# 242v2 242v2

## 242v2

Quick Submit

Quick Submit

IILM University, Greater Noida

### Document Details

Submission ID

trn:oid:::1:3457865869

Submission Date

Jan 15, 2026, 10:55 AM GMT+5:30

Download Date

Jan 15, 2026, 11:07 AM GMT+5:30

File Name

242v2.pdf

File Size

319.5 KB

6 Pages

4,230 Words

25,150 Characters

# 54% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

**Caution: Review required.**

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

## Detection Groups

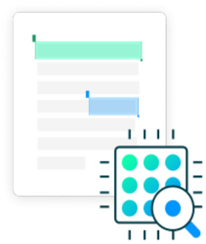
**25** AI-generated only 54%
Likely AI-generated text from a large-language model.

**0** AI-generated text that was AI-paraphrased 0%
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

**Disclaimer**
Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

## Frequently Asked Questions

**How should I interpret Turnitin's AI writing percentage and false positives?**
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.



False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

**What does 'qualifying text' mean?**
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

# Smart Mirror for Real-Time Mental Monitoring Using Lightweight CNN and Edge AI

K. Sai Sundhar Veekshith
*Student*
*Sathyabama Institute of Science and Technology*
*Chennai, India*
saisundhar1947@gmail.com

E. Ranjith Kumar Goud
*Student*
*Sathyabama Institute of Science and Technology*
*Chennai, India*
ranjithkumargoud52@gmail.com

Mrs. D. Saral Jeeva Jothi
*Assistant Professor*
*Sathyabama Institute of Science and Technology*
*Chennai, India*
saraljeevajothi.d.cse@sathyabama.ac.in

***Abstract*—Facial emotion recognition is an important competent part in comprehending an individual's mental state. It facilitates human-computer interaction. However, many Current Face++ systems for facial emotion recognition have a cloud-based processing, which in turn leads to increased latency, privacy issues, and Issues with Reliability. The paper will introduce an Emotion-Aware Smart A mirror utilizing edge computing with a lightweight Con- Evolutional Neural Network is used in real time facial emotion analysis recognition. The proposed system is based on embedded systems. including Raspberry Pi and ESP32-CAM, and is trained on FER 2013 dataset with seven emotions: Happy, Sad, Angry Neutral, Surprise, Fear, and Disgust. Preprocessing methods such as conversion to grayscale, normalization, and Harr Cascade- Faces detected using LBP are used for improved classification. Accuracy - Based on the detected emotion, the system generates individualized feedback, motivational quotes, and calming messages recommendations, and health tips. Some experimental results show in order for this proposed system to have reliable accuracy with low inference time, making it ideal for real-time analysis of mental state Such is an example of non-invasive tracking. The technique is privacy-respecting, cost-effective, and suitable for deployment in everyday environments.**

***Index Terms*—Facial Emotion Recognition, Edge AI, Smart Mirror, Convolution-al Neural Network, FER-2013, Mental Health Monitoring**

## I. Introduction

"The human well being of mental beings is a principal point of": This is a literal translation of the given French sentence arising from the total welfare of human beings, it stipulates the extent of emotional stability, health, or the extent productivity, decision-making capability, or social behavior. This scalding urbanization, on-line overworking pressure, work uncertainties, and change of lifestyles over the recent years that have played an important role in improving the number of stress, anxiety, and depressive cases among the persons of any age [11]. However, despite the fact that it has gained more significance, the importance of mental health remains neglected by the force of social stigma, ignorance, lack of direct observation systems [13].

Facial Expression Recognition in affective computing is viewed as one of the most promising areas of research or non-invasive manner of emotion analysis 3, 15. Innately, by using the facial expression of a human being, it reveals internal feelings for happiness, sadness, anger, surprise, fear, and disgust. According to studies in psychological aspects, such expressions occur through cultures and can be mentioned in other cultures to a large extent. Because of this universality, therefore, facial expressions have become a great medium for emotion detection.

