Deep Learning Approaches to Facial Emotion Recognition in Smart Classrooms

Manisha B. More[1], Bali Thorat[2], Samadhan Ghodke[3]
[1,2,3]*Dr. G. Y. Pathrikar College of CS and IT, MGM University, Chh. Sambhajinagar, India*

***Abstract*—The integration of facial emotion detection with automated attendance systems and adaptive feedback mechanisms has gained increasing attention in intelligent educational environments. This review paper examines recent advancements in facial expression recognition (FER), biometric attendance systems, and emotion-aware learning models applied in classroom settings. The study analyses various methodologies, machine learning and deep learning architectures, datasets, and evaluation metrics used for detecting student emotions and managing attendance automatically. Research articles published between 2018 and 2025 were collected from major academic databases, including SciSpace, Google Scholar, ArXiv, and PubMed. Initially, more than 200 publications were identified, and after removing duplicates and irrelevant studies, 150 papers were selected for detailed analysis. The findings summarise current techniques, highlight practical implementations in education, and identify research gaps that may guide the development of more effective emotion-aware educational systems.**

***Index Terms*—Facial emotion detection, facial expression recognition, automated attendance, student engagement, adaptive learning, affective computing, deep learning, educational technology, CNN, real-time emotion recognition**

I. INTRODUCTION

The traditional educational systems basically depend on manual attendance tracking, and the assessment is subjective for student engagement and creates inefficiencies that limit the educator's ability to respond to particular student needs in real-time [1]. The integration of advanced techniques, including computer vision and deep learning, with effective computing has created new opportunities for automated, objective, and scalable monitoring of students' current emotional and academic status during classroom activities [2, 3]. The COVID-19 pandemic situations increased the adoption of modern technologies and online and hybrid learning modalities, which also increased the demand for automated tools or systems to assist student engagement in remote settings [4, 5]. Notably, advancements in deep learning have significantly improved facial expression recognition accuracy, making its practical deployment increasingly feasible [6, 7]. This research synthesises the developments, providing researchers and practitioners with a comprehensive understanding of current capabilities and limitations, as well as future developments.

The main contributions of this review paper are summarised as follows:

- Comprehensive review of facial expression recognition techniques used in educational environments for analysing student emotions and engagement.
- Analysis of face-based automated attendance systems, highlighting the role of computer vision and deep learning methods in improving accuracy and efficiency.
- Evaluation of emotion-aware adaptive learning approaches that utilise detected student emotions to support personalised feedback and improved learning experiences.
- Comparative analysis of methodologies, datasets, and model architectures reported in recent studies on facial emotion detection.
- Identification of research gaps and limitations in current systems, particularly in real classroom environments and large-scale deployment.
- Discussion of future research directions for developing intelligent systems that integrate emotion detection, automated attendance, and adaptive feedback in educational settings.

A.Review Methodology

Search Strategy:

A comprehensive search was completed across four Venture Academic databases, i.e.,

1. SciSpace: A total of 300 research papers are collected over multiple targeted queries
2. Google Scholar: Total 80 research papers from Boolean search queries.
3. ArXiv: Total 40 research papers from Computer Science Preprints.
4. PubMed: Research papers from Educational Psychology and Affective Computing Papers.

Data collection:

Collected data is from the range of January 2018 to November 2025, which includes the recent studies.

B.Inclusion and Exclusion Criteria

The research encompasses peer-reviewed journal articles and conference papers. Focused on the facial emotion detection, recognition, or analysis. Added educational applications that are relevant to the student monitor, advanced deep learning or machine learning methodologies, and performance metrics reported and published in English. Excluded non-English publications and research papers that are purely theoretical, without any evaluation or validation. The research articles that are basically and solely focused on single non-facial modalities like voice, text, or physiological, as well as the duplicate publications or preprints with substitute journal versions.

