International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1462-1469
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0199

# AI-Based Smart Attendance and Behaviour Monitoring System

*Rahul Kumar Ray Kurmi[1], Sudeep Chaudhary[2], Anish Antony[3]*
*[1,2] UG Scholar, Dept. of Computer Science And Engineering,KPR Institute Of Engineering and Technology, Coimbatore-641407, Tamil Nadu, India. [3]Assistant Professor II, Dept. of CSE(Artificial Intelligence and Machine Learning),KPR Institute Of Engineering and Technology, Coimbatore-641407, Tamil Nadu, India*
***Emails**: rahulray12213@gmail.com[1], csudeep830@gmail.com[2], anishantony@kpriet.ac.in[3]*

## Abstract
*Traditional attendance systems suffer from time inefficiency, susceptibility to proxy attendance, and a complete absence of engagement insights. This paper presents an AI-based smart attendance and behaviour monitoring system that unifies facial recognition with real-time behavioural analysis into a single pipeline. The proposed system employs Multi-task Cascaded Convolutional Networks (MTCNN) [1] for robust face detection, FaceNet [2] for generating 128-dimensional facial embeddings, and a custom Convolutional Neural Network trained on a hybrid dataset for classifying pedagogically relevant emotional states. Experimental evaluation demonstrates an attendance recognition accuracy of 96.8% in multi-person classroom scenarios and approximately 90% accuracy in engagement-related emotion detection. A Flask-based web dashboard [3] provides real-time monitoring and comprehensive analytical reporting. Deployment across live classroom environments confirms that the system recovers 5–10 minutes of instructional time per session, eliminates proxy attendance, and supports data-driven pedagogical interventions — demonstrating both technical reliability and institutional practicality for real-world academic deployment.*

***Keywords:** Face Recognition, Automated Attendance Systems, Emotion Recognition, Deep Learning, MTCNN, Behavioural Analytics.*

## 1. Introduction
Attendance tracking sits so deep in academic routine that its real costs rarely get examined. In most institutions, calling roll consumes five to ten minutes per session — modest in isolation, but significant when compounded across a full semester. The integrity problem is equally serious: manual registers are trivially easy to manipulate, and the consequences ripple into performance assessments and institutional reporting. What makes this most frustrating is that even a perfectly accurate register captures only one fact — whether a body was in the room. A student can sit through an entire lecture in a state of complete cognitive absence and the sheet will still show perfect attendance, leaving instructors with no reliable signal about where confusion is building or when they have lost the room entirely.Prior modernisation efforts have had an uneven track record. Fingerprint scanners reduced clerical errors but created queuing bottlenecks and hygiene concerns that became harder to ignore after 2020. RFID systems solved the contact problem while introducing a new one — cards can be shared as easily as a pen can sign a name. Neither approach scaled gracefully to large classrooms, and neither addressed the more fundamental limitation: they track presence, not engagement, and a great deal of student struggle accumulates quietly in the gap between those two things. The system presented in this paper was designed around that gap. Rather than running identity recognition and behavioural analysis as separate pipelines, we built a unified architecture operating in real time from standard classroom cameras, requiring no deliberate action from students. The emotion classifier was developed in consultation with instructors and grounded in educational psychology, producing four classroom-specific states — attentive, neutral, confused, and disengaged — that map directly onto decisions instructors can act on. Validated through real-world classroom deployment. The remainder of this paper is structured as follows: Section 2 reviews prior work, Section 3 details system architecture, Section 4 presents experimental results, Section 5 discusses findings and

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1462-1469
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0199

limitations, and Section 6 concludes.

## 2. Related Work

Keeping track of who shows up to class — and whether they are actually paying attention — has frustrated educators for decades. What started as a purely administrative challenge has gradually drawn the interest of technologists, and the resulting body of research spans everything from fingerprint readers bolted to classroom doors to neural networks parsing facial expressions in real time.

### 2.1.Automated Attendance Systems

Biometric hardware was the first serious attempt to take attendance out of human hands. Fingerprint scanners had obvious appeal: they were harder to cheat than a signature sheet and did not rely on anyone remembering to do something [4]. But the practical reality in a classroom of fifty students was less clean. Everyone queuing at a single scanner at the start of a lecture is not a solution to administrative friction. Camera-based systems looked attractive partly because they required nothing from the student at all — no card, no fingerprint, no deliberate action. Early implementations leaned on Haar Cascade classifiers [5] and Local Binary Pattern descriptors, both of which had genuine strengths: they ran quickly and did not demand much from the hardware. Their weakness was a fairly narrow operating envelope. Push them outside the conditions they were designed for — strong afternoon light, a student wearing a hood, a cheap webcam at an awkward angle — and accuracy fell off sharply.

### 2.2. Deep Learning for Face Recognition

Deep learning did not so much improve face recognition as transform what the problem looked like. The shift from handcrafted feature descriptors to learned convolutional representations turned out to matter enormously. DeepFace [6], which came out of Facebook's research group, was one of the earlier systems to demonstrate just how much ground had been covered — its accuracy on standard benchmarks was close enough to human performance to make the comparison uncomfortable. FaceNet [2] took a different approach by using triplet loss to push the embedding space toward a geometry where the same person's face always lands close to itself and far from everyone else's, producing a compact 128-dimensional representation that can be compared with a simple distance calculation.

### 2.3. Emotion Recognition and Behaviour Analysis

The psychology underlying most automated emotion recognition systems traces back to Ekman and Friesen's work in the late 1960s and early 1970s, which argued for a small set of universal facial expressions tied to discrete emotional states [7]. CNN-based approaches trained on datasets like FER-2013 [8] and AffectNet [9] improved recognition in messier, real-world conditions, and adding recurrent layers to capture temporal dynamics pushed accuracy further [10]. However, a more fundamental issue persists: standard emotional taxonomies were never designed with classrooms in mind. Whether a student looks happy or sad tells a teacher very little. Whether they look confused, disengaged, or suddenly focused — those distinctions matter, and they do not map neatly onto standard frameworks [7].

### 2.4. Integrated Monitoring Frameworks

The natural next step would seem to be combining attendance tracking and engagement analysis into a single system — yet it is a little surprising how rarely this has actually been done. The more common pattern in the literature is two separate pipelines sitting next to each other, sharing nothing and duplicating work. Multi-task learning research [11] has shown fairly clearly that joint training over shared representations can reduce this redundancy without hurting performance on either task, but the insight has not translated into many real deployed systems.

### 2.5. Research Gap and Motivation

There is, in short, no existing system that satisfactorily handles all of this at once: contactless operation, genuine classroom validation, privacy-aware design, engagement monitoring grounded in learning-relevant emotional states, and longitudinal analytics with practical value for instructors. Each piece exists somewhere in the literature, but they have not been brought together. That integration is what this work sets out to achieve.

## 3. SYSTEM DESIGN AND IMPLEMENTATION

This section presents the design principles, architectural structure, and implementation details of the proposed system. The system is implemented as a coordinated processing pipeline comprising modular

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1462-1469
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0199

subsystems to ensure scalability, robustness, and real-time performance in classroom environments.

### 3.1. Architectural Overview

The system is built around four main components that work together as a continuous pipeline: Video Acquisition, Face Detection and Recognition, Behavioural Analysis, and Data Management with a Web Interface. Keeping these as separate modules was a deliberate choice — it means any one component can be reworked or improved without disturbing the rest of the chain, which matters considerably once a system like this is running in a live environment. Camera feeds from standard classroom-mounted webcams or IP cameras enter through the acquisition module first, which normalises differences between input sources before anything meaningful happens to the footage. Each normalised frame then moves to the face detection module [1], where facial regions are localised before being passed forward. From that point, processing splits into two simultaneous tracks — one handling identity verification through facial recognition [2], the other assessing behavioural state through emotion classification. Both outputs feed into a centralised database, and the web interface sitting on top of that database gives authorized instructors access to live session monitoring as well as historical engagement and attendance data shown in Figure 1 .

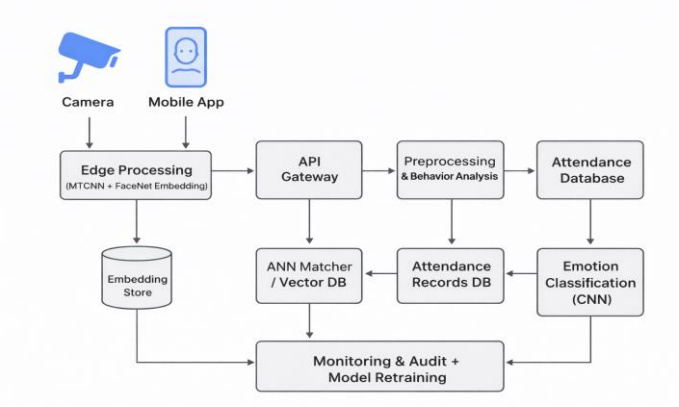


**Figure 1. Overall architecture of the proposed AI-based smart attendance and behavior monitoring system.**

### 3.2.Face Detection Pipeline

Face localisation is handled by MTCNN [1], a cascaded architecture chosen because it progressively refines its own detections across three stages rather than committing to a single-pass result. The first stage — the Proposal Network — scans image pyramids across multiple scales to generate an initial pool of candidate face regions, biased toward high recall so that no real face gets dropped early. Those candidates then move to the Refine Network, which uses deeper convolutional layers to reject false positives more reliably and tighten bounding box positions. The final Output Network confirms each detection and extracts five facial landmarks — both eye centres, the nose tip, and both mouth corners — which feed directly into the geometric alignment step, ensuring that faces arriving at the recognition and emotion modules are consistently normalised regardless of student head position.

### 3.3.Face Recognition and Identity Verification

Identity verification runs through FaceNet [2], which maps each detected face into a 128-dimensional embedding vector using a triplet loss objective that pulls same-person embeddings closer together while pushing different identities apart — making recognition at runtime a simple distance calculation. Enrolment takes approximately thirty seconds per student, capturing twelve to fifteen images across small head movements and expression variations to build a reference embedding that holds up across the range of angles a classroom camera realistically encounters. Each face is aligned using MTCNN landmarks, normalised, and averaged into a single stored vector. No photographs are retained at any point. During live sessions, detected faces go through identical preprocessing before their embeddings are compared against stored references, with a distance threshold of 0.6— determined empirically by balancing false acceptance against false rejection rates — triggering automatic attendance logging with a confidence score and millisecond-precision timestamp.

### 3.4.Database Architecture and Data Management

The backend runs on a normalised MySQL schema built around four core tables — Students, Attendance Records, Behaviour Observations, and Admin Users — connected through foreign key constraints to maintain data integrity. Student records hold academic metadata and serialised facial embeddings rather than photographs; attendance entries capture

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1462-1469
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0199

session context and confidence scores with millisecond-precision timestamps; and behavioural observations are stored at the same resolution to allow meaningful temporal analysis of how engagement shifts across a session shown in Figure 2.

Security was treated as a structural concern rather than an add-on: data at rest is encrypted with AES-256, transmission runs over SSL/TLS, and role-based access control ensures instructors can only retrieve records from their own courses. Storing embeddings instead of images was a deliberate privacy decision — a 128-dimensional vector carries no recoverable visual information if the database is ever compromised. This approach aligns with the data minimisation principles outlined under GDPR [12].

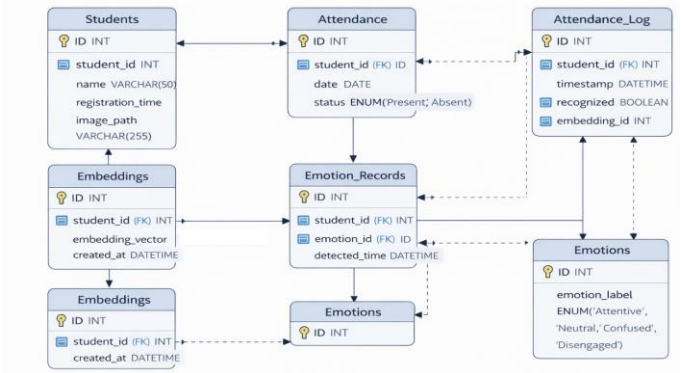


**Figure 2. Relational database schema used for attendance management and behavioral monitoring in the proposed system.**

### 3.5. Web Dashboard Interface

The dashboard is built on Flask [3] following a Model-View-Controller structure, with SQLAlchemy handling database queries and Jinja2 managing templating. The frontend uses HTML5, CSS3, and jQuery for asynchronous updates — the priority was something instructors would actually find intuitive rather than technically impressive. The live monitoring view shows the annotated camera feed in real time, with bounding boxes around detected faces, identity labels, and colour-coded emotion indicators updating continuously over a WebSocket connection. On the analytics side, instructors can browse attendance histories, track how engagement patterns shift across weeks, and drill into individual student trajectories when needed. PDF and CSV export is available directly from the interface for institutional reporting, without requiring any technical knowledge to operate shown in Figure 3.


**Figure 3. Web-based attendance and enrollment interface used for student registration and attendance capture.**

## 4. Experimental Evaluation And Results

This section presents a comprehensive evaluation of the proposed system, covering face recognition accuracy, emotion classification performance, processing efficiency, and real-world classroom deployment feasibility.

### 4.1.Testing Infrastructure and Methodology

All experiments were conducted on workstation hardware representative of institutional computing resources:

- Processor: Intel Core i7-9700K (8 cores, 3.6 GHz base frequency)
- Memory: 32 GB DDR4 RAM
- Graphics: NVIDIA GeForce RTX 2060 (1920 CUDA cores, 6 GB VRAM)
- Cameras: Logitech C920 HD webcams at 1920×1080 resolution, 30 fps
- Operating System: Ubuntu Linux 20.04 LTS
- Programming Language: Python 3.8
- Deep Learning Frameworks: TensorFlow 2.6 and Keras
- Computer Vision Library: OpenCV 4.5 [13]
- Web Framework: Flask 2.0 [3]

A total of 150 students were enrolled in the system, with 12–15 facial images captured per individual

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1462-1469
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0199

during guided enrolment sessions, resulting in approximately 2,100 enrolment images. The emotion classification model was trained on a hybrid dataset of 35,887 images from FER-2013 [8] combined with 2,400 classroom-collected samples.

### 4.2. Face Recognition Performance Analysis

Face recognition accuracy was evaluated across multiple operational scenarios representative of real classroom conditions. Table I summarizes the observed performance shown in Table 1.

**Table 1: Face Recognition Accuracy Under Different Scenarios**

| Scenario | Total Detections | Correct IDs | Accuracy (%) |
|---|---|---|---|
| Single individual | 50 | 49 | 98.0 |
| Five simultaneous students | 250 | 242 | 96.8 |
| Reduced lighting conditions | 40 | 37 | 92.5 |
| Partial facial occlusion | 30 | 27 | 90.0 |

Under standard conditions the system held consistently above 96%, with performance dropping predictably under poor lighting and occlusion — the latter only becoming a real problem when facial coverage exceeded roughly 40%. Spoofing attempts using photographs and video recordings failed entirely, since flat images lack the embedding characteristics that live faces produce. Even similar-looking siblings were correctly distinguished through embedding distance analysis, though with noticeably narrower confidence margins than typical matches.

### 4.3. Processing Efficiency Metrics

End-to-end latency across the pipeline came out at 160ms per frame — 85ms consumed by MTCNN detection [1] and 75ms by FaceNet embedding generation [2] — translating to roughly 6–7 frames per second. That throughput is modest but practically sufficient; students enter a classroom over a window of several minutes rather than all at once, so the system never faces the kind of burst demand that would make 6–7 fps a genuine bottleneck.

### 4.4. Emotion Classification Performance

**Table 2: Emotion Classification Accuracy**

| Emotion Category | Test Samples | Correct | Accuracy (%) |
|---|---|---|---|
| Attentive | 50 | 45 | 90.0 |
| Neutral | 50 | 48 | 96.0 |
| Confused | 50 | 43 | 86.0 |
| Disengaged | 50 | 44 | 88.0 |

Neutral states scored highest at 96%, which is not surprising — a relaxed face is more consistent across individuals than any expressive state shown in Table 2. Confused and disengaged were the harder cases, scoring 86% and 88% respectively, largely because both involve reduced activation and the distinguishing signals — gaze direction and brow tension — are subtle enough to overlap in ambiguous frames [11].

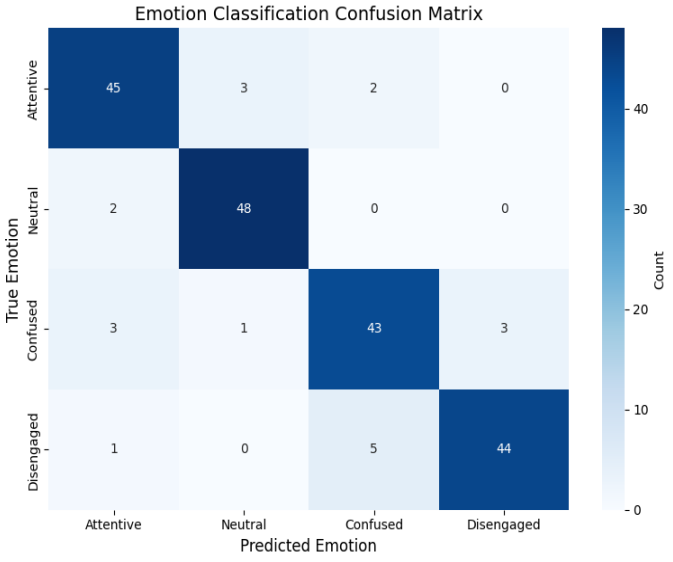


**Figure. 4. Confusion matrix for four-class emotion classification model.**

### 4.5. Real-World Classroom Deployment

The pilot ran across three classroom sections over four weeks, covering 36 instructional sessions in total. Overall attendance accuracy came out at 97.2%

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1462-1469
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0199

against manual verification — most errors traced back to brief camera obstructions or students sitting at angles the cameras struggled with rather than any systematic failure. The time saving was more significant than expected: automated attendance completed within 30 seconds of a session starting, against an average of 7 minutes for manual roll-call, recovering roughly 6.5 minutes per session and adding up to 5.2 hours of instructional time per week across the three sections. Beyond attendance, the behavioural monitoring flagged five students showing confusion patterns across more than 40% of session duration — a signal that would have been invisible under normal classroom conditions. Targeted interventions followed, and three of those students showed measurable gains in subsequent assessments. Student surveys returned a 92%. Response rate, with 78% reporting positive reception and citing reduced administrative disruption as the main benefit.

### 4.6.Comparative System Analysis

The proposed system was benchmarked against existing attendance approaches [1], summarised in Table 3.

**Table 3: Comparison with Existing Attendance Systems**

| System Type | Accuracy | Engagement Monitoring | Processing Time |
|---|---|---|---|
| RFID cards | 99.1% | Not supported | 45s (queue) |
| Fingerprint scanners | 98.5% | Not supported | 60s (queue) |
| Basic CNN recognition | 89.3% | Not supported | Real-time |
| Proposed system | 96.8% | Yes (90%) | Real-time (30s) |

RFID and fingerprint systems do edge out the proposed system on raw accuracy, and that is worth acknowledging honestly. What they cannot do is operate without physical interaction, scale without creating entry queues, or provide any information about what students are actually doing once they are in the room. The proposed system's combination of contactless operation, real-time engagement monitoring, and practical time savings represents a qualitatively different kind of contribution shown in Figure 5.

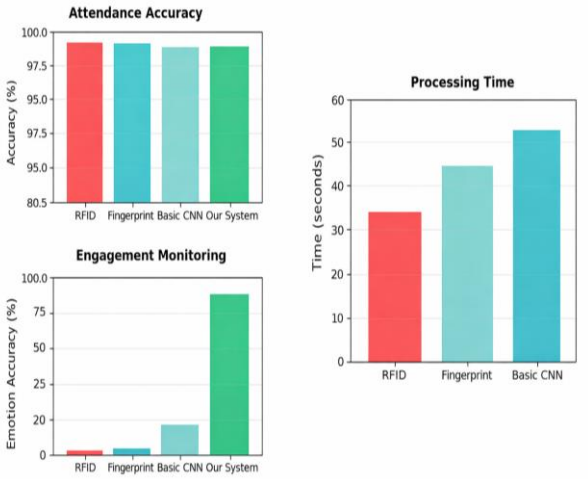


**Figure . 5. Comparison of attendance accuracy, engagement monitoring capability, and processing time across different systems.**

## 5. Discussion And Analysis

This section interprets the experimental findings, discusses their pedagogical and operational implications, examines system limitations, highlights unexpected observations, and outlines directions for future research.

### 5.1. Key Findings and Their Implications

The results suggest that combining facial recognition with behavioural monitoring produces something meaningfully more useful than either component alone [11]. Attendance accuracy at 96.8% met institutional requirements, with occasional errors handled through routine instructor oversight. The behavioural component delivered the more compelling findings — during one calculus session, confusion levels jumped from 15% to 45% within three minutes of a new concept being introduced, prompting immediate instructor intervention that brought confusion back to baseline after a brief clarification. That kind of real-time feedback has no equivalent in a conventional classroom, where comprehension gaps typically stay invisible until assessments surface them [7]. Operationally, attendance automation recovered an estimated 260–400 minutes weekly across a 40-section department.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1462-1469
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0199

Faculty feedback pointed consistently toward the engagement analytics as the feature that felt genuinely new — something traditional observation simply could not produce**.**

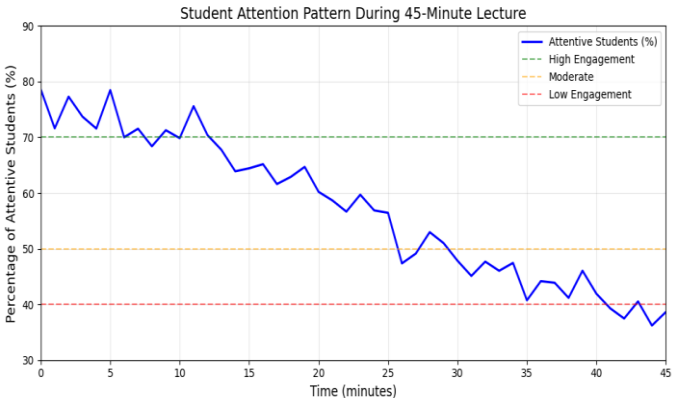


**Figure. 6. Temporal variation of student attentiveness during a 45-minute lecture session.**

### 5.2.. System Limitations and Challenges

Several practical limitations emerged during deployment. Reduced lighting dropped recognition accuracy to 92.5%, with early morning and late afternoon sessions most affected. Partial occlusion from masks pushed accuracy down further to around 90%, producing 3–4 misidentifications in a typical 40-student classroom. Demographic bias auditing was not fully completed due to limited population diversity in the deployment environment. Given documented concerns about bias in facial recognition systems [6], rigorous evaluation across diverse populations remains a prerequisite before broader institutional adoption can be responsibly recommended.Hardware requirements present a separate barrier. Real-time operation needs a dedicated GPU per classroom — an NVIDIA RTX 2060 or equivalent — at a cost of roughly $1,500–$2,000, which may be prohibitive for budget-constrained institutions. CPU-only configurations managed 3–4 frames per second, adequate for attendance but unreliable for behavioural monitoring; model optimisation approaches such as MobileNets [14] could help address this. The four-category emotion taxonomy also struggled at the boundary between confused and disengaged states, where facial cues overlap considerably [9]. Facial expressions are ultimately imperfect proxies for internal cognitive states, and automated engagement metrics should be interpreted with that limitation clearly in mind.

### 5.3. Future Development Directions

Several directions stand out for future work. Integrating additional modalities — body posture, gaze direction via tools such as OpenFace 2.0 [15], and gesture analysis — would help resolve the ambiguities that facial expressions alone cannot reliably distinguish, particularly at the boundary between confused and disengaged states. Incorporating temporal sequence models such as LSTMs [10] could allow the system to differentiate between fleeting expressions and sustained behavioural patterns. On the privacy side, federated learning [16] offers a credible path forward, allowing models to improve across diverse institutional populations without raw biometric data ever leaving local systems — which would simultaneously address the demographic representation gaps.

## Conclusion

This work presented an AI-based attendance and behavioural monitoring system that tackled two persistent classroom problems: time lost to manual roll-call, and the absence of any real-time signal about student engagement. Combining MTCNN-based face detection [1] with FaceNet recognition [2] and a custom learning-oriented emotion classifier produced 96.8% attendance accuracy and roughly 90% accuracy across four pedagogically relevant engagement states. The multi-task architecture [11] improved computational efficiency by approximately 40%, and storing facial embeddings rather than raw images [2] reduced biometric privacy exposure meaningfully, in line with data minimisation principles under GDPR [12]. Faculty responses were broadly positive, with several noting that starting sessions without a roll-call preserved instructional momentum in ways that felt more valuable than the time saving alone. Limitations remain and deserve honest acknowledgement. Accuracy dropped under poor lighting and partial occlusion, demographic bias auditing was incomplete due to limited population diversity, and the hardware cost of roughly $1,500–$2,000 per classroom may restrict adoption for budget-constrained institutions. Student feedback also drew a clear distinction between attendance tracking — which was widely accepted — and

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1462-1469
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0199

continuous behavioural monitoring, which raised privacy concerns that technical safeguards alone cannot fully resolve.

## Acknowledgment

The authors express sincere gratitude to Dr. A. M. Natarajan, Chief Executive; Dr. R. Devi Priya, Principal; and the faculty of the Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), KPR Institute of Engineering and Technology, for their guidance and institutional support. Special thanks are extended to the project supervisor, Mr. Anish Antony, for invaluable mentorship throughout this research. The authors also acknowledge the student participants involved in the pilot deployment, whose cooperation enabled real-world system validation.

## References

[1]. K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multi-task cascaded convolutional networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503, 2016.

[2]. F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815–823.

[3]. A. Ronacher, Flask: A lightweight WSGI web application framework, version 2.0, Pallets Projects, 2021. [Online]. Available: https://flask.palletsprojects.com

[4]. R. Patel, N. Shah, and A. Patel, "Fingerprint based attendance management system," International Journal of Computer Applications, vol. 179, no. 23, pp. 15–19, 2018.

[5]. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2001, pp. 511–518.

[6]. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1701–1708.

[7]. P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," Journal of Personality and Social Psychology, vol. 17, no. 2, pp. 124–129, 1971.

[8]. I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," Neural Networks, vol. 64, pp. 59–63, 2015.

[9]. A. 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, 2019.

[10]. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[11]. S. Li and W. Wang, "Multi-task learning for face recognition and emotion detection," in Proc. International Conference on Pattern Recognition, 2021, pp. 1245–1252.

[12]. European Parliament and Council, General Data Protection Regulation (GDPR), Official Journal of the European Union, 2016.

[13]. G. Bradski, "The OpenCV library," Dr. Dobb's Journal of Software Tools, 2000. [Online]. Available: https://opencv.org

[14]. A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[15]. T. Baltrušaitis, P. Robinson, and L.-P. Morency, "OpenFace 2.0: Facial behavior analysis toolkit," in Proc. IEEE International Conference on Automatic Face and Gesture Recognition, 2018, pp. 59–66.

[16]. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. International Conference on Artificial Intelligence and Statistics (AISTATS), 2017, pp. 1273–1282.