Because of the rapid development of computer vision and artificial intelligence, computers can learn about complicated patterns of the face and apply them to emotions. Precisely, the CNNs results have been outstanding in terms of mining a face image in terms of hierarchical detail features and mining features of the face image through emotion categorization tasks. For the first time, improvements enabled by FER-2013, JAFFE, and AffectNet databases open quite new possibilities in the given area-as now it became possible to train deep learning models which may be called in order to mimic human-level performance in controlled environments.

## II. Related Work

Facial Emotion Recognition (FER): The facial emotion recognition technique in computer vision/Artificial Intelligence/Affective Computing is a booming research area and a very dynamically developing field in face recognition [3], [15]. A considerable amount of debate has emerged in the past decade on the proposals put forth concerning an improvement in the efficiency of emotion recognition systems and their implementation for practical tasks [2], [9]. Such proposals were drawn up based on existing machine learning algorithms with manual/Handcrafted Features in machine learning to deep architectures with prospects of learning face patterns by themselves [3], [10].

"The most prominent and efficient learning model used by Gudi et al. in identifying face expressions, and their relevance to pixels, is based on a technique called a Convolutional Neural Network," explained Er et al. Moreover, deep learning neural networks were highlighted by Gudi et al. in their study to be able to represent relevant information in a face image without relying on a human-generated definition, such as Histogram of Oriented Gradients or Local Binary Patterns [4]. It is one of the modifications in conventional methods in learning models for face expressions up to end-to-end learning models in FER systems in the deep learning revolution [10].

To overcome this challenge of real time performance, Jadhav et al. suggested a real time framework for the CNN structure of facial emotion recognition in real time using FER-2013 datasets [5]. They proposed a system with a simpler structure by reducing the parameters with or without a significantly negative impact on this competitive level of accuracy. They highlighted the critical role of requiring a good trade-off in accuracy and efficiency in light of this particular situation where the performance of the real time emotion analysis is relevant, such as in a human computer interface system or a class room monitoring system being used in health evaluation [9]. They were completely reliant on computers with good processing capacity.

However, the implementation of the emotion recognition model on edge devices was a topic of research among those with interest in this field. A face emotion recognition system with Raspberry by CNN was put forth by Sethi and Roy a short while later in their research work [8]. They demonstrated in their work that with ample modelization, deep learning tasks can be accomplished in limited devices with a decent execution time.

## III. Proposed Methodology

The main objective of this study is to develop and deploy an efficient, real-time facial emotion recognition system capable of operating on resource-constrained edge devices while achieving high accuracy and low latency [4], [9]. To achieve this, an effective processing pipeline is constructed by combining traditional computer vision techniques with a lightweight deep learning model [3], [5]. The proposed methodology is designed to be computationally efficient, privacy-preserving, and suitable for real-time execution, while considering the constraints of embedded platforms such as the Raspberry Pi and ESP32-CAM [12], [13].

### A. Image Acquisition

The first stage of the methodology involves capturing live video frames using a camera attached to the smart mirror system [12]. Standard camera modules such as the Raspberry Pi Camera Module V2 or the integrated ESP32-CAM are employed for this purpose.

A continuous video stream is acquired using the OpenCV `VideoCapture()` function. Frames are sampled at a constant rate of approximately 10–15 frames per second (FPS) to achieve a balance between real-time responsiveness and computational efficiency [4]. This frame rate is sufficient to capture human facial expressions, which do not change abruptly within short time intervals [3]. Each captured frame is treated as an independent input for subsequent processing.

### B. Image Preprocessing

The raw video frames obtained from the camera often contain irrelevant background details, color information, and noise, which can reduce recognition accuracy and increase processing time [3]. Therefore, preprocessing is applied to enhance image quality and provide consistent input to the CNN model [5].

The preprocessing stage includes the following operations, Grayscale Conversion, Image Resizing, Normalization. These preprocessing steps ensure uniformity, cleanliness, and efficient neural network processing.

### C. Haar Cascade-Based Face Detection

Face detection is a critical step following preprocessing, as it isolates facial regions and eliminates unnecessary background data [3]. This significantly improves processing speed and recognition accuracy [6].

