (CNN) architectures: Chezer used CNNs to classify the facial expressions of children in a gamified learning environment with high accuracy (90–94%) [5], while MOODMUSIC [6] and the system proposed in Ref. [7] employed lightweight MobileNet-based architectures for real-time facial emotion detection coupled with music recommendation. Although these systems demonstrate strong performance and computational efficiency, they are typically developed for task-specific applications, limiting their applicability to broader mobile emotional support systems.

Recently, the field has moved towards transformer-based architectures that employ the self-attention mechanism to model long-range spatial relations between facial regions instead of only local convolutional filters [8]. These models have been reported to achieve high accuracy on in-the-wild benchmark datasets in comparison with traditional CNNs and mark a larger shift towards the development of higher-capacity, attention-based models for FER. This development has been paralleled by the creation of large-scale benchmark databases like AffectNet [9] that go beyond controlled databases to include a broader spectrum of naturalistic facial expressions.

Although considerable progress has been made, two major challenges continue to impede real-world deployment of FER systems. Firstly, many of the highest-performing FER models, including transformer-based architectures, are more compute-heavy and tend to be tested in cloud or server environments and not on mobile devices. Similarly, the cloud-based Arabic autism-support system in Ref. [10] introduces additional privacy and latency concerns. Secondly, recent works also pointed out the issue of fairness and demographic bias as new research challenges for facial emotion recognition problems, as recognition performance is different among

age groups, genders, ethnicities, and cultural contexts [11]. While significant strides have been made towards the recognition accuracy, achieving reliable generalization across different populations is a crucial avenue for future research.

Recent research has investigated lightweight on-device inference using TinyML and optimizing deployment frameworks to fill the gap between model complexity and deployability on mobile devices. TinyML allows machine learning models to be executed directly on resource-constrained devices, without the need for cloud infrastructure, providing benefits with regard to latency, bandwidth efficiency, and privacy [12]. TensorFlow Lite is an especially popular framework for converting and deploying deep learning models on mobile and embedded devices [13]. Model compression techniques, such as pruning, quantization, and knowledge distillation, have also been adopted to decrease model size and computational complexity and maintain predictive performance of the compression [14]. In general, facial-based systems have achieved high recognition accuracy; however, there is a trade-off between recognition performance and lightweight, privacy-preserving mobile deployment. This constraint inspires the use of light-weight on-device inference frameworks in emotion recognition for mobile devices.

B. Multimodal and Voice-Based Systems

Several studies have explored multimodal architectures by fusing multiple data sources in order to increase reliability on emotion recognition when relying on a single modality. Felii is an interactive artificial therapist that integrates facial and voice analysis [15], but it needs a considerable amount of computing power, which could be a challenge to use on a mobile device. MANAS combines natural language processing and voice analysis to identify depression and suicidal ideation but is limited to a select number of mental health conditions [16].

More advanced fusion models integrate physiological signals with visual data: the system in Ref. [17] combines facial video and Electroencephalogram (EEG) signals to detect depression with high accuracy but requires specialized equipment, while EmoData combines audio and motion sensors in an educational setting, achieving comparatively low and unstable accuracy (27–73%) [18]. In addition to these application-specific systems, recent research in multimodal information fusion has demonstrated that the fusion of complementary modalities, such as speech and facial expressions, can significantly outperform the single-modality baselines; a recent fusion system reported an improvement from 76–78% accuracy for single-modality detection to 95% accuracy after fusion [19]. It has been observed that the multimodal fusion generally leads to better recognition performance; however, the extra sensing modalities and fusion methods significantly increase the computation load, hardware needs, and implementation complexity, which makes it hard to implement on lightweight mobile devices. These findings suggest that lightweight mobile frameworks may offer a practical alternative when computational efficiency and deployability are prioritized over complex multimodal fusion.

C. Text-Based Sentiment Analysis Systems

Text-based emotion detection is still the commonly used method for its low hardware requirement and ease of implementation. Traditional approaches rely on rule-based or lexicon-based methods: MediBot is a conversational NLP-based chatbot that offers emotional support and self-care guidance using rule-based dialogue logic [20], while Padate *et al.* [21] shows that informal messaging platforms can be analyzed using Pointwise Mutual Information (PMI) and normalized PMI to detect emotions such as anger and surprise, although linguistic ambiguity and sarcasm reduce classification reliability.

Recently, transformer-based language models, like BERT and RoBERTa, have received growing attention for text classification tasks related to sentiment and mental health since they can learn word relations in context rather than relying on a static weight assigned to each keyword. A recent comparative evaluation of BERT and RoBERTa for sentiment classification of social media data pertaining to mental health indicated that the two models gave significantly better performance on contextual and informal language than standard lexicon-based sentiment analysis models [22]. This is a general trend in sentiment analysis towards transformer-based models

In addition to general sentiment classification, safety-related NLP has become a research field for the detection of potentially dangerous expressions like suicidal thoughts in crisis-related text messages. Broadbent *et al*. [23] assessed machine learning models, such as neural models, for the purpose of detecting client suicide risk within text-based crisis counseling interactions and concluded that the neural models were more sensitive in detecting client suicide risk than traditional term-frequency-based approaches, with the downside of a higher false positive rate. This work illustrates the precision–recall trade-off of the automated crisis detection task, especially for use in a safety-critical context.

While these developments are promising, existing methods of text-based methods (rule-based, transformer-based, or safety-focused) have their drawbacks, as they are insufficient to account for nonverbal emotional cues, such as facial expressions and vocal tone, that may be critical for a thorough evaluation of a user's emotional state.

D. Physiological and Wearable-Based Systems

Physiological monitoring approaches offer objective and biosignal-based measures of emotional states (e.g., heart rate, electrodermal activity). Biometric fusion techniques have been used to obtain high detection accuracy, as in the case of MindCare [24], and in Ref. [25] a fusion method is proposed, while in Refs. [26, 27] workplace stress monitoring systems have been successfully tested for stress reduction in real time. Although relatively accurate, all these methods depend on an external sensor or wearable device and cannot be considered cost-effective and scalable in everyday mobile scenarios.

With the sensitivity of the data, privacy-preserving approaches have emerged as an important research direction for physiological and biometric sensing. The federated learning framework proposed by Devi *et al.* [3] provides a promising direction for secure data processing in wearables-based stress monitoring. This approach has since been followed up by a specific federated learning model for mobile mental health sensing by Suruliraj and Orji [28], which focuses on federating a learning model for distributed devices in mental health sensor applications without aggregating the raw sensor signals from the devices. Complementary work has further integrated differential privacy into this pipeline, with Cui *et al.* [29] suggesting a framework for a privacy-preserving, speech-based depression diagnosis system using federated learning and differential privacy and secure aggregation to further ensure user information security during training. This sequence from physiological sensing to federated learning to differential privacy represents an emerging paradigm of distributed and privacy-preserving model training beyond the extremes of all-cloud and all-local scenarios. Other experimental systems such as the nutrition-based emotional intervention from [30] highlight the early stages of development of emerging trends across various disciplines. While physiological signals are objective measures of emotion, reliance on wearable sensors and specialized hardware restricts accessibility for everyday mobile use, making camera-based mobile solutions appealing.

E. General Wellness and Self-Management Applications

General wellness applications tend to focus on habit tracking, mood journaling, and peer support rather than automated diagnosis. Mindset [31] provides users with a way to manually record emotions and reflections, and Mindful Meadows [32] combines the tracking functionality with a community-based support group, but both still require a lot of manual data entry and do not have any built-in capability for adaptive emotional analysis. User interface design has also been found to be a key determinant of engagement: Goyal *et al.* [33] shows that intuitive dashboards and emotionally supportive interface designs have a significant impact on retention rates.

