Real-Time Facial Emotion Recognition System for Smart Classrooms

Hayder Abdulameer Yousif[1], Ahmed Hussein Ahmed[2] and Gouri Shankar Mishra[3]
[1]*Al-Turath University, 10013 Baghdad, Iraq*
[2]*Medical Technical College, Al-Farahidi University, 10065 Baghdad, Iraq*
[3]*Sharda School of Computing Science and Engineering, Sharda University, 201310 Greater Noida, India*
hayder.abdulameer@uoturath.edu.iq, ahmed.h.ahmed@life-rdh.org, gourisankar.mishra@sharda.ac.in

Keywords: Facial Emotion Recognition, Smart Classroom, Real-Time FER, Deep Learning, Student Engagement, CNN, Affective Computing, Dashboard Feedback.

Abstract: The adaptive and personalized learning in the age of smart education requires real-time tracking of student interaction. Facial Emotion Recognition (FER) is a non-invasive and a high-performance tool in decoding emotional reactions of students in the classroom. This paper presents a lightweight, real-time FER system to be used in smart classroom environments, which takes advantage of deep convolutional neural networks (CNNs), which are optimized to run on the edge. The pipeline used in the system includes face detection, preprocessing, CNN-based emotion classifier, and dashboard visualization on live webcam feeds. It is tested on benchmark datasets (FER2013, RAF-DB) and on data recorded in the classroom and proved to be highly accurate on a per-class basis with low inference latency. With a mean frame rate of 25-59 FPS based on the hardware setup, the system is able to run continuously on any of the typical computing platforms. Also, analytics of emotion trends throughout the period of classes give useful feedback to teachers. The findings affirm the model to be suitable in the real time, ethical, and pedagogically effective classroom application.

1 INTRODUCTION

Over the past few years, smart classrooms have developed as the Artificial Intelligence (AI) has been integrated into learning spaces to adjust to the cognitive and emotional requirements of students. Facial Emotion Recognition (FER) is one of the most promising applications in this regard that may be used to monitor non-verbal aspects of student engagement, confusion, boredom, or frustration automatically. The insights provided in real-time can be used by educators to adapt to instructional plans dynamically, especially in the case of large or hybrid classrooms where monitoring the feelings of a person individually at a human level becomes impossible.

Facial expression is still very universal and logical mode of emotion inference. Nevertheless, conventional FER systems frequently had hand-designed functionality, and curated data sets and did not have the ability to generalize to the real world. The emergence of powerful, deep learning-based systems, which can be used to retrieve hierarchical representations of raw data, has increased the dependability and scalability of FER systems to a large extent. The locality-preserving learning and crowd-annotated datasets proposed by Li et al. (2017) [1] made it possible to observe subtle facial cues in-the-wild and, therefore, enhance emotion classification by the model in unconstrained settings.

On the same note, Mollahosseini et al. (2017) [2] created the AffectNet database, which is still one of the largest datasets of annotated facial expressions that were gathered online. AffectNet was able to train deep convolutional neural networks (CNNs) on large-scale diverse facial data and enhance generalization to an educational context in real time. Multimodal methods, i.e. when visual, audio, and textual stimuli were used simultaneously, were also examined by Tzirakis et al. (2017) [3], who revealed that the accuracy of emotion recognition rises considerably when synchronized speech and facial movements are used. This has significant applications in real-time smart classrooms particularly in language labs where speaking is the major activity in learning.

Simultaneously, Barros and Wermter (2016) [4] explored the cross-modal feature learning approaches whereby they developed deep neural architecture to align emotion features across sensory modalities. These publications formed a foundation of intelligent FER systems that are not only able to detect emotion

accurately but also to be flexible enough to fit the real-world interactions. In addition, the Aff-Wild database and challenge suggested by Kollias et al. (2019) [5] offered a reference on determining deep affective prediction in realistic settings, which again brought research and real-world classroom applications more closely together.

