International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

SAGE: Student Attention and Gaze Evaluator for Smart Classroom Monitoring

1. Ms. Priyanka Chopkar, 2. Mr. Manish Patle, 3. Mr. Mihir Deshmukh, 4. Mr. Nachiket Lokhande, 5. Mr. Nishant Bhure, 6. Mr. Piyush Pawade, 7. Ms. Purva Rewatkar

Department of Electronics and Telecommunication Engineering

1 Assistant Professor, Electronics and Telecommunication Engineering, G. H. Raisoni College of Engineering and Management, Nagpur, India

2–7Students, Electronics and Telecommunication Engineering, G. H. Raisoni College of Engineering and Management, Nagpur, India

Abstract - We developed SAGE (Student Attention and Gaze Evaluator), a computer-vision prototype that combines face-recognition attendance, visual behavioural cues, mobile-phone detection, and a local instructor dashboard. We implemented the workflow on a Raspberry Pi 4 with 4 GB RAM. OpenCV and MediaPipe support image and facial-landmark processing, ONNX Runtime is used for face recognition/re-identification, TensorFlow Lite is used for mobile-phone detection, Flask serves the web application, and SQLite stores local records. In the current project materials, the system is reported to process approximately 10–15 frames per second, and selected functions including camera reconnection, face matching, offline dashboard operation, and attendance reconciliation were checked. These checks show that the functions have been integrated, but they do not yet establish model accuracy or educational benefit. The visual outputs are treated as behavioural indicators, not direct measurements of cognition.

Index Terms - Student attention, gaze estimation, computer vision, smart classroom, face recognition, mobile-phone detection, Raspberry Pi, edge AI.

I. INTRODUCTION

We designed SAGE to address a practical classroom problem: instructors may need to record attendance while also monitoring classroom activity and possible device-related distractions. Manual roll calls take time, while informal observation can be inconsistent. Camera-based tools can automate some routine tasks, but their outputs depend on image quality, camera placement, lighting, occlusion, and model design. Gaze direction and head pose are observable behaviours; neither alone establishes whether a student is listening, understanding, or learning.

Prior work has explored sensor-based attention estimation, facial analysis, face recognition for attendance, and camera-based classroom behaviour analysis. Deshpande et al. describe a system combining electronic-sensor information, facial analysis, and attendance processing [1], while Patil et al. report an AI-based attention and alertness monitoring approach [2]. Facial landmarks and eye-related measurements have also been studied for drowsiness detection [3]. Face-recognition attendance systems and Raspberry Pi-based attendance approaches provide context for the attendance component [4], [5]. Classroom behaviour detection and student-engagement studies using object detection or deep-learning methods are also relevant [6]–[9]. These studies address different tasks, so their results should not be compared directly unless their labels, datasets, test conditions, and metrics are sufficiently similar.

Our contribution is an integrated prototype rather than a new attention-estimation algorithm. We combined attendance management, visual behavioural cues, mobile-phone detection, and local dashboard reporting on a resource-constrained edge platform. The available project materials do not include a labelled classroom dataset, independent ground-truth annotations for attention, or formal model-level evaluation metrics. We therefore report the implementation and functional checks completed so far and identify model validation as future work.

II. PROBLEM STATEMENT AND OBJECTIVES

The project addresses the time and effort involved in recording attendance while observing classroom activity and possible mobile-device distractions. Manual roll calls interrupt instruction, while informal observation may be inconsistent and difficult to review. SAGE is intended to provide an integrated workflow for attendance records and visible behavioural cues. It is not designed to determine a student's comprehension, motivation, or academic ability.

We set five objectives: (1) automate attendance using face recognition; (2) derive attention-related behavioural indicators from facial landmarks, gaze-related features, and head orientation; (3) detect visible mobile phones as a possible indicator of device-related distraction; (4) provide live system information through an instructor dashboard; and (5) store attendance and session records locally for later review. The current evaluation covers implementation checks; validated accuracy has not yet been established.

III. RELATED WORK

Sensor-based approaches estimate aspects of attention from physiological signals, whereas camera-based approaches use facial or behavioural cues. Deshpande et al. describe a system combining electronic sensors, facial analysis, and attendance processing [1]. Patil et al. present an AI-based attention and alertness monitoring system [2]. These studies illustrate different sensing and analysis strategies. The papers available for this review do not provide a common benchmark for a direct performance comparison; such a comparison would require compatible labels, participant populations, test conditions, and metrics.