Personalization, in addition to interface design, has been recognized as an important, though less consistently applied, tool to ensure long-term user engagement with mental health apps. In a recent scoping review on personalization and recommendation in mental health apps, the authors reported that while personalization is a commonly reported feature of all apps, its application and effectiveness within the apps in terms of engagement were found to be limited and inconsistent throughout the apps included in the review [34]. This implies that there is an emerging need for both usability and personalization as well as adaptive recommendations for long-term adoption, which are often not sufficiently covered by existing wellness applications.

F. Research Trends and Identified Gaps

In the literature overviewed, facial and physiological systems generally get good recognition accuracy but need a lot of computation or special external hardware, while multimodal and fusion-type systems increase detection reliability but increase system complexity and decrease system accessibility. In contrast to the vision- and biosignal-based systems, text-based ones are light in weight and scalable but are typically less emotional and strong, and many applications in all of these categories are still reliant on cloud infrastructure that can lead to latency and privacy concerns. Recent progress in transformer-based modeling, on-device deployment frameworks like TensorFlow Lite, and privacy-preserving training methods like federated learning and differential privacy are indicative of a growing trend in the field, moving away from the cloud to a more lightweight, privacy-focused, and increasingly context-aware emotion recognition.

TABLE II. COMPARISON OF REPRESENTATIVE EMOTION RECOGNITION AND MENTAL HEALTH SUPPORT SYSTEMS

Study	Modality	Approach	Limitation
Chezer [5]	Facial	CNN/RNN	Educational setting
MOODMUSIC [6]	Facial	CNN	Entertainment only
Music Rec [7]	Facial	MobileNetV2	Limited scope
Transformer FER [8]	Facial	Transformer	High computation, cloud-evaluated
Autistic App [10]	Facial	Cloud AI	Privacy risk
Depression Detection [17]	EEG + Face	CNN+LSTM	Needs equipment
Mindfulness Therapy [15]	Multimodal	LSTM	Heavy computation
MANAS [16]	Text + Voice	ML classifier	Limited conditions
EmoData [18]	Audio + Motion	Fuzzy logic	Low stability
Multimodal Fusion [19]	Facial + Speech	HMM-based fusion	High complexity
MediBot [20]	Text	NLP chatbot	No nonverbal cues
WhatsApp Emotion [21]	Text	PMI/NPMI	Language ambiguity
BERT/RoBERTa Sentiment [22]	Text	Transformer	Not optimized for mobile deployment
Crisis Risk Detection [23]	Text	ML/Neural classifier	High-false-positive rate
MindCare [24]	Physiological	ML fusion	Requires sensors
VR Emotion [25]	Physiological	KNN	Hardware dependent
Wearable Stress [26]	Physiological	Biofeedback	Limited scalability
Call Center Stress [27]	Physiological +Manual	Feedback	Workplace only
Educator Stress FL [3]	Wearables + FL	ML	Small population
Moblie FL Framework [28]	Mobile sensing	Federated Learning	Framework-level, not deployed as full app
Speech FL+DP [29]	Speech	FL + Differential Privacy	High computational
Food Healing [30]	Heart rate	Regression	Experimental
Mindset [31]	Manual	None	Low engagement
Mindful Meadows [32]	Manual + AI	Recommender	Manual only
UI/UX Study [33]	UI survey	Stats	Design only

However, this shift remains incomplete. Overall, such studies suggest that research on emotion recognition has already focused on optimizing individual components such as recognition accuracy, preserving the privacy of the recognized emotions, sensing multimodal emotions, and mobile deployment so far. Few studies try to combine all these requirements in a single mobile solution, leading to real-world compromises in performance, accessibility, computational efficiency, and user privacy. Meanwhile, there has been a growing focus on clinical validation, demographic considerations, and safety-focused assessment, bringing to the forefront critical areas for future research in mobile mental health systems. The trends discussed are summarized in Table II, which compares the modality, technical approach, and main limitation of the systems reviewed.

The results of the studies outlined in Table II can be analyzed critically with three conclusions regarding the nature of the research gap. First, there is a trade-off between the accuracy and the accessibility of the approach considered: the first of three categories (physiological systems [17, 26]; transformer-based facial expression recognition models [8]; or multimodal fusion approaches [19]) is accurate but is intrusive or computationally demanding or requires additional sensing modalities, while manual journaling apps such as [31, 32] are accessible but do not provide objective verification of their accuracy. Second, there is a separation between emotion recognition and user support; existing approaches, such as recent emotion recognition models based on transformers [22] or safety-focused emotion detection systems [23], tend to emphasize detection or recommendation, but not both in a closed-loop emotion recognition or user support system. Third, the use of cloud computing for processing [8, 10] of the biometric data presents latency and privacy issues; although recent studies in federated learning [28] and differential privacy [29] have shown potential solutions, these methods have only been applied to physiological or speech-based sensing pipelines and have not yet been adapted for mobile, multimodal emotion recognition systems like the one suggested in this work. The observations inspire the design of a lightweight mobile framework that integrates emotion recognition with privacy protections on the device, flexible user interaction, and adaptable emotional support in a single architecture. The proposed framework introduced in the following section is aimed at tackling these identified research gaps.

III. Materials and Methods

This section describes the system architecture, multimodal interaction design, machine learning setup, safety mechanism, deployment framework, and experimental protocol of the proposed Feeling Diagnostic Application (FDA). The system has been designed and deployed as a system that preserves privacy and supports two complementary modalities of user interaction: facial expression recognition and sentiment analysis of text, where the emotion analysis is done entirely on-device, in real-time.

A. System Architecture

The Feeling Diagnostic Application was developed by modular three-tier architecture to ensure separation of concerns, scalability, and privacy preservation. Fig. 1 illustrates the overall system architecture.

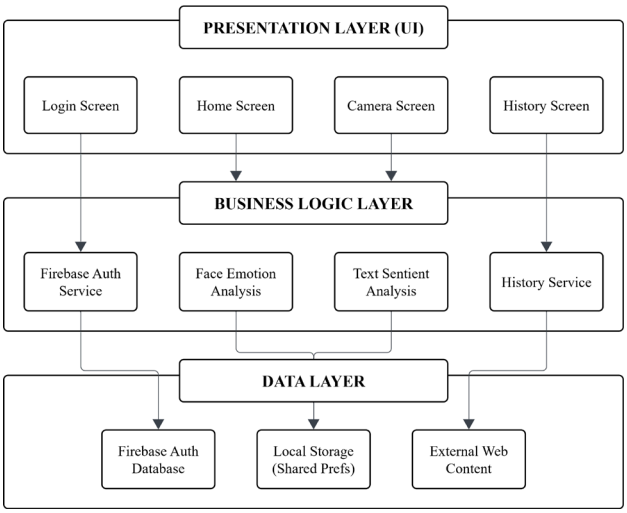


Fig. 1. Overall system architecture.

Three-tier architecture consisting of:
(1) Presentation Layer;
(2) Business Logic Layer;
(3) Data Layer.

The presentation layer handles the interaction with users and consists of the multimodal dashboard, AI face scan interface, text input interface, recommendation page, and emotional history timeline.

The business logic layer conducts all the inference and processing operations, such as Facial Expression Recognition (FER), rule-based sentiment analysis, crisis detection, temporal stabilization, and adaptive recommendations generation. This layer processes all machine learning inference in it with TensorFlow Lite so that there is no need to feed biometric information to the outside server and the analysis of emotions is performed locally.

The data layer is composed of two functionally separate components. Firebase Authentication is only used for user login and account management; it does not receive, process, or store the facial image, emotional record, or sentiment data. All emotion-related data is stored locally on-device in a private sandbox using Android SharedPreferences and is serialized as JSON, including dominant emotion label, probability distribution, and timestamp. Rough face images are not stored to persistent storage at any point along the pipeline. This separation makes sure that while the application relies on a cloud-based identity supply, no biometric or emotional information ever leaves the gadget..

