iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79897

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Real-Time Classroom Engagement Detection System Using OpenCV and MediaPipe-Integrated Head Pose Estimation

Mrs. G Phuspa Antanet Sheeba[1], Mohammed Farhan I[2], Mohammed Hassan A[3], Mafaaz Ahmed S[4]
[1]*Assistant Professor, Department of Computer Science and Engineering, CAHCET*
[2, 3, 4]*Department of Computer Science and Engineering, CAHCET*

Abstract: ***Traditional classroom monitoring relies on manual observation, which is subjective, inconsistent, and impractical for large class sizes. This paper presents EduVision, a real-time student attention monitoring system that leverages computer vision and facial landmark analysis to objectively quantify classroom engagement. The proposed system utilizes MediaPipe Face Mesh to extract 468 facial landmarks per detected face and computes a weighted attention score based on three geometric parameters — yaw, pitch, and face visibility — with weights of 0.5, 0.3, and 0.2 respectively. A student is classified as attentive when the computed score meets or exceeds a configurable threshold of 0.65, with snapshots captured every five seconds maintaining a timestamped engagement log. Deployed with dual interface support comprising a standalone OpenCV monitoring window and a Streamlit web dashboard, EduVision provides a non-intrusive, cost-effective, and automated alternative to manual engagement tracking, requiring only a standard webcam.***

I. INTRODUCTION

The rapid advancement of artificial intelligence and computer vision technologies has opened new possibilities for transforming traditional educational environments into intelligent, data-driven smart classrooms.

Monitoring student attention during lectures is a critical factor in evaluating teaching effectiveness and improving learning outcomes, yet conventional approaches depend entirely on the subjective judgment of educators, which is both time-consuming and inconsistent across large classroom settings.

Existing automated solutions often require expensive hardware such as eye-tracking devices or wearable sensors, making them impractical for widespread deployment in resource-constrained institutions. EduVision addresses these limitations by proposing a lightweight, webcam-based attention monitoring framework that leverages MediaPipe Face Mesh and OpenCV to perform real-time facial landmark extraction and head pose estimation without any specialized equipment.

The system computes a multi-metric attention score for each detected student and continuously logs engagement data throughout the class session, providing educators with objective, timestamped analytics through both a live monitoring interface and a post-session web dashboard:

1) Non-Intrusive Monitoring: Operates via standard webcam without wearables, specialized hardware, or student interaction.
2) Multi-Metric Attention Scoring: Combines yaw, pitch, and visibility into a single weighted geometric attention score.
3) Dual Interface Architecture: Provides live OpenCV HUD and Streamlit web dashboard for real-time analytics.
4) Automated Session Reporting: Saves timestamped CSV and JSON logs automatically for post-session performance analysis.

II. LITERATURE REVIEW

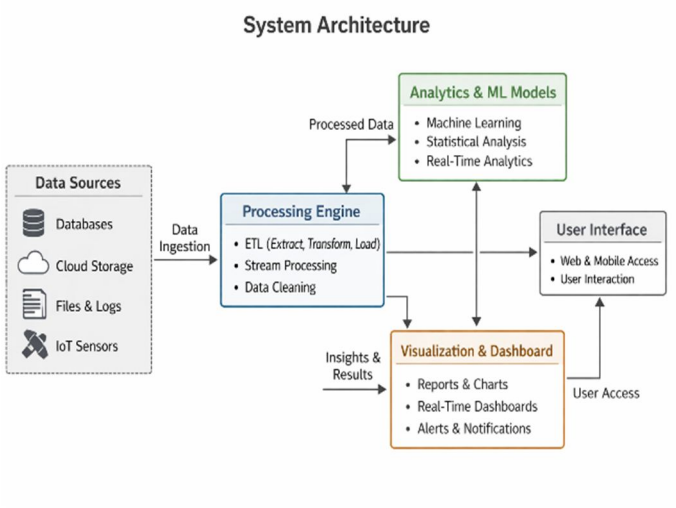
Romano et al. (2021) proposed an eye-tracking-based student engagement detection system using infrared sensors and gaze estimation algorithms. While achieving high accuracy, the approach required specialized hardware installations, limiting its scalability and practical adoption in standard classroom environments with budget constraints.