Facial landmarks and eye-related measurements have been used in drowsiness-related applications. Sawant et al.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

describe a MediaPipe- and eye-aspect-ratio-based driver-drowsiness approach [3]. Face recognition for attendance has also been discussed in smart-classroom settings [4], and Raspberry Pi-based attendance systems have been reviewed [5]. Related classroom behaviour and student-engagement studies use object-detection or deep-learning methods [6]–[9]. These works inform the design of SAGE, but results from driver drowsiness, attendance, and classroom engagement are not interchangeable because the tasks and target labels differ. Based on this literature, we focused on whether attendance management, visual behavioural cues, phone detection, and dashboard reporting could be combined on a local edge device. The work reported here describes that integrated prototype and its functional checks. We have not established that SAGE is more accurate or effective than existing systems; that would require a controlled comparison using a common evaluation protocol.

IV. SYSTEM ARCHITECTURE

We use a camera as the primary input and a Raspberry Pi 4 as the local processing platform. The pipeline covers frame acquisition, visual inference, identity and attendance management, local data storage, and dashboard presentation. OpenCV handles frame acquisition and processing; MediaPipe provides face detection and facial landmarks; ONNX Runtime supports face recognition/re-identification; and TensorFlow Lite supports mobile-phone detection. The Python/Flask application serves the dashboard, Waitress serves the application, and SQLite stores attendance and session records. The workflow is designed to run locally. We have not conducted a formal security assessment or reported a quantified end-to-end latency measurement.

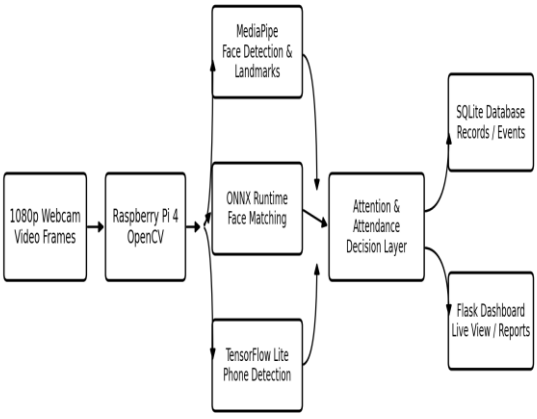

Fig. 1. SAGE system block diagram.

V. METHODOLOGY

A. Video Acquisition and Face Processing

We acquire camera frames and process them with OpenCV. MediaPipe face detection and facial landmarks are used to locate facial regions. The recognition workflow includes face alignment and a minimum face-size constraint. We also checked camera reconnection and safe shutdown. The current paper does not list all preprocessing parameters, camera capture settings, software versions, or model configurations, so the pipeline cannot yet be reproduced exactly from the manuscript alone.

B. Attention-Related Visual Cues

We use gaze-related features, eye movement, facial landmarks, and head orientation as observable behavioural cues. Looking away from the camera may indicate a change in gaze direction, but it does not show that a student is not listening. The project materials report four attention-related output states, but do not fully define the states or provide validated thresholds and independent ground-truth annotations. We therefore treat these outputs as heuristic behavioural indicators, not as a validated classifier of cognitive attention.

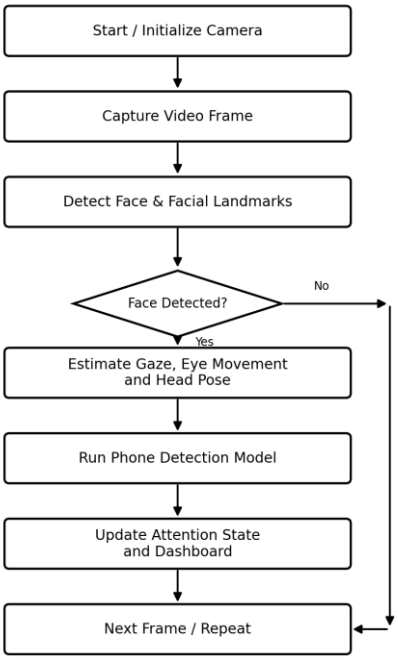

Fig. 2. Flowchart for attention-related cue estimation.

