Affective and Pleasurable Design, Vol. 204, 2026, 228–235
https://doi.org/10.54941/ahfe1007348

An Intelligent Smart Mirror Framework for Real-Time Emotion Awareness and LLM-Based Reflective Guidance

Reema Albarrak, Waleed Al-Nuwaiser, Norah Bin Ajlan, and Lamees Alghanem

Computer Science Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, 11623, Saudi Arabia

ABSTRACT

Recognizing and interpreting emotional states in real time remains a challenge for individuals seeking to improve psychological resilience and emotional self-awareness. This paper presents Istibtan, a human-centered mobile framework designed to support reflective emotional awareness through real-time facial expression recognition (FER) and personalized feedback. The system employs a Convolutional Neural Network (CNN) trained on the FER-2013 dataset, augmented with facial landmark detection to enhance robustness under varying environmental conditions. Beyond emotion classification, Istibtan distinguishes itself through the integration of large language models (Gemini LLMs), which transform raw affective data into context-aware, psychology-informed reflective guidance, supporting meaningful user interpretation rather than passive monitoring. Developed using the Flutter SDK with secure cloud-based data management, the application further supports sustained emotional engage ment through digital journaling and longitudinal emotion tracking modules. Evaluation on external validation datasets achieved a classification accuracy of 76.86%, indicating reliable generalization performance. By combin ing affective computing, reflective interaction design, and AI-driven personalization, this work contributes a novel mobile framework for enhancing emotional intelligence and proactive mental well-being, with clear implications for human-centered mental health technologies.

Keywords: Exemplary paper, Human systems integration, Systems engineering, Systems modeling language

INTRODUCTION

Background

Emotional self-awareness serves as a fundamental pillar of psychological resilience and behavioral regulation. Foundational research in psychology posits that the capacity to recognize and interpret one's own affective states is a prerequisite for emotional intelligence, directly influencing decision-making and social adaptation (Jordan and Ashkanasy, 2013). Furthermore, structured reflection techniques, such as expressive writing, have been shown to significantly improve mental health outcomes by reducing cognitive load (Pennebaker and Chung, 2011). With the global proliferation of mobile technology, smartphones have transformed into vital platforms for delivering these health interventions,

Received February 5, 2026; Revised March 29, 2026; Accepted April 18, 2026; Available online July 20, 2026

transitioning care from clinical settings to everyday life (Donker et al., 2013). Meta-analyses indicate that smartphone-based interventions can be effective in reducing symptoms of anxiety and depression, particularly when designed with high user engagement features (Firth et al., 2017).

Technological Landscape

Parallel to these psychological insights, the field of Affective Computing has witnessed a paradigm shift driven by Deep Learning. Convolutional Neural Networks (CNNs) have largely replaced traditional handcrafted feature extraction methods, demonstrating superior performance in Facial Expression Recognition (FER) tasks (Zhang, 2017). Recent surveys confirm that deep learning models can effectively classify core emotional states—such as happiness, sadness, and anger—even under challenging environmental conditions (Li and Deng, 2020). The development of large-scale annotated datasets, such as FER-2013 (Goodfellow et al., 2013) and AffectNet (Mol lahosseini et al., 2017), has been instrumental in training these models to generalize across diverse user demo graphics. To ensure robustness in real-world mobile deployment, modern frameworks increasingly rely on facial land mark detection. Algorithms utilizing libraries like Dlib and OpenCV allow for precise geometric alignment of facial features (e.g., eyes and mouth regions), which is critical for maintaining accuracy when the user is not di rectly facing the camera (King, 2009; Zhang et al., 2014). Recent comparative studies suggest that aligning faces based on these landmarks significantly reduces error rates caused by head pose variations (Wu and Ji, 2019).

The Research Gap

Despite these technical advancements, current digital mental health tools face a clear limitation: they often workseparately rather than together. Popular mobile applications like Daylio, Reflectly, and Youper depend mainly on manual self-reporting (Daylio, 2025; Reflectly, 2025; Youper Inc., 2025). Although these apps help users build habits, they rely on memory and cannot detect real-time or unconscious emotional changes. In contrast, specialized systems like Affectiva and OpenFace offer strong objective analysis but lack the interactive support needed for user well-being (McDuff et al., 2013; Baltruˇ saitis et al., 2016). Additionally, while some smart mirrors and conversational agents exist (Alboaneen et al., 2020; Sun et al., 2018), they often act only as data displays and lack deep, context-aware guidance (Vaidyam et al., 2019).