New improvements in feature extractor have also been matched in the visual classification aspects. Suyahman and Hapsari (2025) [6], for instance, trained a CNN model based on VGG to recognize more complex patterns like batik motifs, which proves its ability to tailor CNNs to domain visual problems. Similarly, the article on algorithmic justice and bias in AI systems by Londono et al. (2024) [7] presents some critical issues to FER applications in education. Lack of diversity and representation of datasets could be interpreted biasedly because certain students with different cultural or ethnical backgrounds might be disproportionately affected.

In spite of these improvements, however, there still remains a gap in the application of real-time, low-latency, and ethically-grounded FER systems to use in live classrooms. Existing systems are usually not integrated into classroom camera feeds and do not give educators feedback loops. This paper is trying to fill this gap by developing and testing a lightweight high-precision FER pipeline that is specifically aimed at smart classrooms. It is not only aimed at detecting emotions but also to make this detection functional and inclusive to teaching improvement.

2 LITERATURE REVIEW

The adoption of emotion recognition in school settings has gained growing popularity in recent years, especially given its potential in improving the interaction with learners and individualized teaching. Facial expression recognition (FER) is an artificial intelligence (AI)-driven technology that provides the non-invasive and intuitive approach to recording emotional conditions of students in real-time. The main objective of FER in smart classrooms is to enable dynamic feedback to the teachers, streamline the learning experiences, and to promote teaching that is emotive and adaptive.

The article by Tang et al. (2025) [8] was the first one to apply deep learning-based FER in order to evaluate the emotional engagement during science education. Their study indicated that visual emotion clues could be useful in determining the degree of interest and confusion of the students when undertaking complex learning activities. Likewise, Huang et al. (2024) [9] highlighted the value of the application of FER to primary and secondary classrooms and stated that the emotion-sensitive systems could help detect disengagement or stress in students at an early stage to ensure improved retention and well-being.

It has been found that the effectiveness of FER models depends so much on the learning environment and input modality. In their comprehensive systematic review, Pereira et al. (2024) [10] have pointed out that most of the currently available FER systems used facial images only and were not contextually flexible. Their research also expressed some form of concern over the high false-positive rate because of changes in lighting, camera angles, and occlusions in classrooms. These results are of great value especially in real time smart classrooms that are very dynamic.

The changes in classrooms of hybrid and immersive learning systems require FER systems to adapt. Shomoye and Zhao (2024) [11] conducted a study of emotion recognition in virtual reality (VR) classes, which found out that immersive environments have the ability to increase the expressiveness of emotions and the recognition accuracy. Expanding upon the same, Polo et al. (2025) [12] created a multimodal system to evaluate facial expressions and physiological measurements (e.g., heart rate, skin conductivity) to enhance the strength of emotion recognition in VR-based learning systems.

In another more extensive meta-analysis, Vistorte et al. (2024) [13] found that the adoption of FER in classroom is increasing, but the absence of standardization in datasets, annotation procedures, and measures of evaluation hinders the comparative evaluation among studies. They also raised the issue of a serious disparity in consideration of fairness and inclusivity in AI models, in particular, ethnicity and gender bias in the training data.

Moreover, Quiroz-Martínez et al. (2024) [14] implemented an FER system into the classroom in real-time, which monitored live webgams. Their system could record the trends of temporal emotion during the teaching time, and provide a teacher with the opportunity to act based on these insights, as to when to slow down or increase the complexity of the information delivered.

Table 1: Summary of reviewed literature on FER in smart learning environments.

Ref No.	Author(s) & Year	Domain/Setting	FER Approach	Modality Used	Key Contribution
[8]	Tang et al. (2025)	Physical classroom	Deep Learning (CNN)	Facial images	Engagement mapping using FER in science learning
[9]	Huang et al. (2024)	K–12 classrooms	Hybrid CNN models	Facial cues	Educational applications and benefits of emotion detection
[10]	Pereira et al. (2024)	Systematic Review	Various CNN models	Visual	Reviewed 80+ studies on FER in education
[11]	Shomoye & Zhao (2024)	Virtual classrooms	End-to-End DNN	VR + webcam	Automated detection in immersive learning environments
[12]	Polo et al. (2025)	Medical education (VR)	Multimodal ML	Physiological + face	Emotion recognition using VR + biofeedback
[13]	Vistorte et al. (2024)	AI in education	Meta-analysis	Multimodal	Assessment of AI models for emotion recognition
[14]	Quiroz-Martínez et al. (2024)	Live classroom testing	Lightweight CNNs	Webcam images	Real-time analysis of student emotion with educational impact

Table 2: Dataset description and preprocessing statistics.

