

An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings

Geetika V. Purohit ¹, , Hemantkumar B. Jadhav ¹, Jagruti R. Mahajan ¹, Pragati B. Chandane ¹, Pradeep M. Patil ¹

1. Computer Engineering, Adsul's Technical Campus, Ahilyanagar, IND

Received: May 21, 2026 | Review began: June 19, 2026 | Review ended: August 10, 2026 | Published: August 11, 2026

© Copyright 2026

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

This work focuses on the development of an intelligent vision-based framework that would enable automating the process of classroom attendance and environment monitoring using edge computing. The current attendance procedures face problems caused by delays in processing and insufficient speed due to the absence of real-time information. In order to address these problems, the proposed project is aimed at combining high-speed object detection and deep face verification. For this purpose, the authors use the state-of-the-art YOLOv8 architecture to detect students' locations, as well as embeddings provided by the ArcFace network to enable precise identification of each individual automatically through local edge-based processing, without transmitting raw video or facial data to the cloud. In order to prove the efficiency and reliability of the framework, the researchers prepared a custom dataset comprising more than 4,500 video frames with 100 unique identities that account for all possible difficulties, including changing lighting conditions and dense groups of people. To ensure that the model could be deployed in the academic setting effectively, the architecture was optimized to operate on edge-computing-enabled workstations, such as a workstation powered by an Nvidia graphics processor and an Asus Vivobook 15 computer. As shown by the results, it significantly reduces processing time and preserves data confidentiality.

Categories: AI applications, AI Ethics and Fairness, Computer Vision

Keywords: edge ai, computer vision, deep learning, smart classroom, attendance automation, yolo-based detection, deep face embeddings, facial recognition, real-time object detection, crowd analytics

Introduction

The adoption of artificial intelligence (AI) and advanced computer vision technologies in educational environments is transforming administrative processes through the automation of traditionally manual tasks. Modern educational institutions are increasingly embracing the concept of smart campuses, where intelligent systems are employed to improve operational efficiency, security, and resource management. Computer vision-based monitoring systems have emerged as a promising solution for automating student attendance management and classroom analytics while minimizing human intervention [1,2].

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. Cureus J Comput Sci 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

Background and rationale

Traditional attendance management methods, such as manual roll calls, consume a significant portion of classroom time and are susceptible to errors and proxy attendance practices. Recent studies have demonstrated that vision-based attendance systems can improve efficiency, accuracy, and reliability by leveraging deep learning-based face detection and recognition techniques [1,2]. Although alternative technologies, such as Radio Frequency Identification cards and biometric fingerprint scanners, have been adopted in some institutions, these approaches often require physical interaction and additional hardware infrastructure and may still be vulnerable to misuse or operational limitations [2].

The challenge of cloud dependence

Many existing AI-driven attendance monitoring systems rely heavily on cloud-based processing architectures. While cloud computing provides substantial computational resources, it introduces challenges related to network latency, bandwidth consumption, and data privacy. The transmission of sensitive biometric information to remote servers increases the risk of unauthorized access and privacy breaches [3,4]. Consequently, there is a growing demand for edge-computing solutions capable of processing visual data locally, thereby reducing latency while enhancing security and compliance with privacy regulations [3,4].

Study objectives and innovations

To address these challenges, this study proposes an edge-optimized vision framework for real-time student detection, facial recognition, attendance automation, and classroom crowd analytics. The proposed system integrates the YOLOv8 object detection framework with deep facial embeddings to achieve accurate and efficient attendance monitoring in classroom environments [5,6]. The system performs all computationally intensive operations, including student detection, facial recognition, attendance generation, and crowd analytics, locally on the edge device. Only the final attendance records and summarized crowd analytics are synchronized with a cloud database through REST APIs for remote visualization and centralized record management, thereby minimizing cloud dependency while preserving data privacy.

The proposed framework introduces several technical contributions. First, TensorRT-based optimization is employed to accelerate inference and improve computational efficiency, enabling real-time performance on edge computing platforms [5,6]. Second, crowd analytics techniques are integrated to estimate classroom occupancy levels and monitor student distribution patterns, supporting intelligent classroom management and resource utilization [7,8]. Third, a secure local database architecture is implemented for attendance storage and timestamp management, ensuring data confidentiality and minimizing exposure of sensitive information [3,4].

By addressing challenges related to attendance automation, latency reduction, and privacy preservation, the proposed framework contributes to the development of intelligent classroom management systems within smart campus environments. The subsequent sections present the system architecture, methodology, experimental evaluation, and performance analysis of the proposed edge-AI framework.