B. Multimodal Input Design

Fig. 2 shows the system enforces dual-modality emotional input comprising:
- Active textual input;
- Passive visual input.

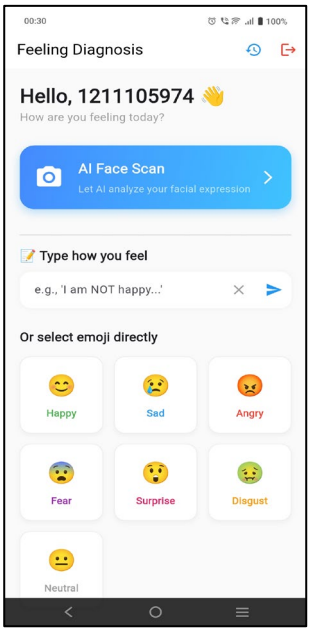


Fig. 2. Multimodal input interface showing text and visual emotion capture modules.

As described above in this paper, the two modalities are not combined as a fused multimodal pipeline but rather as alternate and mutually exclusive interaction modes that the user can select between in any sequence: the user can interact using a facial scan or via text input or interact with both in any order, but the system never makes a combined prediction of facial and textual representations. No early or late fusion or confidence-weighted fusion is performed, and in the event that the emotion results from different modalities are contradictory, the two results are not passed through a reconciliation logic but presented separately as a modality-specific output. This design decision emphasizes flexibility in interactions and lightweight on-device deployment over improvements in accuracy that fusion-based methods might provide, but at the expense of the joint-inference capability that these methods offer.

1) Text-based input pipeline

When users typed in textual descriptions of their emotional state, the system processed the input through a structured multi-stage pipeline:

(1) Tokenization and phrase-level pattern detection;
(2) Keyword-based weighted scoring;
(3) Emoji-to-emotion mapping;
(4) Negation window processing;
(5) Intensity modifier processing;
(6) Contrastive conjunction weighting;
(7) Crisis keyword scanning.

Each detected feature contributed to a weighted score of each emotion. Final scores were normalized in a probability distribution in seven emotion categories.

2) Visual input pipeline

For the visual analysis, users activated a module on the AI system that enabled the camera on the front of the device to scan faces. On the visual pipeline, there were:

(1) Facial boundary detection using Google machine learning kit;
(2) Bounding box-based image cropping;
(3) Grayscale conversion;
(4) Resizing to 48×48 pixels;
(5) Pixel normalization to [0, 1];
(6) Tensor preparation;
(7) On-device CNN inference using TensorFlow Lite.

Predictions were processed by means of a temporal smoothing mechanism, which is described in Section III.

C. Facial Expression Recognition Model

1) Dataset and preprocessing

The facial expression recognition model was trained on the FER-2013 dataset (35,887 grayscale, 48×48-pixel images across seven emotion categories), split into 72% training, 18% validation, and 10% testing. All images were normalized to [0, 1]. Class imbalance of FER-2013 was not explicitly handled when training the data; no class weighting or resampling was applied.

Although the CNN was trained using all seven FER-2013 emotion categories (angry, disgust, fear, happy, sad, surprise, and neutral), only five categories (angry, happy, neutral, sad, and surprise) were included in the deployment evaluation. Preliminary results indicated that the disgust and fear classes had relatively unstable predictions and often mixed up with similar-looking expressions during real-world mobile capture. To enhance robustness and consistency of the prototype system, these two classes were thus excluded from the deployment evaluation.

2) CNN architecture

The face expression recognition model is a sequential CNN consisting of four convolutional blocks, including a fully connected classification head, which sums to 44 Keras layers for the model. Each convolutional block consists of two Conv2D layers (3×3 kernels, "same" padding) with ELU activation and batch normalization, followed by 2×2 max pooling and dropout (rate = 0.2). The four blocks use filter depths of 32, 64, 128, and 256, respectively. The flattened output of this model is then fed into two fully connected layers with 64 neurons that each include ELU activation, batch normalization, a dropout rate of 0.5, and a final softmax layer over seven emotion classes.

3) Training configuration

The training of the models was performed with the following parameters:

- Optimizer: Adam (Adaptive Moment Estimation);
- Loss Function: Categorical Cross-Entropy;
- Batch Size: 64;
- Epochs: 100;
- Early Stopping: Monitored validation loss with a patience of 6 epochs;
- Learning Rate Schedule: ReduceLROnPlateau (factor = 0.5, patience = 3).

Training was done on Google Colab using an Nvidia Tesla T4 GPU. The trained model was converted to the TensorFlow Lite format with the FP32 format for mobile deployment.

4) Model size, computational complexity, and memory footprint

To characterize the model's suitability for edge deployment, the trained network was analyzed for parameter count, computational cost, and memory footprint. There are 1,328,167 total parameters in the model, with 1,325,991 trainable parameters and 2,176 non-trainable batch-normalization moving statistics. The weight footprint for inference is about 5.07 MiB; the saved checkpoint (Adam optimizer state used only during training) is about 15.39 MiB, of which about 10.12 MiB is optimizer state discarded before deployment.

The estimated inference compute is about 117.6 million multiply-accumulate operations (MACs) per forward pass, which is equivalent to about 235 million Floating-point Operations (FLOPs). The maximum single activation tensor at inference (batch size = 1) is 48×48×32, around 288 KiB, and major layer outputs' cumulative footprint is around 1.20 MiB in float32. Actual memory usage will be greater than these estimates because of overhead associated with the TensorFlow Lite interpreter, storage for intermediate computations, and camera and preprocessing pipeline image buffers.

The model is used with standard float32 (FP32) precision throughout and has not been post-trained with quantization or structured/unstructured pruning; the weight sparsity inspection confirmed that the model is not sparse. This is intentional in the current implementation: Although the reported sub-100 ms inference latency is achieved using FP32 TensorFlow Lite deployment, a reduction in weight size from around 5.1 MiB to around 1.3 MiB, potentially at the cost of classification accuracy, can be expected by applying dynamic-range or full-integer (int8) post-training quantization. In the current implementation, neither quantization nor pruning was applied and is labeled as future work.

To test the deployment performance, the TensorFlow Lite inference call was instrumented with a millisecond resolution timer to capture the execution time of the forward pass at runtime. The implementation continuously monitors inference latency throughout operation, providing direct empirical verification of real-time on-device performance.

5) Post-processing and temporal stabilization

To add more stability to the predictions in the real-time scenario, some heuristic rules were performed on the unrefined probability outputs:

- Angry-Neutral Filter: Predictions of "Angry" were conditionally suppressed in favor of "Sad" whenever the "Neutral" confidence was greater than 0.20 to reduce false-positive detection of aggression under ambiguity distributions of confidence.
- Sadness Boost: If the "Neutral" was dominant, but the difference between "Neutral" and "Sad" was less than 0.05 (5%), the system was biased towards "Sad".
- Rare Class Suppression: Predictions ("Disgust" or "Fear") with a confidence score less than 0.70 were suppressed and replaced with the second highest probability class.
- Temporal Smoothing: A rolling buffer containing the most recent 6 predictions was kept, and the final presented emotion was based on the statistical mode of the buffer.

Fig. 3 demonstrates real-time on-device facial emotion detection.

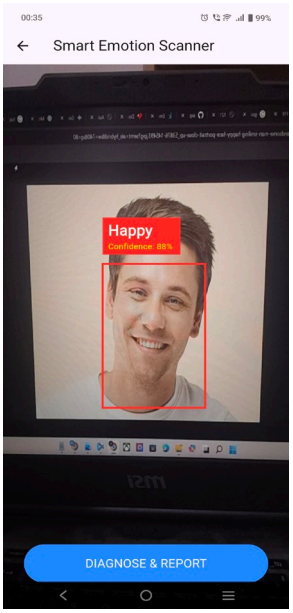


Fig. 3. Real-time on-device facial expression recognition.