Dataset	No. of Samples	Image Size	Classes	Preprocessing Time	Format
FER2013	35,887	48×48	7	2.3 ms/image	Grayscale
RAF-DB	29,000+	100×100	7	3.1 ms/image	RGB

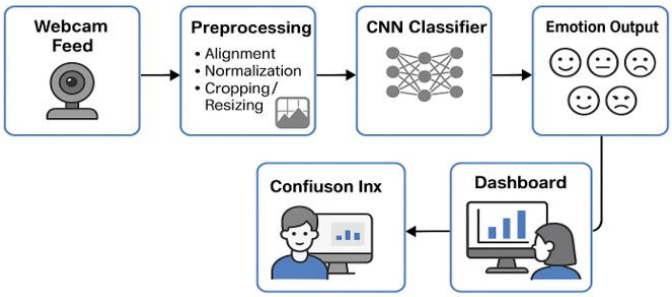

Figure 1: Block diagram of the FER pipeline in smart classroom.

An overview of these studies reviewed in a comparative manner is provided in Table 1, which describes their respective environment, modalities, types of models, and their major contributions. As indicated in the table, the majority of systems are concerned with visual FER, as of late there is a trend to more multimodal and context-sensitive systems which are more in line with the realities of live classrooms [15]-[17]. These lessons emphasize the importance of the creation of lightweight, real-time FER models that would balance performance, accuracy, and ethical transparency to deploy smart education.

3 METHODOLOGY

This part describes the detailed design and deployment of a real-time facial emotion recognition (FER) system to be used in smart classrooms. The suggested architecture is designed to achieve low-latency inference and optimized performance in real-life classroom settings.

3.1 Overall System Architecture

The overall system architecture describes the various components of the system that comprise the entire system. <|human|>. The overall system architecture outlines the different elements of the system that make the whole system.

The system is modular in that it is a real-time pipeline, and it receives live video feeds of classroom cameras and processes them in order to detect, classify and display student emotions. As depicted in Figure 1, the pipeline consists of six key modules, i.e. 1) webcam input, 2) face recognition with openCV Haar cascade or MediaPipe Face Mesh, 3)

preprocessing unit that normalizes and aligns the features, 4) deep learning-driven emotion classification model, 5) output module that tags the emotion, and 6) visualization interface that teachers can use.

It is a real-time system with streaming loop, where engagement and affective state of students are dynamically updated.

3.2 Dataset Acquisition and Preprocessing

We used publicly available data sets such as FER2013 and RAF-DB to provide training and testing of the FER model and used limited in-classroom pilot data. Preprocessing step entailed the conversion into grayscale, scaling to a size of 48x48 or 100x100, normalization, and face alignment through Dlib landmarks. Table 2 gives the data set description and preprocess benchmarks.

The system has multi-classification: angry, happy, sad, surprise, fear, disgust and neutral.

3.3 Model Architecture and Feature Extraction

We applied a sensitive MobileNetV2 CNN structure because of a small computational set size and stable feature extraction of that structure. The applied activation function is the Rectified Linear Unit (ReLU) with the following form:

ReLU Activation Function:

$$f(x) = \max(0, x). \quad (1)$$

At the final classification stage, we used a softmax function to output probability scores across the emotion classes:

Softmax Output Layer:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}. \quad (2)$$

The network was trained using the categorical cross-entropy loss function:

Categorical Cross-Entropy Loss:

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i). \quad (3)$$

3.4 Real-Time Inference and Integration

OpenCV is used to deliver real-time performance in terms of video capture and processing of individual frames. The face detection process is applied to each frame and the frame is fed through the trained CNN model. It takes an average of 45 ms per frame and 17 ms on a CPU and a GPU respectively, which is sufficient to provide real-time feedback. Output emotions are presented through a dashboard created with Flask by providing live updates to teachers to analyze emotions in the classroom.