C. Face Recognition and Attendance

We implemented face recognition/re-identification using ONNX Runtime. Detected faces are matched against stored student representations to support attendance marking. Attendance records are stored in SQLite and can be reconciled and exported. We checked face matching functionally and reconciled attendance records during testing. We have not reported recognition accuracy, false-acceptance or false-rejection rates, or performance across lighting, seating distance, occlusion, and class size.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

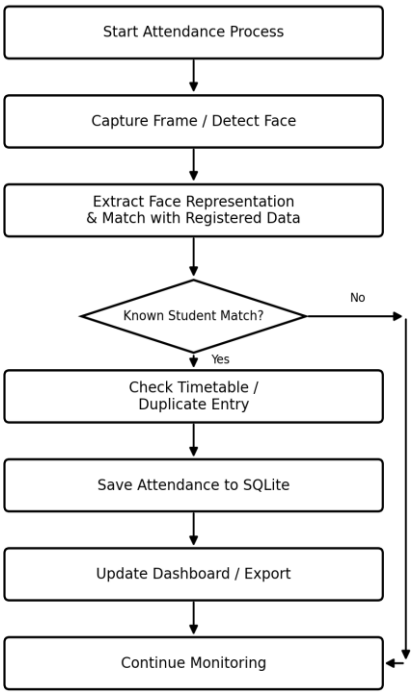

Fig. 3. Face matching and attendance recording workflow.

D. Mobile-Phone Detection

We use a TensorFlow Lite model to detect mobile phones in camera frames. We treat phone detection as a separate visual signal that may alert an instructor to possible device-related distraction. The current project materials do not identify the model architecture, training dataset, confidence threshold, or test-set precision and recall. These details are needed to assess detection performance independently.

E. Backend, Storage, and Dashboard

Our software stack includes Python, Flask, Waitress, PyYAML, threading, and SQLite. The dashboard provides a live camera stream and system updates, with JavaScript and charting components for presentation. Project testing covered offline operation, automated lecture start/stop, timetable integration, crash recovery, attendance export, and automated tests, as documented in the project materials. SQLite provides local persistence and reduces dependence on continuous cloud connectivity, but local storage alone does not establish that the system is secure or privacy-compliant.

VI. IMPLEMENTATION AND TEST SETUP

Our reported hardware setup consists of a Raspberry Pi 4 with 4 GB RAM, a power adapter, and a 1080p Logitech webcam. The estimate does not state whether storage media, cables, an enclosure, peripherals, maintenance, or other deployment costs are included. The software components include OpenCV, MediaPipe, ONNX Runtime, TensorFlow Lite, Flask, Waitress, and SQLite. The operational loop acquires frames, runs visual inference, updates local records, and serves the dashboard. Camera reconnection and safe-shutdown checks were included in implementation testing.

Prototype Setup

The SAGE prototype is assembled around a Raspberry Pi 4 single-board computer, with a Logitech webcam mounted above it to provide the live video input. The camera connects to the processing unit through USB, while the Raspberry Pi serves as the local platform for image processing, attendance management, and the instructor dashboard. The compact arrangement is suitable for a proof-of-concept demonstration and allows the main components to be carried and set up without a dedicated workstation. Figure 4 shows the physical prototype used for the project. The photograph documents the hardware arrangement only; it does not by itself establish recognition accuracy, attention-estimation performance, or classroom-scale reliability.



Fig. 4. Physical prototype of SAGE using a Logitech webcam and Raspberry Pi 4.

We organised the project into five phases: (1) database and core setup; (2) face detection and attention-related cues; (3) face recognition; (4) dashboard development; and (5) automation and attendance. The project presentation marks these phases as completed and records functional checks including database stress testing, camera reconnection, safe shutdown, face matching, offline dashboard operation, attendance reconciliation, and automated tests. However, the manuscript does not give a formal test protocol, trial counts, participant count, labelled test set, or independent ground-truth attention annotations. The reported checks should therefore be understood as functional verification rather than evidence of model accuracy.

VII. RESULTS AND DISCUSSION

In the implemented setup, project documentation reports a processing rate of approximately 10–15 frames per second and four attention-related output states. It also records stable operation, face-matching checks, camera auto-reconnection, offline dashboard operation, multiple-user testing, attendance reconciliation, and automated tests. These are implementation and functional-verification outcomes, not a full evaluation of model accuracy. The documentation does not specify a repeatable measurement protocol, test duration, number of trials, end-to-end latency, or resource utilisation for the frame-rate result.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

TABLE I. SUMMARY OF IMPLEMENTATION RESULTS REPORTED IN THE PROJECT DOCUMENTATION.

Item	Reported outcome
Processing speed	Approximately 10–15 FPS reported; measurement protocol and variability not specified
Attention-state output	Four states reported; definitions, thresholds, and validation not fully specified
Face recognition	Functional matching check reported; no accuracy or error rates reported
System reliability checks	Camera reconnection and safe-shutdown checks reported; trial counts not specified
Dashboard	Offline operation and multiple-user testing reported; protocol and scale not specified
Attendance workflow	Reconciliation, export, and automated tests reported; error rates not reported

The reported frame rate indicates that the edge platform can run the combined workflow at a modest speed under the project's test conditions, although those conditions are not described in enough detail to reproduce the measurement. Suitability for a classroom will depend on camera resolution, the number of visible faces, model complexity, and simultaneous inference tasks. We have not measured attention-estimation accuracy, phone-detection accuracy, face-recognition accuracy, or any reduction in attendance-taking time. Those outcomes require labelled data, task-specific metrics, and a repeatable test protocol.