D. Rule-Based Sentiment Analysis Engine

The textual sentiment engine determines weighted scores for 7 emotional categories based on a deterministic rule-based approach. Emotion keywords and emojis were given heuristic weights between +1.0 and +2.0 depending on intensity and negation modifiers will reverse the polarity (weight × −1.0) of the immediately succeeding term. Fig. 4 illustrates the text processing workflow.

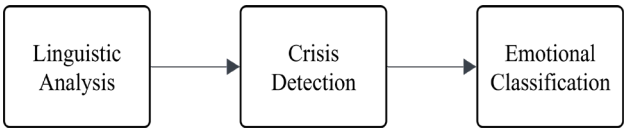


Fig. 4. Text sentiment analysis pipeline.

Processing stages include:

(1) Pattern detection on the phrase level;
(2) Emoji Mapping;
(3) Tokenization;
(4) Intensity Modifier Adjusting;
(5) Negation Window Processing;
(6) Conjunction of Contrastive Re-weighting.

Each element that was identified contributed its pre-defined weight to the pertinent emotion category. Final scores were aggregated and normalized as percentage-based probabilities.

E. Crisis Detection Protocol

A safety-first mechanism is activated prior to the regular sentiment classification, and it follows a two-step deterministic decision-making process.

1) Tier 1 (Immediate-risk pattern matching)

Each sentence in the normalized text is matched against a collection of regular-expression patterns that are known, high-confidence indicators of suicidal intent or self-harm (such as a direct statement of suicide intent, explicit mention of a plan, or common shorthand like "kms"). If the pattern matches any Tier-1 pattern, the input will be automatically classified as crisis-positive without the need for additional scoring. This tier is a hard boolean decision boundary; the presence of an explicit pattern is enough and necessary to escalate immediately.

2) Tier 2 (Weighted risk scoring)

If the text does not contain a Tier-1 pattern, a secondary layer assesses a more general pattern of low specificity risk markers (e.g., expressions of hopelessness, language adjacent to self-harm, indications of a goodbye message, or feeling unsafe with oneself). Each matched pattern is assigned a heuristically determined weight of 1 to 3, based on the severity of that pattern, and the total of the matched weights is calculated to get a cumulative risk score.

The cumulative risk score is computed as

$$RiskScore = \sum_{i=0}^{n} w_i I_i$$

where w_i denotes the predefined weight assigned to the i-th risk pattern and I_i is a binary indicator equal to 1 if the pattern is detected in the input text and 0 otherwise.

The final decision rule is defined as

$$Crisis = \begin{cases} 1, & if\ any\ Tier-1\ patter\ is\ matched \\ 1, & if\ RiskScore \geq 3 \\ 0, & otherwise \end{cases}$$

When this cumulative score reaches a threshold of 3, a crisis classification is triggered. It's a layer that is weighted, so if multiple moderate-risk indicators exist in a single message, even when none of the terms individually qualifies as Tier-1 risk, those terms will trigger escalation.

3) Safety trigger

If a Tier-1 or Tier-2 crisis classification is returned:

- Standard classification of emotions was bypassed;
- The main user interface was replaced by a modal crisis safety dialogue;
- The dialog required explicit user acknowledgment.

The interface provided direct access to the emergency services (such as the 988 hotline) and allowed the user to check on safety status. The system does not automatically alert emergency services, guardians, or other parties on the user's behalf, and escalation is reliant on the user's actions after receiving the resources for crises. This design ensures that the human user retains control over the response and does not create extra risk through automated non-consensual emergency contact, which is appropriate given that the app is a support and signposting tool, not an intervention system.

F. Adaptive Recommendation and Emotional History Tracking

Following successful emotion classification (non-crisis cases), the adaptive recommendation engine was switched on.

Emotion labels were mapped to predefined recommendation categories stored in the internal repository of the application containing multimedia references (music, articles).

The system increased and decreased in a dynamic way:

- Interface theme color;
- Validation message;
- Recommended content.

To conduct longitudinal monitoring, the emotional records stored included:

- Dominant emotion label
- Complete probability distribution
- Timestamp
- Input modality source (or optional user notes)

All records were serialized to the local, application-specific, isolated storage sandbox in the form of a JSON string.

G. Mobile Deployment and Privacy Framework

The application was deployed on Android devices with on-device inference with TensorFlow Lite.

Privacy mechanisms include the following: (1) no persistent storage of raw facial images—camera frames are processed in real time and discarded immediately after inference; (2) local, sandboxed storage of processed emotional records (emotion label, probability distribution, timestamp, and optional notes) via Android SharedPreferences, serialized as plain JSON. SharedPreferences was selected because the application stores only lightweight structured metadata rather than large multimedia files, enabling low-overhead local persistence while maintaining complete offline operation; (3) data isolation on a per-user basis using unique identifiers; and (4) volatile-memory-only biometric inference, with no frame buffering to disk. Firebase is only used for user authentication and is not storing or receiving facial, emotional, or sentiment data. Facial images and text input are only collected explicitly when the user initiates a corresponding interaction; no passive background collection of emotional data is carried out.

Note that there is no transmission of raw biometric data externally or persistence to storage anywhere, however the locally stored emotional metadata is not encrypted at rest in the current implementation; it is stored in the application's private sandbox, but is not protected by an additional encryption layer like EncryptedSharedPreferences or SQLCipher. As the application is not a medical application, but rather a research prototype, encrypted local storage was not a requirement for the current version. Future work will implement Android EncryptedSharedPreferences or other secure storage solutions to further enhance the security against device-level attacks.

The proposed FDA is intended to be an emotional self-monitoring and support app and is not meant to be a

medical diagnostic or alternative to a professional psychological evaluation. The crisis detection module acts as a "caution" safety mechanism that will urge users to contact the proper emergency or professional help when high-risk language is detected. Clinical assessment, diagnosis, and emergency intervention remain the responsibility of qualified healthcare professionals. The application does not make autonomous clinical decisions, and users are responsible for obtaining medical or medical emergency assistance if needed.

Performance measurements were obtained on a Vivo V40 (Snapdragon 7 Gen 3, 12 GB RAM) during continuous scanning for five minutes. The TensorFlow Lite inference latency was directly measured by wrapping the Interpreter.run() function with a millisecond-precision timer and overall CPU utilization was measured through Android Profiler. This approach to direct runtime latency measurement and system-level profiling allows for empirical proof of real-time performance of the framework. Over this five-minute session, 1650 inference samples were recorded, with a mean latency of 13.61 ms, a minimum of 12.27 ms, and a maximum of 51.43 ms—remaining consistently below the 100 ms real-time threshold throughout. CPU utilization was within 12% to 30%. All tests were performed on a single Android device with sequential, single-user operation. Therefore, the performance differences resulting from device-to-device variability, concurrent user scenarios, and hardware-dependent performance differences were not evaluated and are candidates for future investigation.

H. Experimental Protocol

1) Facial expression recognition evaluation

A total of 50 participants were recruited for the controlled facial expression testing, most of whom were ethnic Chinese university students and were gender-balanced. There was not a balance across demographics (age, ethnicity, cultural background) in the evaluation cohort, as recruitment was largely in one university population; therefore, the reported performance should not be interpreted as evidence of demographic fairness or generalizability across different populations.

To standardize the conditions in which the images were captured, each participant acted out 5 predefined emotional expressions under indoor light following a standardized capture prompt on screen, which advised to use good light, remove glasses where possible, and keep hair and bangs off the forehead. Instead of an individual being asked to perform an expression on command, participants were given short conversational cues in front of each capture to provoke natural expressions. Participants were seated approximately one arm's length from the device camera. 250 facial samples were taken altogether. The overall classification accuracy, per-class accuracy, and confusion matrix analysis were used to measure model performance. A cross-validation and leave-one-subject-out evaluation was not performed due to the limited number of subjects.