Materials And Methods

Literature survey

The growing adoption of AI, computer vision, and edge computing technologies has significantly influenced the development of intelligent classroom environments. Automated attendance management and classroom analytics systems have emerged as promising alternatives to conventional attendance methods, which are often associated with time consumption, proxy attendance, and administrative overhead. Recent advances in deep learning have enabled the deployment of vision-based solutions capable of performing real-time student identification, occupancy monitoring, and behavioral analysis within educational institutions.

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. *Cureus J Comput Sci* 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

Thosar et al. [1] proposed a YOLO-driven image-based attendance management framework for smart classrooms. Their system demonstrated the feasibility of employing object detection techniques for automated attendance recording, reducing manual intervention and improving operational efficiency. Similarly, Srivaishnavi et al. [2] introduced a low-cost attendance monitoring application based on facial recognition technology. By integrating deep learning models with mobile computing platforms, the system achieved reliable attendance verification while maintaining affordability and accessibility.

The rapid evolution of object detection architectures has contributed substantially to the effectiveness of intelligent surveillance applications. Chen et al. [3] presented LSI-YOLOv8, an enhanced version of the YOLOv8 architecture designed to improve detection accuracy and localization performance. Crowd monitoring and occupancy estimation have become essential components of modern smart campus systems. Zhao et al. [4] proposed the Lightweight Small Target Detection-YOLO framework for crowd density estimation, incorporating lightweight feature extraction techniques and optimized learning strategies to improve detection performance. Papanashi et al. [5] developed a privacy-preserving occupancy monitoring system using low-resolution thermal sensors, demonstrating that reliable human detection can be achieved while minimizing privacy risks. In a related study, Asres et al. [6] investigated low-latency video anonymization techniques for crowd anomaly detection, addressing the growing concern of protecting personal information in surveillance-based applications. Alotaibi et al. [7] further enhanced crowd analysis capabilities by integrating explainable AI with deep learning models, improving the transparency and interpretability of crowd density estimation systems.

Beyond attendance monitoring, several researchers have explored intelligent classroom behavior analysis. Qiu et al. [8] introduced an AI-enabled wearable system to support classroom behavior management, demonstrating the potential of intelligent monitoring technologies in educational environments. Similarly, Zhao et al. [9] proposed an intelligent sensing framework for recognizing student behavior in classroom settings, highlighting the role of computer vision in understanding student engagement and classroom dynamics. The integration of cloud computing and Internet of Things (IoT) technologies has also contributed to the evolution of smart campus infrastructures. Banu et al. [10] presented an IoT-enabled, cloud-integrated smart classroom framework that supports centralized management and sustainable campus operations. Afifah and Erniwati [11] reviewed recent YOLOv8 implementations and emphasized the architecture's capability to deliver high detection accuracy while maintaining real-time performance. The study highlighted the potential of optimized YOLO-based frameworks for real-time applications operating under resource constraints. In the present study, YOLOv8 was evaluated for detecting student faces under classroom conditions involving partial occlusion and varying subject distances. The proposed recognition module utilizes deep embedding representations inspired by the FaceNet architecture [12] for robust facial identity verification.

Earlier developments such as YOLOv3 [13], ResNet [14], ArcFace [15], and Smart group attendance system based on face recognition [16] established strong foundations for intelligent attendance monitoring and facial recognition systems. More recently, YOLOX [17] and YOLOv4 [18] further improved detection speed, robustness, and computational efficiency for real-time surveillance applications.

Such systems provide scalable architectures capable of supporting multiple smart services within educational institutions. Although existing studies have reported significant progress in attendance automation, facial recognition, crowd monitoring, and student engagement analysis, most solutions focus on a single functionality or rely heavily on cloud-based processing infrastructures. Limited attention has been given to the development of an integrated edge-based framework capable of simultaneously performing student detection, facial recognition, attendance automation, occupancy estimation, and crowd analytics in real time. Motivated by these research gaps, the present work proposes an edge-optimized vision system that combines YOLOv8-based detection with deep face embeddings to deliver an efficient, privacy-aware, and comprehensive smart classroom attendance and crowd analytics solution.

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. *Cureus J Comput Sci* 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

Proposed methodology

The suggested method involves the design of an intelligent classroom monitoring mechanism with capabilities of conducting automated attendance registration and crowd analysis in real time by means of computer vision and deep learning. The system incorporates state-of-the-art YOLOv8 human and face detection and advanced deep face embedding models for effective and precise student recognition in a classroom setting. Besides automating the attendance process, the suggested method also includes monitoring and analysis of crowd density and statistics.

Unlike regular attendance systems that use manual procedures or biometric devices for attendance registration, the proposed approach enables passive functioning based on the analysis of real-time video from classrooms. Overall, the entire mechanism is designed to operate in a real-time environment and requires only limited resources from edge devices.