An important limitation concerns how the attention-related outputs are interpreted. Gaze direction and head pose describe visible behaviour, not a student's full cognitive state. A student may look away while thinking or taking notes, and a forward-facing student may still be distracted. The dashboard should present these outputs as uncertain indicators for instructor review, not as definitive judgements or the sole basis for grading or disciplinary action.

VIII. LIMITATIONS AND ETHICAL CONSIDERATIONS

We have not yet evaluated SAGE using a classroom-scale benchmark or participant-based validation study. The current manuscript contains no confusion matrices or precision/recall measurements, and the range of tested classroom conditions is not specified. These conditions include lighting, seating distance, camera angle, occlusion, and demographic variation. Until these evaluations are completed, our performance claims are limited to the reported implementation checks and approximate frame rate. A reproducible study should define the target labels, data collection and annotation procedure, train/validation/test separation where applicable, test conditions, and evaluation metrics.

Because SAGE processes identifiable faces and behavioural observations, deployment should include appropriate notice and consent or another valid institutional basis, access controls, minimal retention, secure storage, and a way to correct attendance errors. Raw video should be retained only when necessary, authorised, and covered by a defined retention policy. Attention estimates should not be treated as direct measures of comprehension, effort, or academic ability. Before deployment, we need to document any applicable institutional review, evaluate error rates across users and classroom conditions, and clearly distinguish safeguards already implemented from those still proposed. The current project materials do not establish that all these governance measures are in place.

IX. SOFTWARE INTERFACE AND LIVE MONITORING RESULTS

The SAGE interface provides a login page and a home dashboard for accessing the main system functions. The following screenshots are taken from the project interface and show the implemented screens.

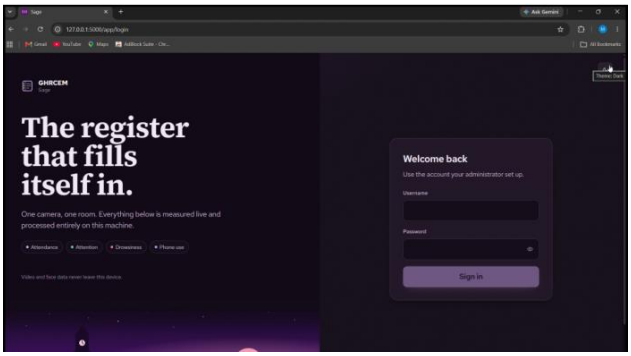

Fig. 5. SAGE login screen for accessing the classroom monitoring application.

After login, the home dashboard provides access to live monitoring, student records, attendance, and session management.

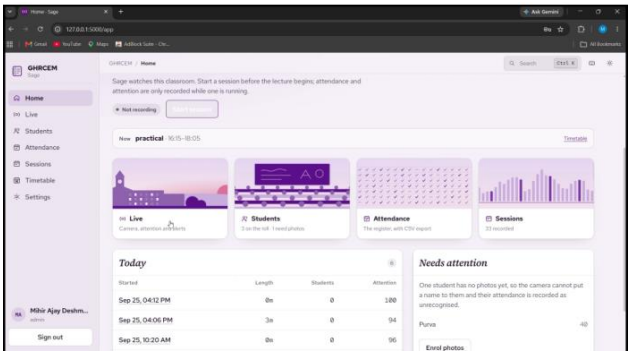

Fig. 6. SAGE home dashboard showing the main monitoring and management modules.

The live monitoring screen presents the camera feed and visual-analysis outputs generated by the running pipeline. These outputs are intended to support instructor review and should not be treated as definitive measurements of a student's internal attention.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

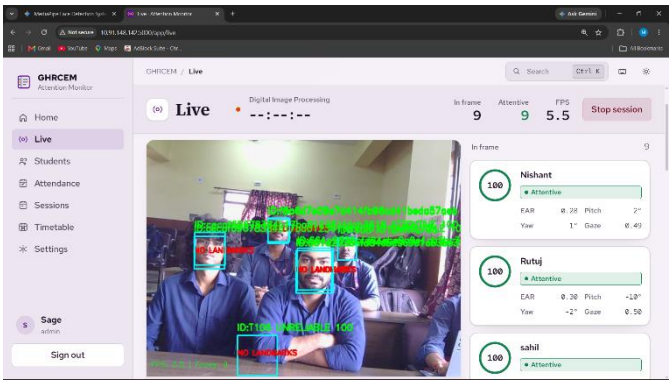

Fig. 7. SAGE live monitoring screen showing the camera feed and analysis indicators.