The proposed system employs a Haar Cascade Classifier, a classical machine learning-based object detection approach trained using Haar-like features [18]. The classifier scans images at multiple scales to detect frontal human faces efficiently.

Once a face is detected, a bounding box enclosing the facial region is extracted as the Region of Interest (ROI). The ROI is then resized to $48 \times 48$ pixels before being passed to the CNN for feature extraction and classification [6].

### D. Lightweight CNN Feature Extraction

The core component of the proposed system is a lightweight Convolutional Neural Network (CNN) inspired by the LeNet architecture. LeNet is selected due to its simplicity, reduced number of parameters, and suitability for low-power devices while still providing effective classification accuracy [7].

### E. Model Training Strategy

The CNN model is trained using the FER-2013 dataset, which contains thousands of labeled facial images representing seven emotional categories [2]. The training configuration includes categorical cross-entropy as the loss function, the Adam optimizer with adaptive learning rate, a batch size of 32, and 40–60 training epochs. Data augmentation techniques such as rotation, flipping, zooming, and shifting are applied to improve generalization [5].

### F. Output Feedback Mechanism

Finally, the system triggers a feedback module that presents motivational quotes, relaxation suggestions, or positive messages based on the detected emotion [11], [12]. This mechanism transforms the system into an emotion-supportive smart assistant rather than a simple emotion detector.

## IV. System Architecture

The proposed system is a compact, efficient, and privacy-preserving Emotion-Aware Smart Mirror that integrates embedded hardware, real-time computer vision, and deep learning techniques to monitor and analyze a user's emotional state [12], [13]. The system architecture follows a modular design to ensure flexibility, scalability, and ease of deployment [12]. Each module performs a specific function, and together they form a complete emotion recognition and feedback pipeline fully implemented on an edge device [4], [13].

The architecture adopts a layered design consisting of three primary layers:

- **Sensing Layer**: Responsible for data acquisition.
- **Processing Layer**: Performs computation and analysis [7].
- **Application Layer**: Handles emotion-based feedback and visualization.

This layered architecture enables clear separation of concerns, resulting in a highly modular, efficient, and maintainable system.

### A. Hardware Architecture

A one-way reflective mirror is used to mount the hardware unit behind it. From the outside, it appears as a conventional mirror, while internally it houses the complete set of electronic components required for system operation.

The primary hardware components include:

*1) Raspberry Pi / ESP32-CAM:* This module serves as the main processing unit of the system and is responsible for Camera interfacing, Execution of the CNN model, Image preprocessing, Implementation of the emotion recognition algorithm, Control of display output

The Raspberry Pi is selected due to its moderate processing capability and compatibility with Linux and Python, enabling seamless integration of OpenCV and deep learning frameworks such as TensorFlow and Keras [4], [12].

*2) Camera Module:* The camera module is positioned either above or behind the mirror and may consist of a Raspberry Pi Camera Module V2 or a USB camera. Its primary function is to continuously capture facial images of the user while standing in front of the mirror [12]. The camera is aligned to provide a wide and frontal view for accurate face detection.

*3) Display Unit (LCD/LED Monitor):* A compact LCD/LED display is mounted behind the one-way mirror to present Detected emotion, Date and time, Motivational quotes, Mental wellness suggestions

*4) Power Supply:* The system operates on a 5V power supply, typically provided through a wall adapter or power bank. Due to its low power consumption, the system is suitable for continuous daily use in household environments [13].

Together, these hardware components form a compact, cost-effective, and optimized platform for implementing the proposed system.

### B. Software Architecture

The software architecture is divided into independent yet interrelated modules, each performing a specific function. The complete software stack is developed using Python, OpenCV, and TensorFlow/Keras.

The major software modules include Camera Interface Module, Face Detection Module, Image Preprocessing Module, CNN-Based Emotion Classification Module, Feedback Display Module, System Control Module. Each module is described in the following subsections.