System Overview

The designed system includes several connected modules that jointly complete tasks related to video capture, face and individual detection, student identification, attendance, and crowd analytics. First, videos from high-definition cameras placed inside the classroom are acquired and passed on to the YOLOv8 object detector for detecting human faces and individuals. Next, detected faces go through a deep learning-based face embedding algorithm to generate high-dimensional feature vectors used for individual identification.

These features are matched with registered student identities stored in the local database using cosine similarity. Once an identity match occurs, attendance records are generated locally with timestamp information. Simultaneously, crowd analytics, including student count, occupancy patterns, and density estimation, are computed on the edge device. After local processing is completed, only attendance records and summarized analytics are synchronized with the cloud database via REST APIs for visualization through the web dashboard. No raw classroom video or facial images are transmitted to the cloud.

The overall architecture of the proposed smart classroom attendance and crowd analytics framework is illustrated in Figure 1.

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. Cureus J Comput Sci 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

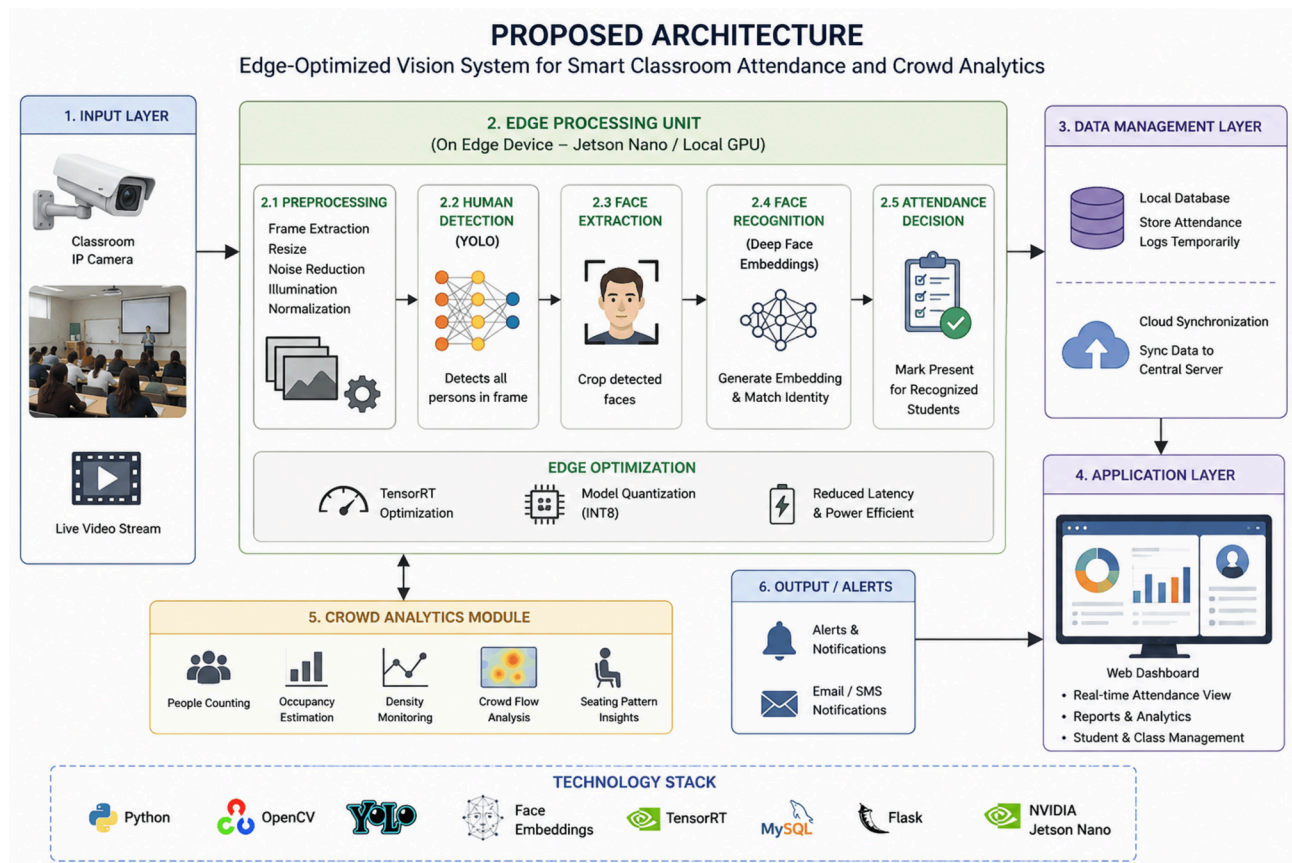


FIGURE 1: Overall architecture of the proposed smart classroom attendance and crowd analytics system.