Proposed Framework

To bridge this digital divide, this paper introduces Istibtan, a unified framework that synergizes objective computer vision with generative psychological guidance. The proposed solution integrates a lightweight CNN optimized for mobile devices (Minaee et al., 2021) with the Gemini 2.5 Flash Large Language Model (Comanici et al., 2025). By leveraging the speed and reasoning capabilities of Gemini 2.5 Flash, Istibtan transforms raw physiological data into meaningful, context-aware dialogue in real-time. This approach aligns with the principles of adaptive interaction, aiming to provide a holistic tool for emotional regulation and self-awareness.

METHODOLOGY

Proposed Framework The Istibtan system is built upon a Layered Client-Server Architecture, designed to ensure modularity, scal ability, and the clean separation of responsibilities between the mobile interface and the core machine learning logic. As illustrated in Figure 1, the framework consists of five distinct layers that work sequentially to process user input.

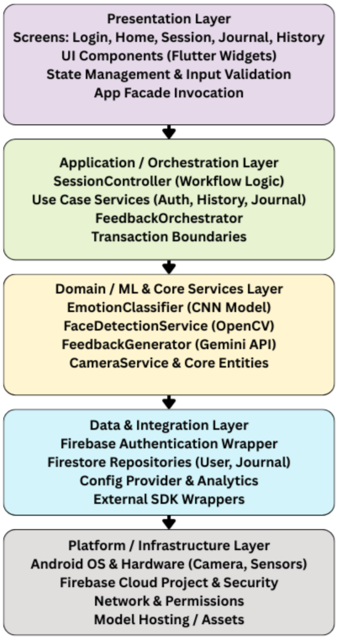

Figure 1: Layered architecture diagram for the istibtan system.

The figure demonstrates the system's modular decomposition, starting from the **Presentation Layer**, which handles the Flutter-based UI and user interactions. This layer communicates with the **Application Layer**, which acts as an orchestrator using controllers like the SessionController to manage workflows. The core intelligence resides in the **Domain/ML Layer**, containing the FaceDetectionService and EmotionClassifier. Data persistence is managed by the **Data Layer** through Firebase integrations, while the **Platform Layer** handles low-level hard ware access such as camera permissions . This structure ensures that the emotion recognition pipeline operates efficiently on mobile devices.

Data Collection and Model Fine-Tuning

To build a robust emotion recognition engine, the project utilized a multi-stage data strategy. The Convolu tional Neural Network (CNN) was initially trained on the FER-2013 dataset, which serves as the primary bench mark for facial expression recognition tasks (Goodfellow et al., 2013). However, to address common limitations regarding lighting variations and head pose angles, the model un derwent a fine-tuning process using supplementary datasets, specifically RAF-DB and FER+ (Li et al., 2017; Barsoum et al., 2016). This extended training allowed the model to better generalize to real-world camera inputs and accommodated variations in expression intensity. Finally, the model's performance was validated against the Extended Cohn-Kanade (CK+) dataset, which was explicitly excluded from the training phase to ensure unbiased evaluation. The fine-tuned model achieved an accuracy of 76.86% on this validation set (Lucey et al., 2010)

Model Architecture

The core classifier is a custom lightweight Convolutional Neural Network (CNN) optimized for mobile de ployment. The architecture consists of four convolutional blocks. Each block comprises a Conv2D layer for feature extraction, followed by BatchNormalization to stabilize learning, and MaxPooling2D for dimen sionality reduction. To prevent overfitting, Dropout layers (rate = 0.5) were integrated after the fully connected dense layers. This specific configuration balances computational efficiency with high accuracy, achieving 76.86% on the CK+ validation set (Minaee et al., 2021).

LLM-Driven Feedback Module

Beyond classification, Istibtan integrates the Gemini 2.5 Flash Large Language Model (LLM) to provide actionable feedback (Comanici et al., 2025). The system employs a dual-prompt strategy to handle both immediate emotional reactions and long-term pattern analysis. As detailed in Table 1, specific prompt engineering techniques were implemented to ensure the AI acts as a "gentle, supportive companion" with a non-clinical, aesthetic tone.

Table 1: Prompt engineering strategy for gemini 2.5 flash.