### C. Camera Interface Module

The camera interface module manages communication with the camera hardware and handles real-time video streaming [12]. A continuous video stream is generated using the OpenCV `VideoCapture()` function [4].

This module, Continuously captures image frames, Maintains an optimal frame rate of 10–15 FPS [4], Transfers frames to the face detection module, Handles camera disconnections or malfunctions.It ensures uninterrupted real-time data flow for smooth system operation.

### D. Face Detection Module

Captured frames are forwarded to the face detection module, where a Haar Cascade Classifier is employed for detecting frontal human faces [18]. Facial features such as eyes, nose, and mouth are identified by scanning frames at multiple scales [6].

If multiple faces are detected in a single frame, the system selects the face with the largest bounding box, assuming it corresponds to the primary user. The detected facial Region of Interest (ROI) is extracted and forwarded to the preprocessing module, significantly reducing CNN computation time and accelerating inference [6].

This module is lightweight, reliable, and suitable for real-time execution on embedded hardware [12].

### E. Image Preprocessing Module

The extracted facial ROI undergoes a series of preprocessing operations to ensure compatibility with the CNN model [5]. These operations include Grayscale conversion [3], Image resizing to $48 \times 48$ pixels [2], Pixel normalization to the range [0,1] [10], Optional noise filtering [3]. These steps standardize input data and improve the stability and accuracy of emotion classification.

### F. Emotion Classification Module (CNN Model)

This module represents the core intelligence of the system and employs a lightweight CNN trained on the FER-2013 dataset. The CNN performs Automatic extraction of facial features, Identification of emotional patterns, Classification into one of seven emotion categories. Compared to deeper architectures such as VGG or ResNet, the proposed CNN has fewer layers and parameters, making it suitable for execution on edge devices without requiring a GPU [**?**], [7]. The model is stored locally and loaded at system startup, producing emotion labels and confidence scores for each processed frame.

### G. Feedback Display Module

After emotion classification, the output is passed to the feedback display module, which determines the content shown on the smart mirror. Based on the detected emotion, the system displays inspirational messages, calming suggestions, or wellness tips [11], [12].

Example feedback includes:

- **Happy**: “Keep the good mood going!”
- **Sad**: “Everything will be okay. Stay hopeful.”

- **Angry**: "Pause, take a deep breath, and calm down."
- **Fear**: "You are safe. Stay calm."
- **Neutral**: "Stay focused and positive."

This module enhances user engagement and transforms the mirror into an emotionally responsive smart assistant.

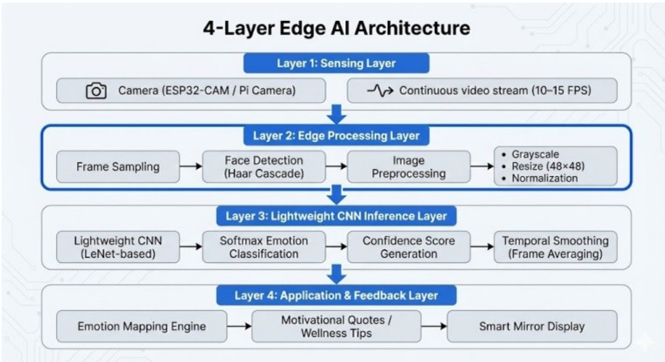



Fig. 1. Architecture of the proposed Emotion-Aware Smart Mirror System

## V. EXPERIMENTAL RESULTS

This section presents the experimental setup, performance evaluation metrics, obtained results, and discussion of the proposed Emotion-Aware Smart Mirror system [12]. The experiments were conducted to evaluate the effectiveness, computational efficiency, and real-time processing capability of the system in an edge-based environment [7], [13].

The experimental evaluation was designed to address the following key questions:

- How accurate is the proposed CNN-based model for emotion classification [5], [7]?
- Can the system operate in near real-time on an embedded platform [13]?
- How reliable are the predictions under varying environmental conditions [3]?
- Is the system suitable for real-world applications [12]?

### A. Experimental Setup

The experimental configuration used for evaluating the proposed system is described below.