Romano et al. (2021) proposed an eye-tracking-based student engagement detection system using infrared sensors and gaze estimation algorithms. While achieving high accuracy, the approach required specialized hardware installations, limiting its scalability and practical adoption in standard classroom environments with budget constraints.

Kamath et al. (2016) developed a vision-based attention monitoring system using head pose estimation from RGB camera feeds in lecture hall settings. The system demonstrated promising results in controlled environments but struggled with varying illumination conditions, partial occlusions, and large class sizes exceeding thirty students simultaneously.

III. SYSTEM ARCHITECTURE

EduVision follows a layered client-server inspired architecture where the webcam feed serves as the primary input, processed through a computer vision pipeline that extracts facial landmarks, computes attention scores, and delivers results through dual interface outputs. The system is entirely local with no cloud dependency.



A. *Frontend*

The frontend consists of two interfaces. The primary interface is a Streamlit web dashboard rendered in the browser, featuring real-time Plotly line charts, donut charts, and minute-by-minute bar graphs. The secondary interface is an OpenCV window displaying a live HUD with color-coded bounding boxes and a side statistics panel**.**

B. *Backend Processing Pipeline*

The backend is a Python-based processing engine comprising three core modules. The Face Analyzer class handles MediaPipe Face Mesh inference, extracting 468 landmarks per face. The Attention Tracker class manages session state, snapshot recording, and file persistence. The HUD Renderer class handles all real-time visual overlay rendering onto the OpenCV video frames.
This system does not use a traditional database. Session data is persisted locally as CSV and JSON flat files inside the session_logs directory, which are read by the report viewer and Streamlit dashboard for post-session analytics.

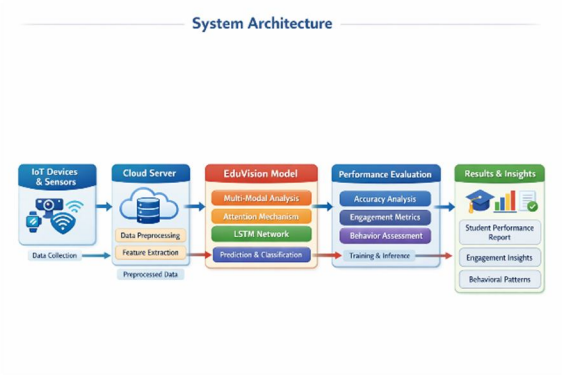
IV. METHODOLOGY AND MODEL ARCHITECTURE

A. *Methodology*

The EduVision system captures live video via a standard webcam at 1280×720 resolution using OpenCV, processing every alternate frame for efficiency. Each frame is passed to MediaPipe Face Mesh, which extracts 468 facial landmarks per detected face. Three geometric parameters — yaw, pitch, and visibility — are computed and combined into a weighted attention score. Faces scoring above 0.65 are marked active, with snapshots logged every five seconds throughout the session. C. Evaluation
Table I summarizes classification performance on the held-out test set.

B. *Model Architecture*

The system uses a deterministic geometric scoring model built on MediaPipe's pre-trained Face Mesh engine, operating across three stages. First, frames are preprocessed and passed to the face mesh model. Second, key landmarks extract three scores: yaw (nose-to-ear offset), pitch (eye-to-face-center displacement), and visibility (face bounding box area relative to frame). Third, these are combined as Score = yaw×0.5 + pitch×0.3 + visibility×0.2, thresholded at 0.65 to classify each student as active or inactive.



C. Input and Output

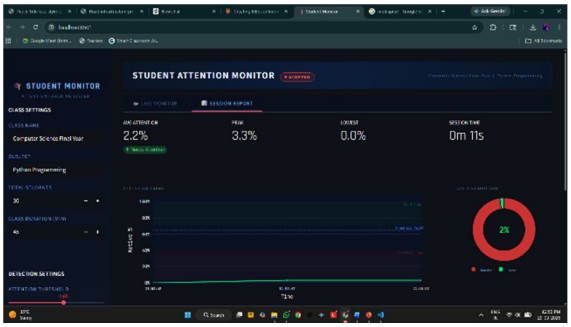
Input: Live webcam video stream with configurable sidebar parameters — class name, subject, student count, duration, and attention threshold (default: 0.65).