Component	Scenario A: Real-Time Feedback	Scenario B: Historical Analysis
Context Source	Single Image Capture (Current State)	Longitudinal Data (History & Current)
Input Data	Detected Emotion Label (e.g., Sad)	• Raw History Logs (JSON) • Most Top 3 Emotions
System Persona	"You are a gentle, supportive companion with a soft and aesthetic tone."	
Core Task	Write exactly 3 short sentences: 1. Warm validation of the emotion. 2. A soft, aesthetically comforting remark. 3. A simple, safe coping suggestion	Write exactly 3 short sentences: 1. Gentle description of the visible pattern 2. Soft encouraging reflection (non-intrusive). 3. Simple and uplifting suggestion.
Safety Guardrails	"Keep the style gentle, non-clinical, and emo-tonally supportive."	"Do not interpret the user's personal life. Keep the tone warm and non-clinical."

RESULTS

Model Performance Evaluation Toensure the reliability of the emotion recognition engine, the fine-tuned CNN model was rigorously evaluated against the CK+ dataset, which serves as a standard benchmark for facial expression analysis. This dataset was strictly excluded from the training phase to provide an unbiased assessment of the model's generalization capabilities. The evaluation results demonstrated that the proposed model achieved a classification accuracy of 76.86% on the validation set. This performance indicates a strong ability to generalize beyond the training data (FER 2013 and RAF-DB), successfully handling variations in facial structures and expression intensities (Minaee et al., 2021). The integration of Dlib's 68-point facial landmarks further enhanced the system's robustness, particularly in stabilizing predictions under varying lighting conditions (Zhang et al., 2014).

User Acceptance Testing

Aquantitative user study was conducted with 60 participants to assess the perceived usefulness and impact of Istibtan. The participants were surveyed after interacting with the system's core features (Reflection, Journaling, and Feedback). The statistical analysis of the responses revealed high user acceptance: • 73.3% of users agreed that the automated emotion detection was a useful feature for self-awareness. • 80%indicated that personalized feedback was a critical factor in their engagement. • 70% believed that regular use of the system could positively impact their mental well-being and stress management (Firth et al., 2017).

Comparative System Analysis

To contextualize the contributions of the Istibtan framework, a comparative analysis was performed against existing state-of-the-art systems reviewed in the literature. Table 2 highlights the functional gaps in current so lutions and demonstrates how the proposed system bridges the divide between objective detection and subjective reflection.

Table 2: Comparison of istibtan with existing solutions.

System	Auto-Detection	Generative Feedback	Journaling	History Tracking	Platform
Affectiva (2013)	Yes (CNN)	No	No	No	Mobile/SDK
OpenFace 2.0 (2016)	Yes (DNN)	No	No	No	Desktop
Reflectly (2025)	No (Manual)	No (Pre-set)	Yes	Yes	Mobile
Youper (2025)	No (Chatbot)	Yes (Scripted)	Yes	Yes	Mobile
Smart Mirror (2020)	Yes (Vision)	Limited	No	No	Hardware
Istibtan (Ours)	**Yes (CNN+Dlib)**	**Yes (LLM-Based)**	**Yes**	**Yes**	**Mobile**

As shown in Table 2, while systems like Affectiva and OpenFace excel in detection accuracy, they lack the reflective components necessary for mental health support. Conversely, apps like Reflectly and Youper offer jour naling but rely entirely on manual user input, missing the objective physiological data. Istibtan uniquely inte grates all four dimensions—automated detection, generative psychological guidance, journaling, and historical tracking—into a single unified mobile platform.

DISCUSSION

Interpretation of Findings

The evaluation confirms the efficacy of Istibtan, particularly regarding the synergy between the CNN model and the Generative Feedback module. Achieving 76.86% accuracy on the CK+ dataset is competitive for a lightweight mobile model, validating the decision to use transfer learning with RAF-DB. Furthermore, the high user acceptance rate (80% for personalized feedback) supports the hypothesis that LLM-driven guidance signifi cantly enhances user engagement compared to static emotional tracking methods found in similar literature.

Implications for Mental Health Tech

This study extends existing research by demonstrating that mobile applications can move beyond passive monitoring. By bridging the gap between objective physiological data (facial expressions) and subjective reflection (journaling), Istibtan offers a more holistic approach to digital mental health. This aligns with the "Human-in-the Loop" design philosophy, ensuring that AI acts as a facilitator for self-awareness rather than a replacement for human judgment.

Limitations and Future Work

Despite the promising results, this study has limitations. First, the evaluation was conducted in a controlled environment; longitudinal studies are needed to assess the long-term impact on psychological resilience. Second, the current model relies solely on visual data; future iterations will aim to incorporate multimodal analysis (e.g., voice tone and text sentiment) to improve emotion detection accuracy in complex scenarios. Finally, expanding the dataset to include more diverse demographics will further ensure the system's inclusivity and robustness.