2) Text sentiment evaluation

The text sentiment analysis module was tested with 80 test sentences manually built, including four categories of sentences with different linguistic complexity: simple sentences, negated sentences, intensification sentences, and complex multi-clause sentences, each containing 20 sentences. Before evaluation, each sentence was classified with a ground truth emotion label. The performance was evaluated by overall classification accuracy as well as category-wise accuracy for the different levels of linguistic complexity.

3) Crisis detection evaluation

To assess the sensitivity and specificity of the crisis detection module, a test set spanning five categories was constructed: explicit ("easy") crisis-positive sentences, indirect or subtly worded ("difficult") crisis-positive sentences, explicit non-crisis ("easy negative") sentences, ambiguous non-crisis ("difficult negative") sentences, and adversarial "false-positive challenge" sentences containing risk-adjacent vocabulary in non-crisis contexts. The design consisted of 5 categories intended to test the module's ability not just to detect explicit risk language, but also to correctly withhold escalation on emotionally charged text that is not a crisis.

All participants provided informed consent prior to participation. The purpose of the study was explained to the participants, and they were given the opportunity to withdraw from the evaluation at any time. Only voluntary facial-expression capture and text-input testing under controlled conditions were undertaken, and none of the medical records or clinical assessments were collected. The study evaluated engineering performance only and was not intended as a clinical investigation.

IV. Result and Discussion

This section introduces the experimental assessment of the proposed multimodal mental health support application. Results are interpreted immediately to show the behavior of the system and practical feasibility and contribution significance.

A. Facial Emotion Recognition Performance

The Facial Expression Recognition (FER) module has been tested with 250 images of 50 subjects, and the test images were taken under controlled lighting conditions. Note that the measurement of only five classes of emotion was made, including neutral, happy, sad, angry, and surprise, as disgust and fear were not used due to the unreliable predictions of real-world emotions under mobile capture conditions.

TABLE III. Facial Emotion Recognition Accuracy (N = 250)

Emotion	Correct Detections	Incorrect Detections	Accuracy
Neutral	49	1	98.0%
Happy	46	4	92.0%
Sad	18	32	36.0%
Angry	25	25	50.0%
Surprise	41	9	82.0%
Overall	179	71	71.6%

The overall classification accuracy of the model came out to be 71.6% (179/250 correct predictions) as shown in Table III.

Class-wise performance is summarized as follows:

- Neutral: 98.0%
- Happy: 92.0%
- Surprise: 82.0%
- Angry: 50.0%
- Sad: 36.0%

The high accuracy for Neutral and Happy suggests that the CNN is well able to extract visually distinctive facial geometries such as the curvature of the smile and the relaxed muscle structure. Surprise was also well done with clear patterns of eye widening and eyebrow elevations.

However, performance for Sad and Angry was much lower. Fig. 5 Confusion matrix showed that most of Sad (81%) and Angry (72%) misclassification results were predicted as Neutral. This neutral bias has thus suggested that subtle negative expressions share overlapping low-intensity features with neutral states, in particular, in non-exaggerated real-world facial behavior.

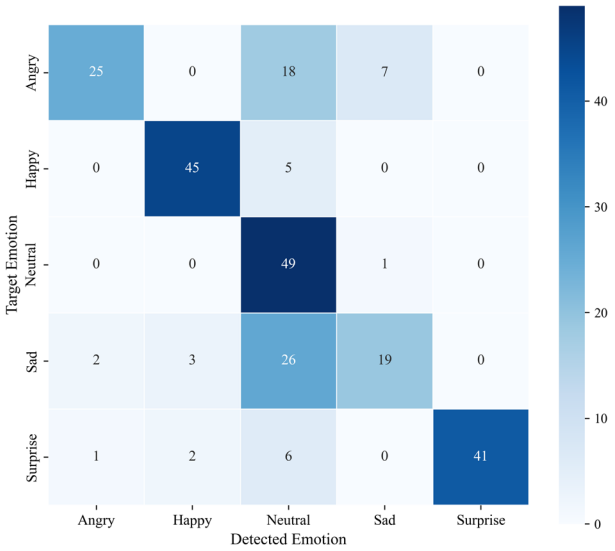


Fig. 5. Confusion matrix.

Even though it has moderate accuracy, it maintained an inference latency below 100 ms in real-time, and the CPU utilisation ranged from 12% to 20%, which is favourable for mobile deployment. To provide direct empirical evidence of this claim instead of relying on profiler-based estimates, the on-device camera interface was instrumented in such a way that every Interpreter.run() call was timed directly, and the results were logged and displayed on the screen as the "TFLite: XX ms | Avg: XX ms" overlay shown in Fig. 3. During the five-minute evaluation period, 1650 samples of inferences were recorded, with an average latency of 13.61 ms, a minimum latency of 12.27 ms, and a maximum latency of 51.43 ms. These directly measured, per-frame figures corroborate the Android Profiler results and confirm that the system consistently operates well within real-time constraints, with even the observed maximum remaining comfortably below the 100 ms threshold.

Fig. 6 shows the distribution of the confidence score of correct predictions and those made erroneously.

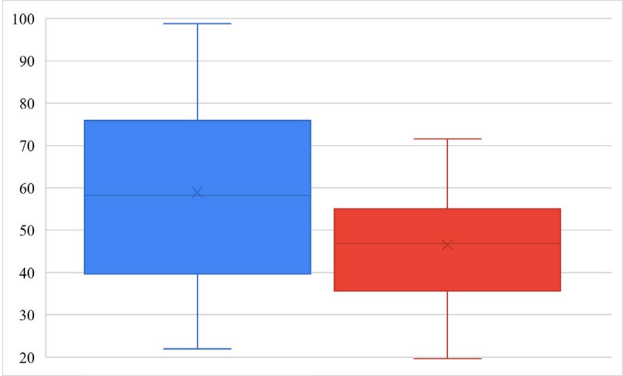


Fig. 6. Confidence score distribution by result.

The FER module is trustworthy in identifying positive and neutral states but not sensitive to negative affect, at least not in small degrees. While far from clinical, it is robust enough for real-time emotional trend monitoring for a privacy-preserving mobile environment.

B. Text Sentiment Analysis Performance

The rule-based sentiment engine was evaluated using 80 text sentences manually annotated into four linguistic complexity groups (simple, intensified, complex (multi-clause), and negated), each containing 20 text sentences, out of which each sentence was given a ground truth emotion label before the evaluation.

Table IV illustrates the overall classification accuracy was 71.3% (57/80) correct predictions.

TABLE IV. TEXT SENTIMENT PERFORMANCE ACROSS DIFFERENT LINGUISTIC COMPLEXITY CONDITIONS (N = 80)

Category	Text Samples	Correct	Accuracy
Simple	20	18	90.0%
Intensified	20	17	85.0%
Complex	20	12	60.0%
Negated	20	10	50.0%
Overall	80	57	71.3%

Performance in the linguistic categories:

- Simple statements: 90.0%
- Intensified expressions: 85.0%
- Complex contrastive sentences 60.0%
- Negated expressions: 50.0%

The engine was able to perform well on explicit and intensified expressions of emotions, showing that the weighted lexical scoring and the combination of intensity modifiers are effective on direct linguistic structures.

However, negated expressions such as "I am not happy" resulted in the lowest accuracy (50%), which suggests that there are limitations in contextual polarity reversal handling. Similarly, the need for rule-based clause weighting constraints in compound sentences with contrasting clauses decreased performance for the complex, multi-clause category.

Despite these limitations, the system works in full data offline and has little computational overhead, which guarantees privacy and feasibility at the device level.

The rule-based approach can provide for satisfactory performance with structured emotional language but has difficulty with contextual complexity. The trade-off from interpretability to contextual depth is worth it for privacy-first mobile deployment. Fig. 7 shows the accuracy by complexity.