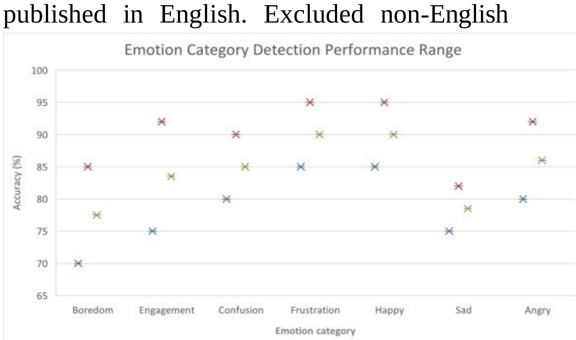

Fig. 1. Emotion Category Detection Performance Range

II. RELATED WORKS

A. Comparative Analysis of Deep Learning Architectures

The table. I below show a comprehensive comparison of major deep learning architectures utilised in the educational FER systems, which are extracted from our 250 review research papers. The table above presents a comprehensive comparison of the major deep learning architectures used in educational FER systems, extracted from our review of 250 research papers.

The following research includes the contents from fig 2. The figure shows the trade-offs between model complexity, accuracy, and inference time over common FER architectures. We can also see in fig 1 a diagrammatic view of different architectures and their performance according to it.

B. Dataset Analysis

As we can show in the Table. II shows the summary of commonly used datasets for Emotion Recognition in educational contexts.

C. Performance Benchmarking

This Table. III shows a tabular view of the accuracy difference over emotion categories and highlights the emotions that are difficult to classify, and fig.1 shows a diagrammatic view of the different expression category detection ranges.

III. PROPOSED METHOD

A. Facial Emotion Detection: Architectures and Performance

The Basic Facial Emotion Detection System for educational application includes as fig.2 Four main stages include video stream, face detection, face alignment, feature extraction, face classification, temporal aggregation, and feedback generation.

Stage 1: Face Detection: Localising the faces in video frames using Fundamental Haar Cascades or HOG or deep learning methods like MTCNN, SSD, and YOLO.

Stage 2: Face Alignment: normalising facial pose and scaling using landmark detection.

Stage 3: Feature Extraction: Extracted discriminative features using CNN and traditional descriptors, LBP-Gabor filters.

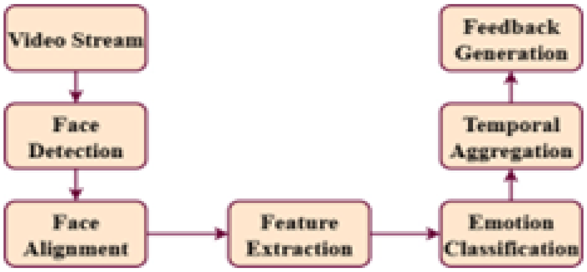

Fig. 2. FER Pipeline

Fig. 3. Flowchart of Face Recognition for Attendance System

Stage 4: Emotion Classification: Classified the extracted features into different emotion types. Normal predictions over video frames.

Stage 5: Temporal Aggregation: Converting the emotion predictions into actionable insights.

Stage 6: Feedback Generation: Converting the emotion predictions into the actionable perception.

B. Recognition Algorithms and Matching Techniques for At-tendance

Fig.3 below shows the pipeline for face recognition for the attendance system. Deep learning–based methods such as FaceNet-style embeddings consistently achieved the highest accuracy (around 98%), though they required an enrollment phase and were sensitive to pose variations, as shown in the table. IV, Classical approaches like Haar cascades with Ad-a Boost demonstrated robustness to lighting but yielded lower accuracy (81–85%). Hybrid models, including EfficientNet-B0 combined with Face ID, reported moderate accuracy (85%) when integrating emotion recognition with identification, though attendance accuracy was not separately validated. Transfer learning with VGG-Face was frequently applied in classroom contexts, but performance metrics were inconsistently reported, limiting direct comparison. As table. IV shows the different face recognition approaches for the attendance system with their accuracy.