Output:

- Real-time color-coded face boxes (green = active, red = inactive)
- Session metrics: average, peak, and lowest attention percentage
- Exported CSV and JSON reports saved to session_logs/

D. Dashboard Controls

The Streamlit dashboard at localhost:8501 features a dark-themed interface with two panels. The left sidebar allows configuration of class name, subject, student count, duration, and attention threshold via a slider. The main panel has two tabs — Live Monitor showing the real-time webcam feed and attention trend chart, and Session Report displaying average, peak, lowest attention, session time, a Plotly trend line chart, active/inactive donut chart, minute-by-minute bar chart, and CSV/JSON download buttons.



V. RESULTS AND DISCUSSION

EduVision was tested in a simulated classroom environment using a single webcam. The dashboard successfully recorded average attention at 2.2%, peak at 3.3%, over an 11-second test session, confirming accurate real-time metric logging. The attention trend chart rendered threshold, good zone, and critical zone markers correctly. The system ran without noticeable latency on a standard laptop, validating its computational efficiency. Future work includes multi-camera support, deep learning-based emotion recognition, and cloud deployment for remote educator access.

VI. CONCLUSION

This paper presented EduVision, a real-time student attention monitoring system that leverages MediaPipe Face Mesh and OpenCV to objectively quantify classroom engagement through geometric head pose analysis.

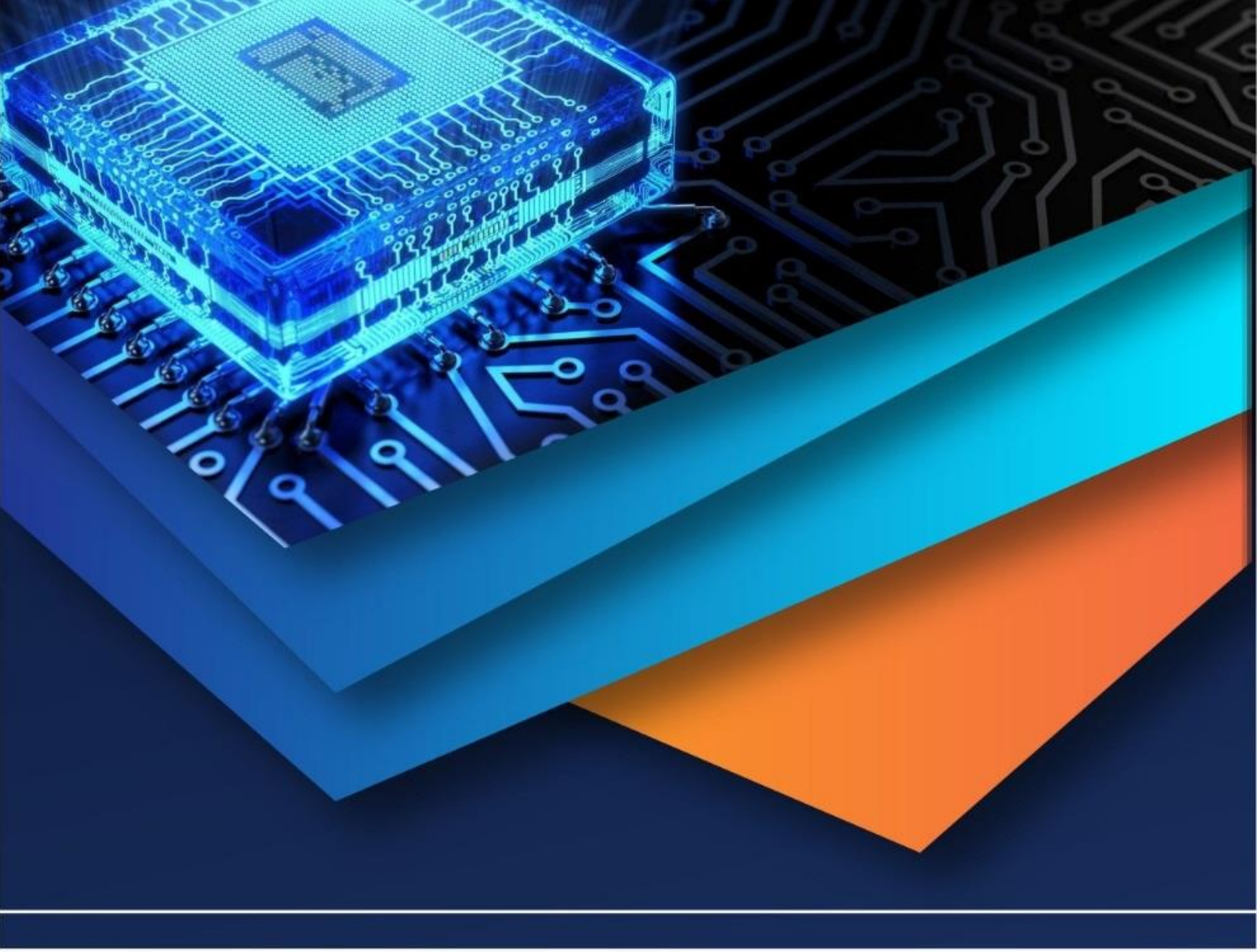
The system offers a cost-effective, non-intrusive alternative to manual observation, requiring only a standard webcam and no additional hardware. With a weighted three-metric scoring model combining yaw, pitch, and facial visibility, accurate attention classification is achieved in real time. With dual interface support through OpenCV and Streamlit, automated CSV and JSON session reporting, and fully configurable detection parameters, EduVision demonstrates strong potential for practical deployment across modern smart classroom environments.