CONCLUSION

This study introduces **Istibtan**, a validated design framework for intelligent emotion awareness. By integrating real-time computer vision with generative AI, the system provides a novel approach to reflective mental health support. The evaluation confirmed the system's technical viability and user acceptance. Future work will focus on enhancing the multimodal capabilities and conducting long-term field studies to validate the clinical efficacy of the proposed solution.

REFERENCES

Alboaneen, D. A., Alsaffar, D., Alateeq, A., Alqahtani, A., Alfahhad, A., Alqahtani, B., Alamri, R., and Alamri, L. (2020). Internet of things based smart mirrors: A literature review. In 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS), pages 1–6. IEEE.

Baltrušaitis, T., Robinson, P., and Morency, L.-P. (2016). Openface: an open source facial behavior analysis toolkit. In 2016 IEEE winter conference on applications of computer vision (WACV), pages 1–10. IEEE.

Barsoum, E., Zhang, C., Ferrer, C. C., and Zhang, Z. (2016). Training deep networks for facial expression recognition with crowd-sourced label distribution. In Proceedings of the 18th ACM international conference on multimodal interaction, pages 279–283.

Comanici, G., Bieber, E., Schaekermann, M., Pasupat, I., Sachdeva, N., Dhillon, I., Blistein, M., Ram, O., Zhang, D., Rosen, E., et al. (2025). Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. arXiv preprint arXiv:2507.06261.

Daylio (2025). Daylio: Journal, diary and mood tracker. Mobile Application, Available at: https://daylio.net/. Accessed: 2025.

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M.-R., and Christensen, H. (2013). Smartphones for smarter delivery of mental health programs: a systematic review. Journal of medical Internet research, 15(11):e2791.

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., and Sarris, J. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World psychiatry, 16(3):287–298.

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., et al. (2013). Challenges in representation learning: A report on three machine learning contests. In International conference on neural information processing, pages 117–124. Springer.

Jordan, P. J. and Ashkanasy, N. M. (2013). Emotional intelligence, emotional self-awareness, and team effectiveness. In Linking emotional intelligence and performance at work, pages 145–164. Psychology Press.

King, D. E. (2009). Dlib-ml: A machine learning toolkit. The Journal of Machine Learning Research, 10:1755–1758.

Li, S., Deng, W., and Du, J. (2017). Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2852–2861.

Li, S. and Deng, W. (2020). Deep facial expression recognition: A survey. IEEE transactions on affective computing, 13(3):1195–1215.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In 2010 IEEE computer society conference on computer vision and pattern recognition-workshops, pages 94–101. IEEE.

McDuff, D., Kaliouby, R., Senechal, T., Amr, M., Cohn, J., and Picard, R. (2013). Affectiva-mit facial expression dataset (am-fed): Naturalistic and spontaneous facial expressions collected. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pages 881–888.

Minaee, S., Minaei, M., and Abdolrashidi, A. (2021). Deep-emotion: Facial expression recognition using attentional convolutional network. Sensors, 21(9):3046.

Mollahosseini, A., Hasani, B., and Mahoor, M. H. (2017). Affectnet: A database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing, 10(1):18–31.

Pennebaker, J. W. and Chung, C. K. (2011). Expressive writing: Connections to physical and mental health. Oxford University Press.

Reflectly (2025). Reflectly: World's first intelligent journal. Mobile Application, Available at: https://reflectly.app/. Accessed: 2025.

Sun, Y., Geng, L., and Dan, K. (2018). Design of smart mirror based on raspberry pi. In 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), pages 77–80. IEEE.

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., and Torous, J. B. (2019). Chatbots and conversational agents in mental health: a review of the psychiatric landscape. The Canadian Journal of Psychiatry, 64(7):456–464.

Wu, Y. and Ji, Q. (2019). Facial landmark detection: A literature survey. International Journal of Computer Vision, 127(2):115–142.

Youper Inc. (2025). Youper: Artificial intelligence for mental health. Mobile Application, Available at: https://www.youper.ai/. Accessed: 2025.

Zhang, T. (2017). Facial expression recognition based on deep learning: a survey. In International conference on intelligent and interactive systems and applications, pages 345–352. Springer.

Zhang, Z., Luo, P., Loy, C. C., and Tang, X. (2014). Facial landmark detection by deep multi-task learning. In European conference on computer vision, pages 94–108. Springer.