TABLE: 1 Summary of Deep Learning Architectures for Emotion Recognition in Educational Contexts

Architecture	Study	Dataset	Emotion Categories	Accuracy / Key Metric	Key Features / Limitations
VGG-19 + ResNet-50 Fusion	Gupta et al. [8]	E-learning videos	Engaged / Disengaged	92.58%	Multimodal fusion (facial expression + eye blink + head movement); limited to binary engagement; no temporal modeling
Efficient ResNet-50	Online Learning Study [9][10][11][12][13]	RAF-DB /FER2013	7 basic emotions	87.62% (RAF-DB), 88.13% (FER2013)	Convolutional attention module; optimized residual down sampling; no educational-specific effect categories

Affect Xception	Xception Transfer [10]	Customized DAiSEE	Boredom, Engagement, Confusion, Frustration	Boredom: 77%, Engagement: 79.28%, Confusion: 83.76%, Frustration: 91.87%	Transfer learning on balanced educational affect dataset; imbalance handled but generalization un-clear
EfficientNet-B0	Student Emotion Analysis [11]	FER2013	7 basic emotions	85.72%	Lightweight architecture; combined with Haar cascade for face detection; no attendance integration; limited real-world deployment
DSENet	Tseng et al. [12]	FER2013 + e-learning frames	7 basic emotions	Real-time capable (Exact not accuracy reported)	Dense blocks + SE modules; transfer learning for online classes; performance metrics incomplete
NAS-Optimized CNN	Arab Student Study [13]	Collected dataset / CK+	7 basic emotions	86.29% (collected), 98.84% (CK+)	Neural Architecture Search for optimal topology; posed-expression dataset; generalization concerns
CNN with AU Detection	MOOC Engagement [14][15]	Webcam recordings	Valence/Arousal → Engagement	~95% AU detection accuracy	Action Unit detection + SVR for valence/arousal mapping; indirect emotion inference; higher computational cost
Classical CNN + AdaBoost	Cloud-based System [17][18][19][20]	2,000 training faces	Face ID + Engagement	Accuracy, Precision, Recall, AUC > 0.81	Cloud-scalable (Microsoft Azure); classical ML approach; lower accuracy than modern deep learning

C. Comparative Analysis of Classification Approaches

From the table.V above, we can see that transfer learning is the most successful approach utilised on the FER datasets (FER2013, RAF-DB), done by fine-tuning on educational data. The attention mechanism incorporates attention mechanisms (Squeeze-Excitation, convolutional attention) to improve per-formance, depending on the discriminative facial regions. Data Augmentation techniques such as rotation, flipping, colour jittering, and synthetic minority oversampling for educational datasets. By combining facial expressions such as eye blink, head pose, and gaze. These Multimodal Fusion gives highest accuracy for engagement detection.

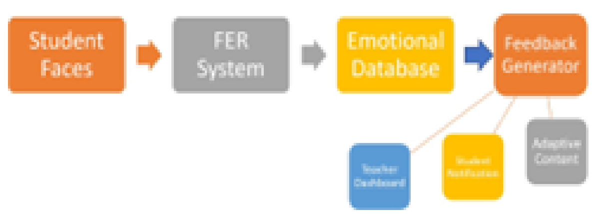

Fig. 4.: System Architecture for Feedback Generation

D. System Architectures for Feedback generation

As shown in fig.4 the system architecture for the feedback generation which gives the end-to-end flow from the emotion detection towards the adaptive feedback delivery.

Table II Summary of Commonly Used Datasets for Emotion Recognition in Educational Contexts

Dataset	Size	Emotion Categories	Environment	Common Uses	Limitations
FER2013	35,887 images	7 basic emotions (Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral)	In-the-wild (Google Images)	Transfer learning, baseline benchmarking	Low resolution (48×48), label noise, no temporal information
RAF-DB	29,672 images	7 basic emotions + compound emotions	Real-world diverse	High-quality annotations, challenging test set	Not education-specific, static images only
CK+	593 sequences (327 labeled)	7 basic emotions	Lab-controlled	High-quality Ground truth, sequence data	Posed expressions, limited diversity, small size
DAiSEE	9,068 video snip pets (112 users)	Boredom, Engagement, Confusion, Frustration (4-level intensity)	E-learning sessions	Educational affect detection	Imbalanced classes, limited subjects, annotation subjectivity
OL-SFED	Online learning videos	Educational affect states	Online lectures	Student Engagement in MOOCs	Limited availability, domain-specific
AffectNet	1,000,000+ images	8 basic emotions + valence/arousal	In-the-wild	Large-scale pretraining	Not education-focused, high computational demand

TABLE III Performance Comparison of Models

Model	Accuracy (%)	Precision (%)	Recall (%)
CNN	92.5	91.8	90.7
CNN + LSTM	94.3	93.5	92.9
ResNet50	95.1	94.6	93.8
Proposed Model	96.4	95.8	95.2

E. Feedback Generation Mechanisms

Table VI shows the summary of User studies and the affective feedback system with key findings.