The screenshots document the implemented interface and integration of the main software modules. They do not, by themselves, establish attention-estimation accuracy, phone-detection accuracy, or face-recognition accuracy; quantitative claims remain limited to the experiments reported in this paper.

X. CONCLUSION AND FUTURE WORK

We developed SAGE as a Raspberry Pi 4-based prototype combining face-recognition attendance, facial-landmark and gaze-related processing, head-pose analysis, mobile-phone detection, local storage, and an instructor dashboard. Project documentation reports approximately 10–15 FPS and completion of selected functional checks. These results show that the functions have been integrated into a local workflow. They do not yet establish attention-estimation accuracy, phone-detection accuracy, face-recognition performance across conditions, or educational benefit.

Our next step is to define each attention-related output state and evaluate the system using a labelled classroom dataset collected and handled under an appropriate governance process. We should report per-class precision, recall, F1-score, confusion matrices, face-recognition error rates, phone-detection performance, end-to-end latency, and resource utilisation. Testing should include varied lighting, camera angles, seating distances, occlusion, and simultaneous faces, supported by a transparent annotation protocol and privacy safeguards. A controlled user study is also needed to assess whether the dashboard supports instructors without encouraging undue surveillance or misinterpretation of student behaviour.

REFERENCES

[1] M. J. Deshpande, A. Atterkar, D. Gangarde, P. Ghadge, V. Golhar, and A. Yenkikar, “Development of Student Classroom Attention System Using Artificial Intelligence and Electronic Sensors,” Cureus Journal of Computer Science, vol. 2, Art. no. es44389-024-02400-y, Apr. 2025, doi: 10.7759/s44389-024-02400-y.

[2] S. Patil, D. Pathak, A. Jani, D. Jondhale, and A. S. Savyanavar, “Student’s Attention and Alertness Monitoring System using AI,” in Proc. 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), Pune, India, Apr. 2024, doi: 10.1109/MITADTSoCiCon60330.2024.10575441.

[3] A. G. Sawant et al., “A Real-Time Driver Drowsiness Detection System Using MediaPipe and Eye Aspect Ratio,” arXiv preprint arXiv:2511.13618, 2025.

[4] A. S. Lateef and M. Y. Kamil, “Facial Recognition Technology-Based Attendance Management System Application in Smart Classroom,” Iraqi Journal for Computer Science and Mathematics, vol. 4, no. 3, Art. no. 12, 2023, doi: 10.52866/ijcsm.2023.02.03.012.

[5] K. R. Krishna, K. J. Chandu, N. Srikanth, K. S. Niteesh, and Y. K. Reddy, “A Review of Smart Attendance Monitoring System using Raspberry Pi,” International Journal of Engineering Research & Technology, vol. 12, no. 3, Mar. 2023. [Online]. Available: https://www.ijert.org/research/a-review-of-smart-attendance-monitoring-system-using-raspberry-pi-IJERTV12IS030012.pdf.

[6] L. Tang, T. Xie, Y. Yang, and H. Wang, “Classroom Behavior Detection Based on Improved YOLOv5 Algorithm Combining Multi-Scale Feature Fusion and Attention Mechanism,” Applied Sciences, vol. 12, no. 13, Art. no. 6790, 2022, doi: 10.3390/app12136790.

[7] N. Mahmood, S. M. Bhatti, H. Dawood, M. R. Pradhan, and H. Ahmad, “Measuring Student Engagement through Behavioral and Emotional Features Using Deep-Learning Models,” Algorithms, vol. 17, no. 10, Art. no. 458, 2024, doi: 10.3390/a17100458.

[8] J. Hu, Z. Huang, J. Li, L. Xu, and Y. Zou, “Real-Time Classroom Behavior Analysis for Enhanced Engineering Education: An AI-Assisted Approach,” International Journal of Computational Intelligence Systems, vol. 17, 2024, doi: 10.1007/s44196-024-00572-y.

[9] N. Alruwais and M. Zakariah, “Student-Engagement Detection in Classroom Using Machine Learning Algorithm,” Electronics, vol. 12, no. 3, Art. no. 731, 2023, doi: 10.3390/electronics12030731.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)