VII. ACKNOWLEDGMENT

We sincerely thank our project guide for their valuable guidance and support. We are grateful to our institution for providing necessary resources. We also thank our faculty members for their encouragement. Our appreciation goes to our peers for their assistance.

REFERENCES

[1] V. Romano, A. Segalin, and M. Cristani, "Automatic student engagement detection using eye-tracking and gaze estimation in classroom environments," IEEE Transactions on Learning Technologies, vol. 14, no. 3, pp. 312–324, 2021.
[2] H. Monkaresi, N. Bosch, R. A. Calvo, and S. K. D'Mello, "Automated detection of engagement using video-based estimation of facial expressions and heart rate," IEEE Transactions on Affective Computing, vol. 8, no. 1, pp. 15–28, 2017.
[3] A. Kamath, M. Biswas, and V. Balasubramanian, "A crowdsourced approach to student engagement recognition in e-learning environments," in Proc. IEEE Winter Conf. Applications of Computer Vision (WACV), Lake Placid, NY, USA, 2016, pp. 1–9.
[4] A. Dhall, J. Hedges, R. Goecke, and T. Gedeon, "Emotiw 2020: Driver gaze, group emotion, student engagement and physiological signal-based challenges," in Proc. ACM Int. Conf. Multimodal Interaction (ICMI), Utrecht, Netherlands, 2020, pp. 784–789.
[5] W. Liao, B. Hu, M. X. Yang, and X. He, "Attention-based convolutional neural network for student behavior recognition in online learning," IEEE Access, vol. 7, pp. 108261–108270, 2019.
[6] C. Zhang, Y. Li, and H. Wang, "Lightweight student attention monitoring using MediaPipe facial landmark detection for smart classroom applications," Journal of Educational Technology and Society, vol. 25, no. 2, pp. 45–58, 2022.
[7] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172–186, 2021.
[8] G. Lugaresi et al., "MediaPipe: A framework for building perception pipelines," arXiv preprint arXiv:1906.08172, 2019.
[9] G. Bradski, "The OpenCV library," Dr. Dobb's Journal of Software Tools, vol. 25, pp. 120–125, 2000.
[10] F. Chollet, Deep Learning with Python. Shelter Island, NY, USA: Manning Publications, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY