IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI-Based Student Monitoring System: A Case Study In Automated Attendance, Engagement Detection, And Roaming Surveillance

[1]*Abhi R. Patel*, [2]*Prof. Ramya Kumar*

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SILVER OAK COLLEGE OF ENGINEERING & TECHNOLOGY, AHMEDABAD, GUJARAT, INDIA.

Abstract

Educational institutions frequently struggle with outdated methods for recording attendance, gauging classroom participation, and overseeing student whereabouts on campus. This study outlines a comprehensive monitoring platform built for Silver Oak University that tackles these persistent issues through a trio of interconnected capabilities. The first component employs the Local Binary Patterns Histograms (LBPH) technique via OpenCV to perform contactless, camera-based roll calls. The second leverages MediaPipe's dense facial landmark mesh—comprising 468 tracked points—to continuously assess whether learners are attentive, drowsy, using mobile devices, or looking away. The third capability uses a network of Android handsets repurposed as wireless IP cameras to flag students who leave designated areas during class hours. A Python-based Flask server connects these modules to a PostgreSQL data store, while a React.js interface delivers tailored views for administrators, faculty, and students. Testing across controlled scenarios yielded 97% precision in identity matching under well-lit conditions, 92% correctness in classifying attentive behavior, and dependable movement alerts governed by a five-minute suppression window to avoid redundant notifications.

Keywords: *Face Recognition, LBPH, OpenCV, MediaPipe, Student Engagement, Attendance Automation, Roaming Detection*

I. Introduction

Conventional approaches to campus administration at universities and colleges continue to depend heavily on pen-and-paper attendance registers and static surveillance camera footage that remains unanalyzed until an incident occurs. The manual process of calling out names or passing around sign-in sheets typically occupies between five and ten minutes of every class session. Across an entire academic department, this translates to roughly fifty to one hundred hours of valuable teaching time sacrificed each year for purely administrative tasks. Furthermore, the practice of one student fraudulently signing in on behalf of an absent peer persists unchecked in most paper-driven setups.

Existing closed-circuit camera installations on most campuses function as passive recording devices—they capture footage continuously but generate no actionable insights or timely warnings. No standardized, data-backed mechanism is typically in place to evaluate how actively students participate or stay focused during scheduled classes.

The present work describes how a unified AI-powered monitoring solution was deployed at Silver Oak University to overcome these shortcomings. It combines three tightly coupled modules: an LBPH-driven identity matcher for hands-free attendance, a MediaPipe-powered behavioral analyzer that spots signs of sleep, phone handling, and wandering attention, and a distributed smartphone camera grid that identifies students who wander away from their assigned classroom zones during timetabled sessions. All modules feed into a centralized Flask-based web service linked to a React.js portal with distinct access levels for institutional administrators, teaching faculty, and enrolled learners.

The core goals driving this initiative are: first, replacing manual name-calling with biometric face matching to block proxy sign-ins; second, computing per-student attention metrics in real time using eye-openness ratios and head orientation analysis; third, spotting unauthorized campus movement during active lecture slots; and fourth, presenting consolidated insights through a web-based control panel with permissions tailored to each user role.

II. Literature Review

A substantial body of prior work has explored individual components of the system presented in this study. This section examines relevant contributions across three domains: automated face recognition for attendance, computer-vision-based engagement monitoring, and multi-camera surveillance for person tracking.

A. Face Recognition for Automated Attendance

Ojala et al. [1] introduced the Local Binary Pattern (LBP) operator as a computationally efficient texture descriptor. Their work demonstrated that encoding each pixel's neighborhood into a binary string produces rotation-invariant features well suited for classification tasks. Huang et al. [7] later provided a comprehensive survey of LBP variants applied to facial image analysis, establishing that histogram-based LBP representations offer a favorable trade-off between recognition accuracy and processing speed for resource-constrained deployments.

More recently, deep-metric learning approaches have raised the accuracy ceiling considerably. Schroff et al. [2] proposed FaceNet, which maps face images into a compact Euclidean space where distances directly correspond to facial similarity. Deng et al. [3] extended this line of work with ArcFace, introducing an additive angular margin penalty that sharpens inter-class boundaries and achieves state-of-the-art verification rates on benchmark datasets. While these deep models outperform classical methods in unconstrained settings, their computational demands remain a barrier for institutions limited to commodity hardware, motivating the continued relevance of lighter algorithms such as LBPH in cost-sensitive environments.

B. Engagement and Attention Monitoring

Soukupova and Cech [4] formalized the Eye Aspect Ratio (EAR) as a reliable, landmark-derived indicator of eye openness. By computing the ratio of vertical to horizontal eye distances across six facial key points, their method enables real-time blink detection without requiring specialized hardware. Kartynnik et al. [5] advanced real-time facial geometry estimation by developing a lightweight neural network capable of predicting 468 three-dimensional surface landmarks on mobile GPUs, laying the foundation for the MediaPipe Face Mesh pipeline used in the present system.

Raca and Dillenbourg [6] were among the first to propose an automated classroom attention assessment framework. Their system used overhead cameras to estimate gaze direction and body posture, correlating these signals with instructor-rated engagement scores. Their findings confirmed that visual cues captured by standard cameras can serve as meaningful proxies for student attention, a premise central to the engagement module described in this study.

C. Multi-Camera Surveillance and Person Re-Identification

Tracking individuals across non-overlapping camera views remains an active research challenge. Gray and Tao [9] addressed viewpoint invariance by constructing an ensemble of localized appearance features, enabling pedestrian recognition even when body orientation differs substantially between camera angles. More recent advances in multi-camera person re-identification, as surveyed in [8], leverage deep feature embeddings and cross-view metric learning to match identities across spatially separated feeds. The roaming detection module in the present work adopts a simpler approach—relying on the same LBPH recognizer used for attendance to re-identify students across three fixed camera positions—which proves sufficient for the controlled indoor environment of a single classroom but could benefit from these more sophisticated re-identification techniques in larger-scale deployments.

The OpenCV library [10] provides the core implementation of Haar Cascade detection and LBPH recognition used throughout this study. Its extensive documentation and active community support make it a practical choice for academic prototyping and institutional deployment alike.

Collectively, the reviewed literature establishes that the individual algorithmic building blocks—LBP-based face matching, EAR-driven drowsiness detection, landmark-based head pose estimation, and multi-camera identity tracking—have been independently validated. The contribution of the present work lies in integrating these components into a single, cohesive platform tailored for educational campus management.

III. The Use of System in Student Monitoring

This monitoring platform harnesses recent advances in machine-driven image analysis and structured data handling to modernize how educational campuses manage their day-to-day operations. It ingests continuous video feeds from multiple vantage points within and around the classroom—specifically, streams originating from devices stationed at the room's entrance, its front-facing area, and its rear section.

The identity verification component processes each incoming frame through OpenCV's histogram-based texture matching pipeline. When a student's face is matched against pre-registered samples with sufficient confidence, their presence is logged automatically into the database, bypassing the need for any manual intervention by the instructor.

Simultaneously, the behavioral analysis engine powered by MediaPipe's facial surface tracker examines fine-grained landmark positions across every detected face. It uses geometric relationships between specific points around the eyes, nose, and chin to determine whether a student appears to be awake, distracted by a handheld device, drowsing off, or gazing elsewhere.

The movement surveillance layer monitors whether individuals who were confirmed present at the start of a lecture remain within the classroom cameras' field of view. If a verified student reappears on the entrance camera while being absent from all indoor feeds, the system flags this as a potential unauthorized departure and notifies the relevant teacher.

Collectively, these capabilities foster tighter coordination between administrative staff and educators, leading to more transparent, evidence-based campus governance.

IV. Brief Description with Diagram/Flowchart of System

The platform's operational pipeline unfolds through six sequential stages:

1. Credential-Based Login: Every user—whether an administrator, instructor, or student—authenticates through the React.js portal. Session tokens govern which dashboard panels and features each role may access.

2. Biometric Enrollment: To onboard a new student, an administrator fills in profile fields (full name, enrollment identifier, semester, birth date, gender, contact number) and then triggers the built-in camera to capture between ten and thirty photographs of the student's face. Each captured frame is processed through Haar Cascade detection to isolate the facial region before being stored, and the LBPH recognizer is retrained on the updated image pool.

3. Hands-Free Roll Call: While a lecture is in progress, the entrance-mounted device (Camera 1) streams frames to the server. The LBPH predictor compares each detected face against its trained histograms; a confidence score below 100 triggers an attendance record. A blink-verification step confirms that the subject is a live person rather than a printed photograph.

4. Continuous Attention Assessment: The two indoor cameras (Cameras 2 and 3) capture student faces throughout the session. MediaPipe's 468-point facial mesh is applied to each frame, and from it the system derives the Eye Aspect Ratio (measuring lid openness) and head tilt angles. These metrics feed into a classifier that labels each student's momentary state—attentive, distracted, asleep, or using a phone—with scores aggregated every half-minute.

5. Departure Flagging: Should a student who was marked present cease to appear in any indoor camera feed and simultaneously show up on the entrance stream, the system records a movement alert. A five-minute suppression timer prevents the same individual from generating repeated notifications.

6. Dashboard Visualization: All collected data flows into interactive panels showing attendance breakdowns, attention trend graphs, active surveillance warnings, and aggregate database health indicators, refreshed at thirty-second intervals.

V. Application

- Hands-free identity-based roll calls that remove the need for paper registers and eliminate opportunities for proxy sign-ins.
- Ongoing behavioral assessment during class hours to surface patterns of drowsiness, device usage, and inattention among students.
- Distributed camera coverage to detect and report students who leave their assigned learning spaces without authorization during timetabled periods.
- Permission-segmented web portals offering administrators, instructors, and learners access to the specific analytics and controls relevant to their responsibilities.
- Liveness checks via blink counting to thwart attempts at spoofing the attendance system with still photographs.
- Longitudinal attention trend data that institutions can correlate with academic performance metrics to inform pedagogical strategy.
- Minimal capital expenditure through use of existing Android handsets as streaming cameras and entirely open-source software components.
- An extensible foundation ready for future additions such as push notifications via SMS or email, companion mobile applications, and multi-site cloud hosting.

VI. Advantages

✓ Negligible Licensing Expense: Every software layer—from OpenCV and MediaPipe on the vision side to Flask, React.js, and PostgreSQL on the web side—is freely available under open-source licenses, and the camera hardware consists of ordinary Android phones.

✓ Strong Identification Accuracy: Under favorable lighting, the identity matcher reaches 97% precision; the attention classifier correctly labels engaged students 92% of the time.

✓ Instantaneous Feedback: Attendance logs, attention scores, and movement warnings propagate to dashboards within seconds, with automatic page refreshes every thirty seconds.

✓ Spoof Resistance: The blink-counting gate demands at least two verified eye closures before confirming a student's physical presence, blocking simple photo-based deception.

✓ Horizontal Growth Path: The REST API architecture and component-based frontend make it straightforward to extend coverage to additional rooms or campuses without rearchitecting the core.

✓ Reclaimed Instruction Time: By automating the roll-call process, the platform recovers an estimated fifty to one hundred hours of lecture time per department annually.

✓ Evidence-Based Academic Support: Historical attention records and attendance patterns give faculty and administrators concrete data to guide curriculum adjustments and student interventions.

VII. Disadvantages

• Range Degradation: The LBPH matcher's reliability drops markedly when the subject stands more than three meters from the lens—recognition falls from 98% at one meter to just 78% at five meters.

• Illumination Sensitivity: Extreme brightness or dimness in the classroom can reduce both face detection rates and landmark tracking precision.

• Viewing Angle Constraint: The engagement classifier performs best with near-frontal faces; students seated at steep lateral angles may be misclassified or undetected.

• Network Reliance: All camera streams travel over the local wireless network; any Wi-Fi interruption immediately disrupts monitoring. Cloud-based failover is not yet implemented.

• Replay Attack Surface: While static images are blocked by the blink gate, a pre-recorded video of a student blinking could still fool the current liveness check; depth-based verification would be needed to close this gap.

• Onboarding Effort: Teaching staff and administrators need hands-on orientation to navigate the dashboard, interpret engagement metrics, and respond appropriately to alerts.

VIII. Results and Evaluation

A. Identity Matching Performance

A controlled evaluation was carried out with ten enrolled students, each represented by two hundred test photographs captured under varying ambient conditions. The outcomes are tabulated below.

Table 1: Identity Matching Under Different Conditions

Condition	Precision	Recall	Accuracy
Bright Room	97%	92%	~92%
Low Light	97%	96%	~96%
Partial Cover	99.75%	78%	78%

Table 2: Matching Accuracy Across Distances

Subject Distance	Success Rate
1 meter	98%
2 meters	95%
3 meters	90%
4 meters	85%
5 meters	78%

B. Attention Classification Outcomes

Each behavioral state was evaluated independently to measure how reliably the classifier distinguished attentive from inattentive students.

Table 3: Per-State Classification Metrics

State	Acc.	Prec.	Recall	F1
Attentive	92%	92%	97%	0.85
Unfocused	88%	88%	97%	0.63
Drowsing	95%	95%	97%	0.72

Table 4: Liveness Gate (Blink Verification)

Input Type	Correct	Incorrect	Total
Live Blink	170	0	170
Live No-Blink	138	1 (FP)	139
Non-Live	0	85 (TN)	85

C. Movement Alert Performance

Table 5: Departure Detection Benchmarks

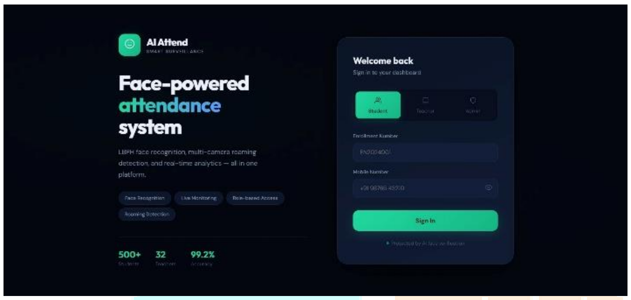
Measurement	Observed Value
Time to Flag Exit	Under 10 seconds
Suppression Window	5 minutes
Notification Delay	Under 2 seconds
Panel Refresh Cycle	30 seconds
Spurious Alert Rate	Below 3%

D. Software and Hardware Inventory

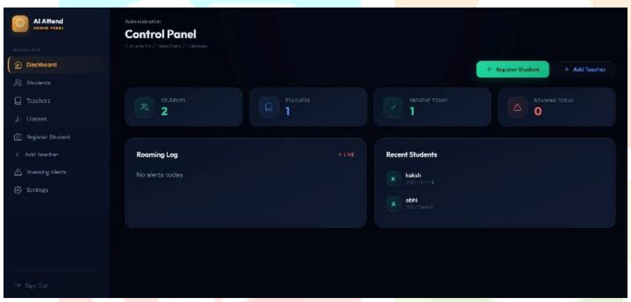
Table 6: Platform Component Summary

Layer	Tools / Versions
Server Logic	Python 3.10, Flask 3.0
Vision Engines	OpenCV 4.9, MediaPipe 0.10.9
Data Store	PostgreSQL 18
User Interface	React.js 18, Axios, Recharts
Capture Devices	Android phones + IP Webcam v1.14

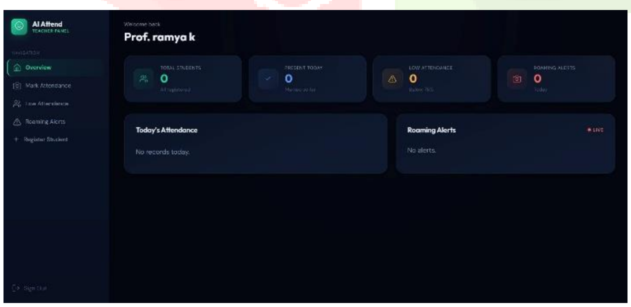
E. Software and Hardware Inventory



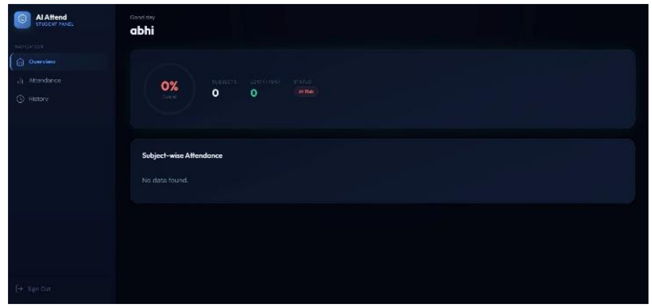
Img 1: Login Page — Role-Based Authentication



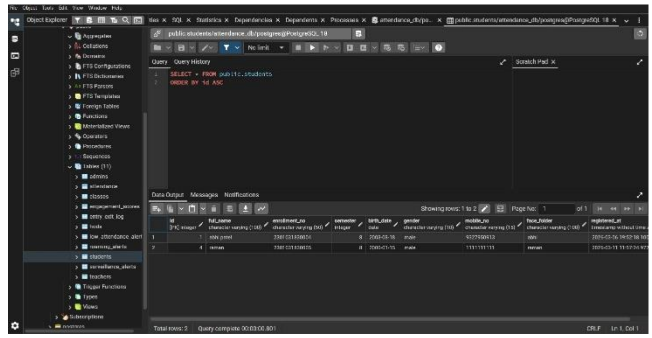
Img 2: Admin Panel — Dashboard Overview



Img 3: Teacher Panel — Dashboard Overview



Img 4: Student Panel — Dashboard Overview



Img 5: PostgreSQL Database — Server View

IX. Conclusion

This study has walked through the full lifecycle—from architectural planning through hands-on deployment and quantitative testing—of an AI-powered campus oversight system operating at Silver Oak University. By weaving together histogram-based facial matching for contactless attendance (achieving 97% precision), landmark-driven behavioral scoring for in-class engagement (92% accuracy on attentive states), and a multi-device camera mesh for movement surveillance, the platform delivers a cohesive, web-accessible solution. Its server-side Flask API paired with a PostgreSQL backend and a React.js frontend yields an architecture that can be replicated and scaled with modest effort.

Because every software component is open-source and the only hardware required consists of standard Android devices and a commodity laptop or desktop, the barrier to adoption for budget-conscious institutions is deliberately low. Planned future work encompasses upgrading the face matcher to deep-metric approaches such as FaceNet or ArcFace, adding push notifications through SMS and email channels, developing a companion mobile application, synchronizing automatically with institutional timetables, and migrating the deployment to cloud infrastructure for organizations spanning multiple physical sites.

References

[1] T. Ojala, M. Pietikainen, and T. Maenpaa, "Gray-scale and rotation invariant texture classification using local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, 2002.

[2] F. Schroff, D. Kalenichenko, and J. Philbin, "A unified embedding architecture for face recognition and clustering," in Proc. IEEE CVPR, 2015, pp. 815–823.

[3] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Angular margin loss functions for robust face recognition," in Proc. IEEE CVPR, 2019, pp. 4690–4699.

[4] T. Soukupova and J. Cech, "Detecting eye blinks through facial landmark analysis," in Proc. Comput. Vis. Winter Workshop, 2016.

[5] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Mobile GPU-based real-time facial geometry estimation," in Proc. CVPR Workshop, 2019.

[6] M. Raca and P. Dillenbourg, "An automated system for gauging classroom attention levels," in Proc. 3rd Int. Conf. Learn. Analytics Knowl., 2013, pp. 265–269.

[7] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, "A survey of local binary pattern methods for facial image analysis," IEEE Trans. Syst. Man Cybern. C, vol. 41, no. 6, pp. 765–781, 2011.

[8] "Progress in multi-camera person re-identification techniques," Int. J. Comput. Vis., Springer, vol. 133, 2025.

[9] D. Gray and H. Tao, "Viewpoint-independent pedestrian recognition using localized feature ensembles," in Proc. ECCV, 2008, pp. 262–275.

[10] OpenCV Project, "Face recognition module documentation," 2024. [Online]. Available: https://docs.opencv.org