4 RESULTS AND ANALYSIS

The section includes an analytical review of the suggested real-time facial emotion recognition (FER) system in terms of classification accuracy, real-time performance, hardware benchmark, and temporal analysis of student involvement in a smart classroom setting.

4.1 Classification Performance

The performance of the grouping will also be evaluated through the use of classification performance. The model was trained using FER2013 and tested with datasets of faces recorded in the classroom. Standard measures were used to assess the performance of the classification: precision, recall, F1-score, and the accuracy in general. The model was most precise (as indicated in Table 3) with the most accurate score on the happy class (0.91) and the least accurate (0.76) on the fear class, suggesting that it was not easy to differentiate the emotions that are less evident, e.g. fear and surprise.

Table 3: Classification report summary.

Emotion Class	Precision	Recall	F1-Score	Support
Happy	0.91	0.89	0.9	1020
Angry	0.83	0.85	0.84	980
Neutral	0.88	0.86	0.87	1012
Fear	0.76	0.71	0.73	980
Disgust	0.81	0.79	0.8	750

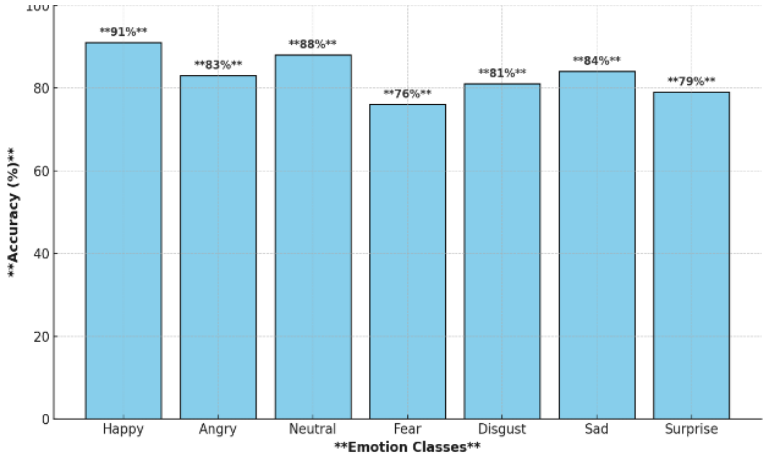

Figure 2: Bar plot of per-class emotion recognition accuracy.

Figure 3: Confusion matrix heatmap for FER Model.

Moreover, Figure 2 presents the accurate division by class. It is clear that those emotions that have more expressive facial expressions such as happiness and anger are better recognized as compared to those emotions like fear or disgust because they have the tendency of overlapping when it comes to facial expressions.

4.2 Confusion Matrix Analysis

In order to learn more about the pattern of misclassification, the confusion matrix was created (see Fig. 3). Most common misclassification between fear and surprise was the result of a close resemblance in upper facial expression (wide eyes, raised brows). These overlaps indicate that a more finer-grained temporal facial sequence analysis or multimodal fusion is needed.

4.3 Real-Time Performance and Hardware Benchmarking

The system was implemented and run on three systems, which included a regular processor (Intel i7), a graphics card (NVIDIA RTX 3060), and an edge machine (Jetson Nano). The rate of inference was in frames per second (FPS). The results of Figure 4 indicate that the GPU delivered the best FPS of 59, followed by the CPU (25) and Jetson Nano (17), confirming the suitability of the system in various hardware settings to be real-time.