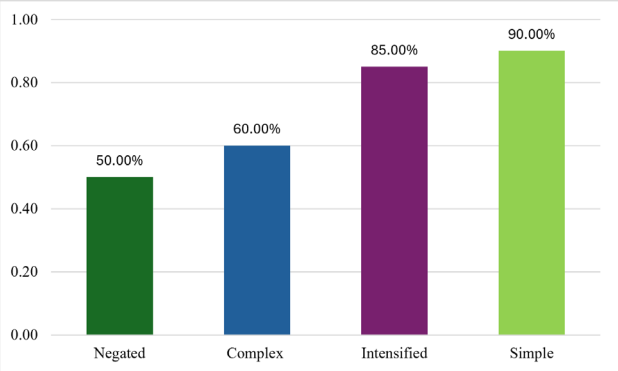


Fig. 7. Accuracy by complexity.

C. Crisis Detection Evaluation

The two-tiered crisis detection module was tested using a 70-sentence test set covering five categories to elicit both sensitivity and specificity responses. Table V illustrates the overall classification accuracy was 88.6% (62/70) correct predictions.

TABLE V. CRISIS DETECTION EVALUATION (N = 70)

Category	Text Samples	Correct	Accuracy
Easy Positive	20	20	100.0%
Difficult Positive	10	4	40.0%
Easy Negative	18	18	100.0%
Difficult Negative	12	12	100.0%
False-Positive Challenge	10	8	80.0%
Overall	70	62	88.6%

The module returned a perfect detection rate of explicit crisis expressions and a perfect rejection rate of both explicit and emotionally negative non-crisis text, demonstrating the success of the Tier-1 immediate-risk pattern set for unambiguous language and Tier-2 weighted-scoring threshold for preventing over-triggering of emotionally negative non-crisis text. Results for the adversarial false positive challenge set (80%) show that the normalization and weighting scheme is fairly, but not fully, resilient to risk-adjacent vocabulary in non-crisis settings.

The most significant result is the 40% detection rate on indirect, subtly worded crisis expressions. Since the module is explicitly pattern matched and uses a weighted keyword scoring system, expressions of distress that lack identifiable risk language elements (such as abstract hopelessness without an explicit reference to self-harm) may be more likely to be missed. This number should be interpreted as the main limitation for the current reliability of the module in a crisis detection scenario, where a false negative has a meaningful real-world consequence compared to a false positive. This discovery aligns with other precision-based crisis detection systems that have been reported, which were able to reliably identify explicit risk language but not indirect distress [31]. More general semantic modeling, such as context-based embeddings besides lexical and pattern-based signals, is still a key area for future research.

This is a conservative threshold for detection, favoring precision over recall, so as to minimize the false alarm rate and build user confidence in the detection. Given that mental health applications need to take false positives into account, and this can cause unnecessary distress to the end user, the system deliberately defaults to cautious escalation for ambiguous language while still relying on the weighted risk score to catch moderate-risk indicators. The crisis module has shown the functioning of a two-tier, auditable decision boundary in a fully offline setting, representing a substantial improvement over a flat-keyword baseline. However, there is a need for greater semantic modeling to boost the sensitivity for indirect distress language.

D. Comparative Analysis

Compared with the existing mental health applications described in Section II, the proposed system has several distinguishing contributions:

(1) Hardware Independence: Unlike EEG-based or wearable-based systems, the application does not need any external sensors.
(2) On-Device Inference: All facial and text processing is done on the local level, removing cloud dependency and privacy concerns.
(3) Dual-Modality, Alternative-Input Design: The system is a combination of FER, rule-based NLP, two-tier crisis detection, adaptive recommendations, and longitudinal tracking in a single mobile architecture.
(4) Real-Time Performance: Inference latency of less than 100 ms for a seamless user interaction.

Table VI provides a critical comparison between the proposed Feeling Diagnostic Application and 25 existing solutions reviewed in the literature.

Of this expanded comparison, two patterns should be noted. First, the recently introduced transformer-based and fusion-based systems [8, 19, 22] generally sacrifice cloud dependency or increase computational complexity or lack mobile-specific optimizations, thus recapitulating the same accuracy-deployability trade-off found in literature. Second, federated learning and differential privacy systems [28, 29] can also provide privacy levels comparable to the proposed framework, but are limited to either physiological or speech-based sensing pipelines and are not deployed as a full user-facing mobile application as the proposed system is. This puts the Feeling Diagnostic Application in a relatively uncharted territory: high privacy level, no external hardware, and a fully developed mobile application instead of a framework-level contribution.

Although not more accurate than laboratory-grade EEG systems (>90%), the proposed solution offers the best balance between accessibility, scalability, and privacy—all important aspects for real-world adoption.

The novelty of this study isn't in the best accuracy of classification, but in proving that an emotional support system that preserves privacy can be deployed entirely on consumer-level smartphones.

TABLE VI. ANALYSIS OF EXISTING SYSTEM VS PROPOSED WORK

Study	Modality	Privacy Level	Limitation
Chezer [5]	Facial	Medium	Educational setting
MOODMUSIC [6]	Facial	Medium	Entertainment only
Music Rec [7]	Facial	Medium	Limited scope
Transformer FER [8]	Facial	Medium	High computation, cloud-evaluated
Autistic App [10]	Facial	Low	Privacy risk
Depression Detection [17]	EEG + Face	High	Needs equipment
Mindfulness Therapy [15]	Multimodal	Low	Heavy computation
MANAS [16]	Text + Voice	Low	Limited conditions
EmoData [18]	Audio + Motion	Medium	Low stability
Multimodal Fusion [19]	Facial + Speech	Medium	High complexity
MediBot [20]	Text	Low	No nonverbal cues
WhatsApp Emotion [21]	Text	Medium	Language ambiguity
BERT/RoBERTa Sentiment [22]	Text	Medium	Not optimized for mobile deployment
Crisis Risk Detection [23]	Text	Medium	High-false-positive rate
MindCare [24]	Physiological	High	Requires sensors
VR Emotion [25]	Physiological	High	Hardware dependent
Wearable Stress [26]	Physiological	High	Limited scalability
Call Center Stress[27]	Physiological+Manual	Medium	Workplace only
Educator Stress FL [3]	Wearables + FL	Very high	Small population
Moblie FL Framework [28]	Mobile sensing	High	Framework-level, not deployed as full app
Speech FL+DP [29]	Speech	Very High	High computational
Food Healing [30]	Heart rate	High	Experimental
Mindset [31]	Manual	High	Low engagement
Mindful Meadows [32]	Manual + AI	High	Manual only
UI/UX Study [33]	UI survey	High	Design only
Proposed Application	Face/Text	Very High	Accuracy lower than EEG; alternative-mode; no external hardware

E. Novelty and Technical Contributions

Beyond verifying the standard functional requirements, the Feeling Diagnostic Application presents a number of architectural novelties and functional combinations that set it apart from existing solutions for mental health use in digital devices. These contributions address specifically the trade-offs of privacy, latency, and accessibility identified in the literature review.

(1) Privacy-Preserving Edge AI Architecture: As opposed to the traditional systems that call the APIs on the cloud to perform the image processing operation, this system is a completely autonomous Edge AI Inference Model. By deploying the Convolutional Neural Network (CNN) within the mobile environment itself with the TensorFlow Lite interpreter, the application does some complex biometric analysis on local hardware. This architecture provides Zero Cloud Biometric Transmission, which means that the raw facial data is only kept in volatile memory (RAM) and is never sent to external networks or third-party servers.

(2) Decoupled Hybrid Data Management Framework: The application makes use of a novel hybrid data architecture that is intended to rigidly segregate sensitive diagnostics from content delivery:

- Local Security: Personal emotional history and psychometric logs are stored within the secure sandbox, application-private environment of the device only (Local Emotion Logs), ensuring the data never leaves the device. This storage currently relies on the operating system's application sandbox for isolation rather than an additional encryption layer, a limitation identified for future work.
- Optimized Delivery: Therapeutic multimedia is consumed through on-demand external streaming, kept entirely separate from the locally stored emotional data.