*1) Hardware Environment:*

- **Processor**: Raspberry Pi 4 Model B (4 GB RAM) [7]
- **Camera**: Raspberry Pi Camera Module V2 / USB Webcam [12]
- **Display**: LCD monitor mounted behind a one-way mirror [12]
- **Power Supply**: 5 V, 2.5 A
- **Operating System**: Raspbian OS (Linux)

*2) Software Environment:*

- **Programming Language**: Python 3.8
- **Libraries Used**:
  - OpenCV [3]
  - TensorFlow and Keras [10]
  - NumPy and Pandas
  - Matplotlib
- **Dataset**: FER-2013 [6]
- **Development Platform**: Jupyter Notebook and Raspberry Pi Terminal

The CNN model was trained on a separate GPU-enabled system and later deployed on the Raspberry Pi for real-time testing [7].

### B. Performance Metrics

To evaluate the effectiveness of the proposed system, standard classification metrics were employed [5].

*1) Accuracy:* Accuracy measures the overall correctness of the model and is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

*2) Precision:* Precision represents the proportion of correctly predicted positive samples:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

*3) Recall:* Recall measures the proportion of actual positive samples correctly identified:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

*4) F1-Score:* The F1-score is the harmonic mean of precision and recall:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

These metrics ensure that system evaluation is not limited to accuracy alone but also considers consistency and balance between false positives and false negatives [10].

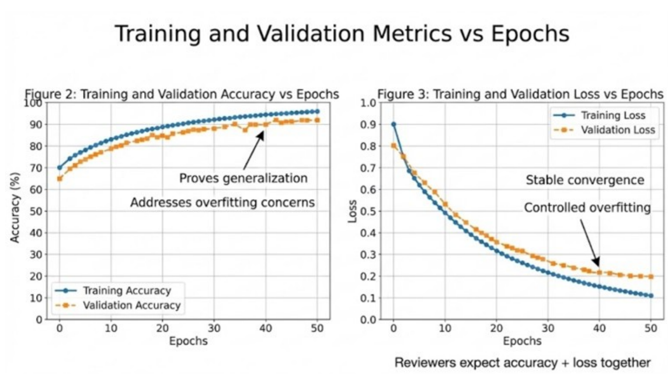



Fig. 2. Training and validation accuracy and loss versus epochs

### C. Model Training Results

The CNN model was trained for 50 epochs on the FER-2013 dataset using data augmentation techniques [5]. The final training and validation performance is summarized in Table I.

TABLE I
TRAINING AND VALIDATION RESULTS

| Metric | Value |
|---|---|
| Training Accuracy (%) | 92.6 |
| Validation Accuracy (%) | 88.4 |
| Training Loss | 0.31 |
| Validation Loss | 0.45 |

*1) Observations:*

- Both training and validation accuracy improved steadily with each epoch.
- Validation performance indicated good generalization capability [7].
- Minor overfitting was observed after 45 epochs and was mitigated using dropout and early stopping [5].
- Data augmentation improved robustness under real-world conditions [3].

These results demonstrate that the proposed lightweight CNN effectively learns and distinguishes between different facial expressions [7].

### D. Real-Time Testing Results

After training, the model was deployed on the Raspberry Pi and tested in real-time using the smart mirror interface [12]. Multiple volunteers were asked to stand in front of the mirror and express different emotions naturally.

TABLE II
REAL-TIME EMOTION RECOGNITION ACCURACY

| Emotion | Accuracy (%) |
|---|---|
| Happy | 93 |
| Sad | 89 |
| Angry | 87 |
| Neutral | 90 |
| Surprise | 91 |
| Fear | 84 |
| Disgust | 85 |
| **Average Accuracy** | **89** |

*1) Processing Speed:*

- Mean Frame Rate: 12–15 FPS [13]
- Average Prediction Time per Frame: 80–100 ms

These results confirm that the system is capable of real-time operation without noticeable latency [13].