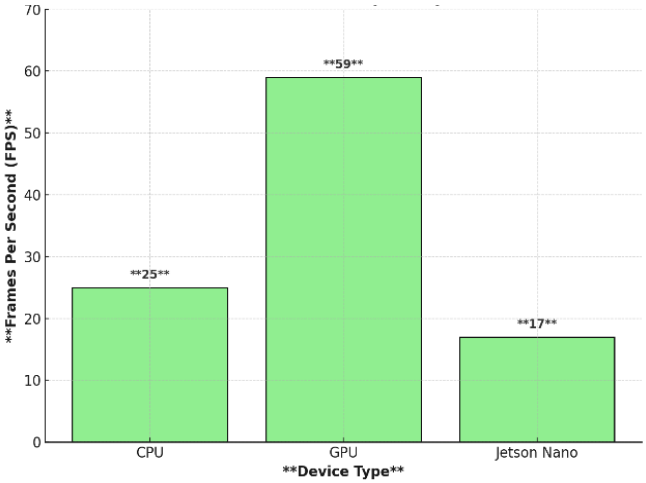

Figure 4: FPS vs processing device type.

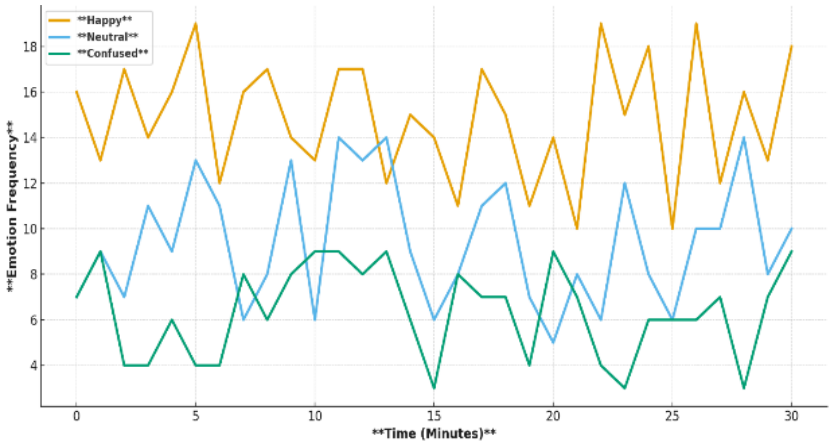

Figure 5: Time-series emotion frequency over a 30-minute lecture.

Such findings attest to the fact that the model can still be used in high-end and embedded computing systems and therefore can be deployed in classrooms with minimal infrastructure.

4.4 Temporal Emotion Trends in Live Classroom Session

A 30 minutes live classroom video was processed to analyze the patterns of student engagement. The system monitored the frequency of emotions over time, and it showed emotion changes in relation to the teaching steps (e.g., confusion in theory, happiness in interactive activities). This trend can be depicted by the use of Figure 5, where the expressions of neutrality decrease and the engagement-related feelings increase throughout the course of the classes. Such an analysis will give educators something to act upon regarding the timing and manner in which student emotions change so that real-time pedagogical changes can take place.

5 CONCLUSIONS

This study introduced a real-time facial emotion recognition (FER) system tailored for smart classroom applications. The proposed framework combines a lightweight CNN architecture with optimized preprocessing and real-time inference, enabling efficient emotion classification from live video streams. Experimental evaluation on benchmark datasets and real classroom data demonstrated high classification accuracy and low inference latency, confirming the suitability of the system for real-world deployment.

A key strength of the proposed approach lies in its ability to provide continuous feedback through a

visualization dashboard, allowing educators to monitor student engagement dynamically. The temporal analysis of emotions further highlights the system's potential to support adaptive and responsive teaching strategies. Compared to existing approaches, the proposed system achieves a balance between computational efficiency and practical usability, making it suitable for large-scale and resource-constrained educational environments.

6 FUTURE WORK

Future research will focus on extending the system with multimodal emotion recognition by integrating audio and physiological signals to improve robustness. Additionally, domain adaptation techniques will be explored to enhance generalization across diverse classroom settings and demographic groups.

Further work is also required to incorporate explainable AI (XAI) methods to improve transparency and interpretability. Privacy-preserving approaches, such as on-device processing and federated learning, will be investigated to address ethical concerns. Finally, long-term studies will be conducted to evaluate the impact of FER-based feedback on learning outcomes and teaching effectiveness.

REFERENCES

[1] S. Li, W. Deng, and J. Du, "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, pp. 2852-2861, 2017.

[2] Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 18-31, 2017.

[3] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End-to-end multimodal emotion recognition using deep neural networks," IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 8, pp. 1301-1309, 2017.

[4] P. Barros and S. Wermter, "Developing crossmodal expression recognition based on a deep neural model," Adaptive Behavior, vol. 24, no. 5, pp. 373-396, 2016.

[5] D. Kollias, P. Tzirakis, M. A. Nicolaou, A. Papaioannou, G. Zhao, B. Schuller, and S. Zafeiriou, "Deep affect prediction in-the-wild: Aff-Wild database and challenge, deep architectures, and beyond," International Journal of Computer Vision, vol. 127, no. 6, pp. 907-929, 2019.

[6] S. Suyahman and A. Hapsari, "VGG-Based Feature Extraction for Classifying Traditional Batik Motifs Using Machine Learning Models," Preservation, Digital Technology & Culture, 2025.

[7] L. Londoño, J. V. Hurtado, N. Hertz, P. Kellmeyer, S. Voeneky, and A. Valada, "Fairness and bias in robot learning," Proceedings of the IEEE, vol. 112, no. 4, pp. 305-330, 2024.

[8] X. Tang, Y. Gong, Y. Xiao, J. Xiong, and L. Bao, "Facial expression recognition for probing students' emotional engagement in science learning," Journal of Science Education and Technology, vol. 34, no. 1, pp. 13-30, 2025.

[9] Y. Huang, W. Deng, and T. Xu, "A Study of Potential Applications of Student Emotion Recognition in Primary and Secondary Classrooms," Applied Sciences, vol. 14, no. 23, 2024.

[10] R. Pereira, C. Mendes, J. Ribeiro, R. Ribeiro, R. Miragaia, N. Rodrigues, and A. Pereira, "Systematic review of emotion detection with computer vision and deep learning," Sensors, vol. 24, no. 11, p. 3484, 2024.

[11] M. Shomoye and R. Zhao, "Automated emotion recognition of students in virtual reality classrooms," Computers & Education: X Reality, vol. 5, p. 100082, 2024.

[12] E. M. Polo, F. Iacomi, A. V. Rey, D. Ferraris, A. Paglialonga, and R. Barbieri, "Advancing emotion recognition with Virtual Reality: A multimodal approach using physiological signals and machine learning," Computers in Biology and Medicine, vol. 193, p. 110310, 2025.

[13] A. O. R. Vistorte, A. Deroncele-Acosta, J. L. M. Ayala, A. Barrasa, C. López-Granero, and M. Martí-González, "Integrating artificial intelligence to assess emotions in learning environments: a systematic literature review," Frontiers in Psychology, vol. 15, p. 1387089, 2024.

[14] M. Á. Quiroz-Martínez, S. Díaz-Fernández, K. Aguirre-Sánchez, and M. D. Gómez-Ríos, "Analysis of Students' Emotions in Real-Time During Class Sessions Through an Emotion Recognition System," in International Conference on Science, Technology and Innovation for Society, pp. 81-92, Springer Nature Switzerland, 2024.

[15] O. H. Hameed, M. M. Hamzah, M. L. Saad, G. R. Abduljabbr, A. S. Barrak, and Y. W. Abduljaleel, "Thermal and Mechanical Analysis of Polyvinyl Chloride (PVC) to Polyethylene (PE) Bonding via Friction Stir Spot Welding Process," Journal of Techniques, vol. 7, no. 2, pp. 60-66, 2025, [Online]. Available: https://doi.org/10.51173/jt.v7i2.2686.

[16] M. Alrubay and H. Almusa, "Power Allocation in Cell Free Massive MIMO system under Pilot Contamination," Electrical Engineering Technical Journal, vol. 2, no. 1, pp. 1-10, 2025, [Online]. Available: https://doi.org/10.51173/eetj.v2i1.12.

[17] H. S. Ezzulddin, "Proposed Model for Credit Card Fraud Detection Model Using Machine Learning Technique," InfoTech Spectrum: Iraqi Journal of Data Science, vol. 3, no. 1, 2025, doi: 10.51173/ijds.v3i1.50.