This "split-state" solution helps to minimize the storage footprint of the application while ensuring that user profiling data will never intersect with content providers from the external world.

(3) Smart Closed-Loop Emotional Support Ecosystem: Unlike the current tools available that serve as disconnected devices for detection and passive apps for journaling, the proposed project creates a closed-loop recommendation system. The architecture is the taking of individual functional modules and creating a unified and automated workflow: Detect (Biometric/Text Input) > Analyze (Inference Engine) > Recommend (Curated Prescription) > Log (Historical Tracking) > Support (Therapeutic Content Consumption). This end-to-end integration is turning the application from a passive monitoring tool to an active and real-time emotional regulation system.

(4) Two-Tier Crisis Safety Protocol: In order to address the ethical risk of automated mental health applications, this system features a two-tiered decision process before performing the sentiment analysis, with a first layer for high-risk language (that has a high level of specificity) and a second layer for low-level specificity indicators of psychological distress that are combined and trigger a second major step when a threshold of cumulative weight is exceeded. This is a structured, auditable decision boundary that allows for the interpretation of the crisis-escalation logic, avoiding the interpretability of standard features, and guiding users to the crisis resources (e.g., the 988 hotline) when a crisis is

detected. The assessment process showed that this mechanism worked well with explicit distress language (100%) and non-crisis rejection responses (100%), with low sensitivity for non-explicit expressions of distress (40%) identified as an area needing further development.

F. Limitations

Several limitations were identified:

(1) Neutral bias in the FER model makes one less sensitive to subtle negative emotions;
(2) Rule-based NLP has trouble with negation, sarcasm and changing slang;
(3) Crisis detection sensitivity to indirect distress is low, a meaningful safety gap given the cost of false negatives;
(4) The FER model uses FP32 precision with no quantization or pruning, increasing model size;
(5) Emotional records are not encrypted at rest as they are stored locally;
(6) Testing was conducted under a single device, and single-user configuration;
(7) The 50-participant cohort was recruited from a single university population and was not demographically stratified; results should not be read as evidence of fairness or generalizability;
(8) There was no cross validation or leave-one-subject-out evaluation performed;
(9) Facial testing was only carried out under controlled indoor lighting.

Future improvements may include:

- Lightweight, compressed transformer models on-device (e.g., MobileBERT, DistilBERT) for better negation and indirect distress understanding
- Post-training quantization to reduce model size
- Encrypted local storage
- Adaptive confidence calibration
- Expanded crisis lexicon
- Cross-validation with larger test sets of more diverse demographics
- Multi-device performance benchmarking

V. Conclusion

This study presents a privacy-preserving, dual-modality mobile framework for real-time emotional self-monitoring and support, proposed and validated as the Feeling Diagnostic Application (FDA). In response to the research question introduced in the Introduction, the findings show that a mobile framework that integrates facial emotion classification and text-based sentiment analysis can be realized as a fully on-device system, providing emotional support in real time, in private and personalized mode, without sending biometric data to external servers. The study achieved its three primary objectives, demonstrating the feasibility of fully on-device mental health computing.

The first goal, creating a privacy-first Android application with Edge AI, has been accomplished using a completely autonomous inference architecture. The custom CNN is deployed via TensorFlow Lite, which means that all biometric analysis is done in the device's local memory, avoiding any dependency on the cloud and external transmission of facial data. Direct runtime instrumentation showed a mean inference latency of 13.61 ms, which is much lower than the 100 ms real-time requirement, ensuring that the on-device processing does not compromise the responsive experience.

The second goal, an emotion-based recommendation engine, was realized through a closed-loop adaptive recommendation module that maps recognized emotional states to curated therapeutic multimedia, alongside a two-tier crisis detection mechanism achieving 88.6% overall accuracy on a 70-sentence evaluation set. This module was successful in its performance of explicit crisis language (100%), but less sensitive to indirect modes of distress (40%), as discussed below. Consistent with its role as a support and signposting tool rather than a diagnostic authority, the system does not replace professional clinical assessment or emergency response.

The third goal, a dual modality support system, was realized by creating two alternative interaction modes, automated facial expression recognition and rule-based text sentiment analysis, which were selected by the user. This design provides users with the flexibility to participate passively through scanning or actively through self-report as per context and preference. 71.6% accuracy for facial emotion recognition ($n = 250$) and 71.3% for text sentiment classification ($n = 80$) were obtained in the experiments, demonstrating that providing multiple lightweight modalities is a viable approach to enhance interaction flexibility and accessibility, but without testing whether multiple modality inputs combine to improve recognition accuracy compared to using either modality alone.

These results are limited by several points. The facial recognition model has a bias towards "neutral" predictions for subtle negative emotions, while the rule-based text engine is less effective when it comes to negation and multi-clause sentence structures. This is also pertinent with respect to the reduced sensitivity of the crisis detection module to indirect distress language, as this is a very important safety consideration for future clinical use. The evaluation cohort was not split by demographic, and locally stored data is currently not encrypted at rest, both of which were priorities for future work and not evidence of the robustness of the present day. In addition, the framework was assessed only from an engineering viewpoint and not with a standardized psychological assessment tool or with the expert clinician's evaluation. Future studies include examining lightweight transformer-based language models (e.g., MobileBERT) for efficient on-device language understanding, analyzing the model across a wider demographic to increase model accuracy, and evaluating privacy-preserving collaborative training methods like federated learning to update the model without violating data sovereignty.

Overall, this study lays out a scalable framework for affective computing at the edge, which is not meant to replace clinical, professional evaluation but rather serve as a viable approach to ethical, accessible, and privacy-preserving digital mental health support.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

D.S.T. conducted the research, analyzed the data, and wrote the manuscript. L.S.L. conceptualized and supervised the study and reviewed the manuscript. J.L.T. contributed to the conceptualization of the study and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENT

The authors would like to thank the participants involved in the experimental evaluation.

REFERENCES

[1] P. Agarwal *et al.*, "Navigating the Mind: Analysis of Stress and Mental Well-being," in *Proc. 2024 International Conference on Innovations and Challenges in Emerging Technologies,* 2024. doi: 10.1109/ICICET59348.2024.10616359

[2] J. Koh, G. Y. Q. Tng, and A. Hartanto, "Potential and Pitfalls of Mobile Mental Health Apps in Traditional Treatment: An Umbrella Review," *Journal of Personalized Medicine*, 2022. doi: 10.3390/jpm12091376

[3] D. S. Devi, M. Sujithra, P. Velvadivu, V. V. Kavya, and S. Jhunandhini, "Predictive Stress Management for Working Women in Educational Institutions using Federated Learning and Wearable Technology," in *Proc. 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, 2024, pp. 1554–1559. doi: 10.1109/I-SMAC61858.2024.10714822

[4] M. Yatheen and D. Ishwara, "Financial vulnerability among scheduled castes and scheduled tribes in India," *EPRA International Journal of Economic and Business Review (JEBR)*, vol. 12 , issue 6, 2024. doi: 10.36713/epra17399

[5] S. Suhan, K. Kalaichelvan, L. Samarage, D. Alahakoon, P. Samarasinghe, and M. Nadeeshani, "Automated Evaluation of Child Emotion Expression and Recognition Abilities," in *Proc. 2022 International Conference on Information Technology Systems and Innovation*, 2022, pp. 388–393. doi: 10.1109/ICITSI56531.2022.9970990

[6] B. Chaudhary, N. Singh, A. Arora, and S. Sengupta, "MOODMUSIC: An Android App for Mood-based Music Player using Deep Facial Expression Detection," in *Proc. the 3rd International Conference on Applied Artificial Intelligence and Computing*, 2024, pp. 818–822. doi: 10.1109/ICAAIC60222.2024.10575740

[7] A. V. Gadagkar, S. Santhosh, S. Begum, and M. Ashwin Shenoy, "Emotion Recognition and Music Recommendation System based on Facial Expression," in *Proc.2nd IEEE International Conference on Advances in Information Technology*, 2024. doi: 10.1109/ICAIT61638.2024.10690441

[8] X. Zhang, M. Li, S. Lin, H. Xu, and G. Xiao, "Transformer-based multimodal emotional perception for dynamic facial expression recognition in the wild," *IEEE Trans. Circuits Syst. Video Technol*., vol. 34, no. 5, pp. 3192–3203, 2024.