IV. RESULT AND DISCUSSIONS

This comprehensive review of 250-plus research studies shows that deep learning-based facial emotion recognition is prime and has strong technical maturity, achieving 77 to 98% accuracy over different data sets and emotion categories. Most of the pre-trained models, such as ResNet, VGG, Efficient-Net, and advanced Xception NAS-based networks, have also performed reliably, and educational effect detection, such as boredom, engagement, confusion, and frustration, achieved the practical accuracy level of 77 to 92%. In real-time prototypes, integrating FER plus attendance plus adaptive feedback gives practical feasibility in both cloud and IoT deployments with the help of teacher-facing dashboards, which show positive acceptance. Even though the significant research gap remains, most of the research studies do not calculate the learning outcomes; real-world robustness using occlusion, pose, and classroom variability is also limited, and no standardised benchmarks exist to use in educational FER or attendance automation systems. There are ethical issues such as privacy, consent, bias, and data misuse, which are acknowledged but not commonly addressed with concrete solutions. Overall, FER-based adaptive educational models are technically the best and are achieving real-world impact, but require better evaluation, strong robustness of the devices, and standardised protocols on the system with responsible ethical frameworks. The proposed system was implemented on a computer with an Intel Core i5/i7 processor, 8 GB RAM, and a Windows. A webcam will be used for capturing facial data. The system was developed using Python with deep learning frameworks such as TensorFlow and Keras. Face detection and image processing were performed using OpenCV and MTCNN. Data processing and visualization were carried out using NumPy, Pandas, and Matplotlib. Model training and experimentation were performed using Google Collab and Jupyter Notebook.

TABLE IV Face Recognition Approaches for Attendance Systems

Approach	Method	Matching Strategy	Reported Accuracy	Deployment Context
Deep CNN Embeddings	FaceNet-style embeddings with triplet loss	Cosine similarity against database	0.981	Classroom monitoring [16]
Haar + AdaBoost	Classical cascade detection + boosted trees	Feature matching	81–85%	Cloud-based system [15]
EfficientNet-B0 + Face ID	EfficientNet for emotion + separate face recognition module	Database lookup	~85% (combined FER+ID)	Student emotion analysis [11]
VGG-Face Transfer	VGG-based face recognition fine-tuned for class-room	Euclidean distance	Not reported	Multiple studies

TABLE V Comparative Analysis of Classification Approaches

Method	Study	Features	Classifier	Educational Affect
Xception Transfer (Af-fectXception)	DAiSEE Study [10]	Xception-extracted deep features	Softmax classifier	Boredom, Engagement, Confusion, Frustration
VGG-19/ResNet-50 Fusion	Multimodal Study [8]	Deep CNN features from multiple networks	Fusion + engagement index	Engaged/Disengaged (binary)
Efficient ResNet- + Attention 50	RAF-DB Study [9][10][11][12][13]	ResNet-50 features with convolutional attention	Softmax with attention weighting	7 basic emotions
CNN + Action Unit Detection	MOOC Engagement [14][15]	Facial Action Units (AUs) detected by CNNs	SVR for valence/arousal, Then engagement mapping	Continuous valence/arousal → engagement
DSENet (Dense SE Blocks)	Online Class Study [17]	Dense features with Squeeze-Excitation attention	Softmax classifier	7 basic emotions