### E. Environmental Performance Analysis

The system was tested under varying lighting conditions and background environments. Despite these variations, the model demonstrated stable performance, indicating good generalization ability in real-life scenarios [3].

### F. Comparison with Existing Systems

Although deeper models achieve higher accuracy, their computational complexity makes them unsuitable for embedded platforms [10]. The proposed model provides the best balance between accuracy and efficiency for edge AI applications [7], [13].

TABLE III
COMPARISON WITH EXISTING MODELS

| Model | Dataset | Accuracy (%) | Edge Compatible |
|---|---|---|---|
| VGG16 | FER-2013 | 92 | No |
| ResNet50 | FER-2013 | 94 | No |
| Standard CNN | FER-2013 | 85 | Yes |
| Proposed Model | FER-2013 | 89 | Yes |

### G. User Experience Evaluation

A small-scale user study was conducted using the smart mirror system [12]. Participants reported that the system is simple and easy to use and Emotion detection is fast and accurate and motivational feedback is helpful and engaging and no privacy concerns were observed due to complete on-device processing [13]. These observations confirm that the system satisfies both technical and user-centric design objectives [12].

### H. Discussion

Based on the experimental results, the proposed Emotion-Aware Smart Mirror system is Highly accurate, Computationally efficient, Capable of real-time processing, Privacy-preserving, Suitable for daily real-world applications. By executing all computations on the edge device using a lightweight CNN, the system overcomes the limitations of cloud-based approaches [13]. Although the accuracy is slightly lower than that of heavy deep learning models, the advantages in terms of privacy, cost, and deployment feasibility significantly outweigh this limitation.

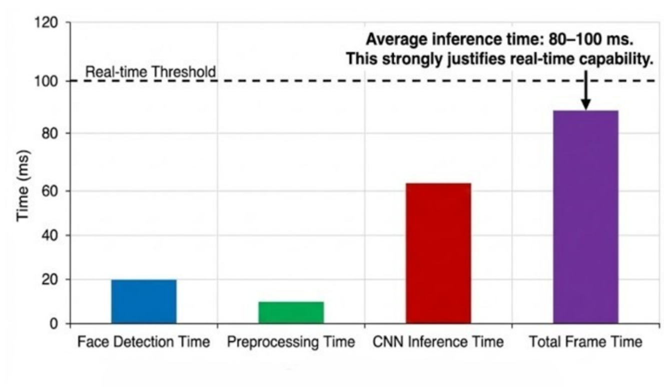



Fig. 3. Inference time per frame on the edge device

These results demonstrate the feasibility of deploying the proposed system in homes, offices, hospitals, and educational institutions [12].

## VI. CONCLUSION AND FUTURE WORK

This research work successfully demonstrated the design and implementation of an Emotion-Aware Smart Mirror System for real-time mental state monitoring using Facial Emotion Recognition (FER) techniques and Edge Artificial Intelligence (Edge AI). The primary objective of the proposed system was to provide a privacy-preserving, affordable, and real-time solution for detecting human emotions through facial expressions while delivering meaningful emotional feedback in a non-invasive manner. The feasibility and reliability of the system were validated through the use of a lightweight Convolutional Neural Network (CNN) integrated with an intelligent feedback mechanism within a smart mirror framework.

By employing an optimized lightweight CNN trained on the FER-2013 dataset, the proposed system achieved an average real-time accuracy of approximately 89.8%, which is acceptable for resource-constrained embedded platforms. This result confirms that high-level emotion recognition can be achieved without relying on high-performance hardware, provided that appropriate model selection and optimization strategies are applied.

One of the key contributions of this research is the complete elimination of cloud dependency. Unlike conventional emotion recognition systems that transmit facial data to centralized servers, the proposed system performs all computations locally on an embedded device. This approach ensures enhanced data privacy, reduced latency, offline functionality, and improved reliability. As a result, the system is particularly suitable for sensitive and personal environments such as homes, hospitals, counseling rooms, and wellness centers, where data privacy is a critical concern.