[9] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Trans. Affective Comput*., 2017.

[10] A. Al-Qahtani, F. Al-Mahawes, F. Al-Qahtani, L. Berriche, and S. L. M. Sainte, "Cloud-Based Emotion Recognition Application for Arabic Autistic Kids," in *Proc. 2025 8th International Conference on Data Science and Machine Learning Applications,* 2025, pp. 174–179. doi: 10.1109/CDMA61895.2025.00035

[11] M. M. Hosseini, A. Pourramezan Fard, and M. H. Mahoor, "Faces of fairness: Examining bias in facial expression recognition datasets and models," arXiv preprint, arXiv:2502.11049, 2025.

[12] P. P. Ray, "A review on TinyML: State-of-the-art and prospects," *J. King Saud Univ.–Comput. Inf. Sci*., vol. 34, no. 4, pp. 1595–1623, 2022.

[13] R. David *et al*., "TensorFlow Lite Micro: Embedded machine learning for TinyML systems," *Proc. Mach. Learn. Syst. (MLSys)*, vol. 3, pp. 800–811, 2021.

[14] Z. Li, H. Li, and L. Meng, "Model compression for deep neural networks: A survey," *Computers*, vol. 12, no. 3, 60, 2023.

[15] C. D. P. Del Rosario, L. M. D. Fernandez, V. A. Agustin, C. J. Centeno, M. B. Mangrobang, and D. M. Cortez, "AI-Powered Intelligent Mindfulness-Based Cognitive Therapy Companion for Emotion Recognition and Context-Aware Therapy," in *Proc. 8th International Symposium on Innovative Approaches in Smart Technologie*, 2024. doi: 10.1109/ISAS64331.2024.10845635

[16] G. Akash, S. Gopinath, S. M. M. Zachariah, and S. Mithuna, "Personalized Recommendations System for Mental Health Support using NLP," in *Proc. 3rd International Conference on Electronics and Renewable Systems*, 2025, pp. 1055–1062. doi: 10.1109/ICEARS64219.2025.10940314

[17] G. Tiwary, S. Chauhan, and K. K. Goyal, "Automatic Depression Detection Using Multi-Modal & Late-Fusion Based Architecture," in *Proc. 7th IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions*, 2023. doi: 10.1109/CSITSS60515.2023.10334084

[18] M. L. Barron-Estrada, R. Zatarain-Cabada, and C. G. Aispuro-Gallegos, "Multimodal recognition of emotions with application to mobile learning," in *Proc. IEEE 18th International Conference on Advanced Learning Technologies*, 2018, pp. 416–418. doi: 10.1109/ICALT.2018.00104

[19] Z. Dai, H. Fei, and C. Lian, "Multimodal information fusion method in emotion recognition in the background of artificial intelligence," *Intelligent Learning Environments*, Mar. 2024. doi: 10.1002/itl2.520

[20] N. Kavyashree and J. Usha, "MediBot: Healthcare Assistant on Mental Health and Well Being," in *Proc. 7th IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions*, 2023. doi: 10.1109/CSITSS60515.2023.10334083

[21] R. Padate, A. Gupta, P. Chakrabarti, and A. Sharma, "Emotion Recognition from WhatsApp Text Messages Using Unsupervised Machine Learning," in *Proc. 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud*, 2024, pp. 1871–1876. doi: 10.1109/I-SMAC61858.2024.10714817

[22] M. I. Fahmi and A. A. Nababan, "A comparative study of BERT and RoBERTa for sentiment analysis on Twitter data related to mental health," *Math. Model. Eng. Probl*., vol. 12, no. 9, pp. 3289–3295, 2025.

[23] M. Broadbent, M. Medina Grespan, K. Axford, X. Zhang, V. Srikumar, B. Kious, and Z. Imel, "A machine learning approach to identifying suicide risk among text-based crisis counseling encounters," *Front. Psychiatry*, vol. 14, 1110527, 2023. doi: 10.3389/fpsyt.2023.1110527

[24] P. Ushashree and A. Naik, "MindCare: Machine Learning based Personalized Mental Health Monitor," in *Proc. 4th International Conference on Sustainable Expert System*, 2024, pp. 1528–1533. doi: 10.1109/ICSES63445.2024.10763122

[25] A. F. Bulagang, J. Mountstephens, and J. Teo, "Investigating Electrodermography (EDG) and Heart Rate (HR) Signals for Emotion Classification in Virtual Reality," in *Proc. 1st International Visualization, Informatics and Technology Conference*, 2022, pp. 29–34. doi: 10.1109/IVIT55443.2022.10033366

[26] L. G. A. Raymondi, J. Armas-Aguirre, F. E. A. Guzmán, and P. A. González, "Technological solution for the identification and reduction of stress level using wearables," in *Proc. 2020 15th Iberian Conference on Information Systems and Technologies*, 2020. doi: 10.23919/CISTI49556.2020.9140971

[27] K. Lee, H. Lim, S. Ahn, T. Kim, and H. Hong, "Designing a Personalized Stress Management System for Call Center Workers," in *Proc. 2023 IEEE International Conference on Big Data and Smart Computing*, 2023, pp. 383–385. doi: 10.1109/BigComp57234.2023.00092

[28] B. Suruliraj and R. Orji, "Federated learning framework for mobile sensing apps in mental health," in *Proc. IEEE 10th Int. Conf. Serious Games Appl. Health (SeGAH)*, 2022, pp. 1–7.

[29] Y. Cui, Z. Li, L. Liu, J. Zhang, and J. Liu, "Privacy-preserving speech-based depression diagnosis via federated learning," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2022, pp. 1371–1374.

[30] N. J. Antonio, Y. R. Nobel, H. C. Ho, and W. Astuti, "Emotion Recognition based on Heart Rate Detection using Logistic Regression for Automation Food Healing Application," in *Proc. International Conference on Education and Technology*, 2022, pp. 118–122. doi: 10.1109/ICET56879.2022.9990056

[31] M. Samuel and C. P. Shirley, "Mindset, An Android-Based Mental Wellbeing Support Mobile Application," in *Proc. 2023 3rd International Conference on Pervasive Computing and Social Networking*, 2023, pp. 989–996. doi: 10.1109/ICPCSN58827.2023.00169

[32] A. Gaikwad, G. Nimbolkar, R. Keswani, V. Kolhe, and G. S. Navale, "Mindful Meadows: A Mental Health App," in *Proc. 2024 8th International Conference on Computing, Communication, Control and Automation*, 2024. doi: 10.1109/ICCUBEA61740.2024.10775202

[33] S. Goyal, J. Sharma, K. Sharma, and A. Kumar, "Impact of UI/UX on usage of mobile apps for remote psychological health monitoring," in *Proc. 2023 14th International Conference on Computing Communication and Networking Technologies*, 2023. doi: 10.1109/ICCCNT56998.2023.10307320

[34] P. Matthews and C. Rhodes-Maquaire, "Personalisation and recommendation for mental health apps: A scoping review," *Behaviour & Information Technology*, vol. 44, no. 10, pp. 2389–2404, 2025. doi: 10.1080/0144929X.2024.2356630

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).