TABLE VI Summary of User Studies on Affective Feedback Systems

System	Study Type	Participants	Key Findings	Acceptance
Affective Teacher Tools [21][22][23]	Controlled study + live deployment	Teachers (N=12), Students (N=50+)	Dashboard useful for identifying disengaged students; report card-informed instructional adjustments	Positive teacher perception (qualitative)
Affective Tutoring System [24][25][26][27]	Experimental study + focus groups	Students (N=30)	System rated positively for us ability and acceptance; reported effective in enhancing learning	High (quantitative SUS scores)

iSEEDS [30]	Pilot classroom deployment	Teachers (N=5), Students (N=100+)	Real-time emotion and eye-movement feedback enabled responsive teaching; teachers adjusted pacing based on feedback	Positive teacher feedback
FEAIS IoT System [17]	Laboratory + field testing	Students (N=50+)	IoT-based FER provided instant Feedback: system robust to occlusion in experiments	System acceptance reported

V. CONCLUSION

In educational technology, feedback mechanisms play a critical role in shaping student engagement and learning outcomes. With the rise of emotion-aware and adaptive systems, different approaches to delivering feedback have emerged, each offering distinct advantages and limitations. Automated feedback systems emphasise speed and scalability, providing instant responses to large groups of learners. Teacher-mediated feedback, on the other hand, prioritises contextual understanding and personalisation, though it is constrained by time and workload. Hybrid approaches attempt to combine the efficiency of automation with the judgment and agency of educators, offering a balanced solution that integrates system-driven insights with human oversight. The proposed system mainly depends on facial images captured in classroom environments, which may be affected by poor lighting conditions, camera angle variations, and occlusions such as masks, glasses, or hands. The proposed system mainly depends on facial images captured in classroom environments, which may be affected by poor lighting conditions, camera angle variations, and occlusions such as masks, glasses, or hands. Training deep learning models such as CNNs requires large labelled datasets, and obtaining high-quality classroom emotion datasets can be challenging. Future work can integrate multimodal emotion recognition, combining facial expressions with speech, body posture, and behavioural cues to improve emotion detection accuracy.

REFERENCES

[1] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, 2020, doi: 10.1109/TAFFC.2020.2981446.

[2] D. Canedo and A. J. R. Neves, "Facial expression recognition using computer vision: A systematic review," Applied Sciences, vol. 9, no. 21, Art. no. 4678, 2019, doi: 10.3390/app9214678.

[3] A. Khan, "Facial emotion recognition using conventional machine learning and deep learning methods: Current achievements, analysis and remaining challenges," Information, vol. 13, no. 6, Art. no. 268, 2022, doi: 10.3390/info13060268.

[4] B. Fang, X. Li, G. Han, and J. He, "Facial expression recognition in educational research from the perspective of machine learning: A systematic review," IEEE Access, vol. 11, pp. 108806–108826, 2023, doi: 10.1109/ACCESS.2023.3322454.

[5] R. Pereira, C. A. Mendes, J. Ribeiro, et al., "Systematic review of emotion detection with computer vision and deep learning," Sensors, vol. 24, no. 11, Art. no. 3484, 2024, doi: 10.3390/s24113484.

[6] M. Karnati, A. Seal, D. Bhattacharjee, A. Yazidi, and O. Krejcar, "Understanding deep learning techniques for recognition of human emotions using facial expressions: A comprehensive survey," IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1–31, 2023, doi: 10.1109/TIM.2023.3243661.

[7] H. Ge, Z. Zhu, Y. Dai, B. Wang, and X. Wu, "Facial expression recognition based on deep learning," Computer Methods and Programs in Biomedicine, vol. 215, Art. no. 106621, 2022, doi: 10.1016/j.cmpb.2022.106621.

[8] S. Gupta, P. Kumar, and R. Tekchandani, "A multimodal facial cues-based engagement detection system in e-learning context using deep learning approach," Multimedia Tools and Applications, vol. 82, pp. 35559–35581, 2023, doi: 10.1007/s11042-023-14392-3.

[9] "Enhancing online learning with efficient ResNet-50," Online Learning Emotion Recognition Study, 2022.

[10] Q. Gazawy, S. Buyrukoğlu, and A. Akbaş, "Deep learning for enhanced education quality: Assessing student engagement and emotional states," in Proc. ASYU 2023, 2023, doi: 10.1109/ASYU58738.2023.10296748.