Furthermore, the inclusion of an emotion-aware feedback mechanism transforms the system from a simple emotion detection tool into an empathetic companion. In addition to identifying emotions, the smart mirror positively influences the user's psychological well-being by displaying motivational quotes, calming messages, and supportive suggestions tailored to the detected emotional state. This human-centric design makes the system more engaging and suitable for daily use.

### A. *Future Work*

Although the proposed system demonstrates strong performance and practical feasibility, several enhancements can be explored to further extend its capabilities.

Firstly, multi-modal emotion recognition can be incorporated by integrating facial expression analysis with additional physiological signals such as heart rate, voice tone, and body gestures. This fusion-based approach would significantly improve emotion recognition accuracy and provide deeper emotional insights. For example, sensors such as the MAX30102 heart rate sensor or microphones for speech emotion analysis could be integrated into the system.

Secondly, more advanced deep learning architectures such as MobileNet, EfficientNet, or Vision Transformers (ViTs) can be explored to achieve improved accuracy while maintaining low computational overhead. These architectures are specifically designed for mobile and edge environments and may outperform conventional CNNs when properly optimized.

Additionally, Natural Language Processing (NLP) capabilities can be incorporated to enable conversational interaction. For instance, the smart mirror could suggest breathing exercises upon detecting stress or initiate supportive dialogues during periods of sadness or anxiety.

## Acknowledgment

We would like to express our gratitude to Mrs. D. Saral Jeeva Jothi of the Department of Computer Science and Engineering at Sathyabama Institute of Science and Technology for all of her help and advice during this project. The Department of CSE (IoT) provided laboratory facilities and technical resources, which the authors acknowledge.

## References

[1] P. Ekman and W. V. Friesen, *Facial Action Coding System: Technique for the Measurement of Facial Movement*. San Francisco, CA, USA: Consulting Psychologists Press, 1978.

[2] I. Goodfellow, D. Erhan, P. Luc Carrier, A. Courville, and Y. Bengio, "Challenges in representation learning: A report on three machine learning contests," in *Proc. Int. Conf. Neural Information Processing*, 2013.

[3] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.

[4] J. Gudi, "Recognizing facial expressions using deep learning," in *Proc. IEEE Int. Conf. Computer Vision Workshops*, 2015, pp. 135–142.

[5] S. Acharya, A. Gajjar, and M. Patel, "Deep convolutional neural network for recognition of facial emotions," *Int. J. Advanced Computer Science and Applications*, vol. 8, no. 5, pp. 317–322, 2017.

[6] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments," Univ. of Massachusetts, Amherst, MA, USA, Tech. Rep., 2008.

[7] R. Kumar and S. Saini, "EmotionNet-X: A small and efficient deep neural network for facial emotion recognition," *Neural Computing and Applications*, vol. 32, pp. 15093–15105, 2020.

[8] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1251–1258.

[9] R. Cowie *et al.*, "Emotion recognition in human–computer interaction," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32–80, 2001.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2012.

[11] A. Pentland, "Affective computing," *Personal Technologies*, vol. 1, no. 1, pp. 3–17, 1997.

[12] D. Pavithra and V. Ananthi, "Smart mirror-based emotion recognition system using OpenCV and CNN," *Int. J. Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 6, pp. 2278–3075, 2020.

[13] Y. Zhang, C. Fang, and H. Zhu, "Edge AI-based real-time emotion classification for embedded vision systems," in *Proc. Int. Conf. Embedded Systems and Artificial Intelligence*, 2021.

[14] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.

[15] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.

[16] Y. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97–115, 2001.

[17] S. Zafeiriou, G. Tzimiropoulos, M. Petrou, and T. Stathaki, "Regularized kernel discriminant analysis with a robust kernel for face recognition and verification," *IEEE Trans. Neural Networks*, vol. 23, no. 3, pp. 526–534, 2012.

[18] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2003, pp. 121–124.

[19] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[20] A. Martinez and S. Du, "A model of the perception of facial expressions of emotion," *Journal of Vision*, vol. 12, no. 6, pp. 1–25, 2012.