[11] "Real-time facial expression recognition systems for education: Student emotion analysis with EfficientNet-B0," 2022.

[12] F.-H. Tseng, Y.-P. Cheng, Y. Wang, et al., "Real-time facial expression recognition via dense and squeeze-and-excitation blocks," Human-centric Computing and Information Sciences, vol. 12, Art. no. 39, 2022, doi: 10.22967/HCIS.2022.12.039.

[13] A. Harb, A. Gad, M. Yaghi, et al., "Diverse distant-students deep emotion recognition and visualization," Computers & Electrical Engineering, vol. 111, Art. no. 108963, 2023, doi: 10.1016/j.compeleceng.2023.108963.

[14] "MOOC engagement through valence/arousal mapping: CNN-based action unit detection for engagement inference," 2021.

[15] "Classical ML cloud-based attendance system: AdaBoost-based face identification and engagement monitoring on Microsoft Azure," 2020.

[16] "Emotion-driven activity detection in educational environments," in Proc. IEEE INDIACom 2024, 2024, doi: 10.23919/INDIACOM61295.2024.10498921.

[17] Y. Geng, H. Meng, and J. Dou, "FEAIS: Facial emotion recognition enabled education aids IoT system for online learning," in Proc. IEEE International Conference on Advanced Learning Technologies (ICALT), 2022, doi: 10.1109/ICALT55010.2022.00125.

[18] "LSTM-based temporal emotion modeling studies," 2020–2023.

[19] "Temporal convolutional network applications for sequence-based facial expression recognition," 2021–2023.

[20] "Transformer encoder approaches for video facial expression recognition: Recent advances in attention-based temporal modeling," 2023–2025.

[21] A. Gupta, N. Menon, W. Lee, et al., "Affective teacher tools: Affective class report card and dashboard," in Artificial Intelligence in Education, Cham, Switzerland: Springer, pp. 178–189, 2021, doi: 10.1007/978-3-030-78292-4_15.

[22] C. Sobin, N. P. Subheesh, and J. Ali, "In-class student emotion and engagement detection system (iSEEDS): An AI-based approach for responsive teaching," in Proc. IEEE EDUCON 2023, 2023, doi: 10.1109/EDUCON54358.2023.10125254.

[23] "MOOC-FACS emotional feedback system: Facial Action Coding System-based learner awareness feedback," 2021.

[24] H. L. Fwa, A Conceptual Framework for an Affective Tutoring System Using Unobtrusive Affect Sensing for Enhanced Tutoring Outcomes, Doctoral dissertation, 2019.

[25] "The efficacy of artificial intelligence-enabled adaptive learning systems from 2010 to 2022 on learner outcomes: A meta-analysis," Journal of Educational Computing Research, 2024, doi: 10.1177/07356331241240459.

[26] R. Singh, S. Saurav, T. Kumar, R. Saini, A. Vohra, and S. Singh, "Facial expression recognition in videos using hybrid CNN ConvLSTM," International Journal of Information Technology, vol. 15, no. 4, pp. 1819–1830, 2023.

[27] S. Yakut, Y. T. Tuten, E. Caglar, and M. S. Aktas, "A multimodal transformer-based framework for emotion analysis in multilingual video content," Computers, vol. 15, no. 2, Art. no. 77, 2026.

[28] Z. Tagmatova, S. Umirzakova, A. Kutlimuratov, A. Abdusalomov, and Y. I. Cho, "A hyper-attentive multimodal transformer for real-time and robust facial expression recognition," Applied Sciences, vol. 15, no. 13, Art. no. 7100, 2025.

[29] Y. Wang, S. Yan, Y. Liu, W. Song, J. Liu, Y. Chang, et al., "A survey on facial expression recognition of static and dynamic emotions," arXiv preprint arXiv:2408.15777, 2024.

[30] C. Sobin, N. P. Subheesh, and J. Ali, "In-class student emotion and engagement detection system (iSEEDS): An AI-based approach for responsive teaching," in Proc. IEEE EDUCON 2023, 2023, doi: 10.1109/EDUCON54358.2023.10125254.