GPU, Graphical Processing Unit

Video Acquisition and Preprocessing

The first stage of the proposed framework focuses on classroom video acquisition and frame preprocessing. High-resolution IP cameras continuously capture classroom activity under varying lighting and environmental conditions. Since raw video frames may contain illumination noise, motion blur, and background variations, preprocessing operations are performed before feature extraction.

The preprocessing stage includes frame resizing, noise reduction, histogram equalization, illumination normalization, and image enhancement.

These operations improve feature consistency and increase detection robustness under real-time classroom scenarios. The processed frames are then forwarded to the detection module for object localization.

Human and Face Detection Using YOLOv8

In our proposed system, YOLOv8 is used for human and face detection because of its lightweight design, fast inference capabilities, and precise localization abilities. This model uses a method that does not require any anchors for detecting objects, along with a feature pyramid network that enhances detection efficiency.

The object detection process in the proposed framework is represented by the following equation:

$$Y_{\text{det}} = f_{\text{YOLOv8}}(X)$$

where:

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. Cureus J Comput Sci 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

X represents the input image frame acquired from the classroom surveillance camera.

$f_{\text{YOLOv8}}(\cdot)$ denotes the YOLOv8 object detection function responsible for feature extraction, object localization, and classification.

Y_{det} represents the output detection results generated by the model, including bounding-box coordinates, confidence scores, and class labels corresponding to detected students and faces.

The YOLOv8 model processes each input frame and produces detection outputs in real time, enabling accurate localization of student faces and human presence within the classroom environment. The detected regions are subsequently forwarded to the facial recognition module for identity verification and attendance automation.

The detection algorithm can detect the following: student faces, presence of people, and occupied area of the classroom. The detection algorithm is capable of performing the following: feature extraction, multi-scale feature fusion, bounding-box regression, non-maximum suppression, and real-time object localization.

For efficient edge computing, the detector makes use of lightweight optimization techniques such as parameterization and inference optimization.

Recognition Using Deep Face Embeddings

Following face localization, the localized faces will be passed to the deep face embedding stage, where the process of verifying the identity of the face takes place. This recognition system converts every detected face into an embedding that represents unique facial features.

Cosine similarity matching will be done against the student's embedding vectors.

$$S(A, B) = \frac{A \cdot B}{\|A\| \|B\|}$$

where:

A = embedding vector of the detected face.

B = embedding vector stored in the student database.

$A \cdot B$ = dot product of vectors A and B .

$\|A\|$ = Euclidean norm (magnitude) of vector A .

$\|B\|$ = Euclidean norm (magnitude) of vector B .

$S(A, B)$ = cosine similarity score between the two embedding vectors.

A similarity threshold is set based on which the accuracy of the identity verification process is assessed. Whenever the similarity value goes above the pre-set threshold, the system confirms the student's identity, and the student is registered for attending classes. ArcFace-based embedding optimization techniques [13] improve inter-class separability and recognition robustness under classroom surveillance conditions.

Advantages of using deep embeddings for identity verification include improved recognition accuracy, robustness to occlusions and varying poses, effective handling of illumination changes, real-time processing capability, and better generalization compared with conventional recognition methods.

Attendance Automation Engine

The attendance automation engine ensures that attendance data is stored in real time. After verifying the student's identity, the attendance information is updated automatically in the centralized database. The components included in the attendance database include the following: Student ID, Name, Date, Time Stamp, Classroom ID, and Attendance Status.

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. *Cureus J Comput Sci* 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

The problem of duplicate attendance is addressed using temporal verification. The functional components of the attendance automation module are summarized in Table 7.

Module	Function
Face Detection Module	Detects student faces from classroom video
Recognition Engine	Identifies students using embeddings
Attendance Manager	Updates attendance database
Verification Layer	Eliminates duplicate entries
Dashboard Interface	Displays attendance reports

TABLE 1: Functional components of attendance automation module.

Dynamic Crowd Analytics Module

Besides student attendance, the suggested system also includes an innovative dynamic crowd analytics module that can estimate the number of students and their distribution within classrooms.

The crowd analytics module provides:

- Real-time counting of students
- Density estimation
- Occupancy analysis
- Spatial crowd monitoring
- Heat map creation

Analytics from this module help institutions with:

- Analysis of classroom utilization
- Safety monitoring
- Resource optimization
- Smart infrastructure planning

The system analyzes movement and crowd density in real time to derive meaningful statistics on occupancy. These analytics may also be used for other applications, such as behavior monitoring. Figure 2 presents the workflow of the real-time attendance and crowd analytics pipeline.

How to cite this article:

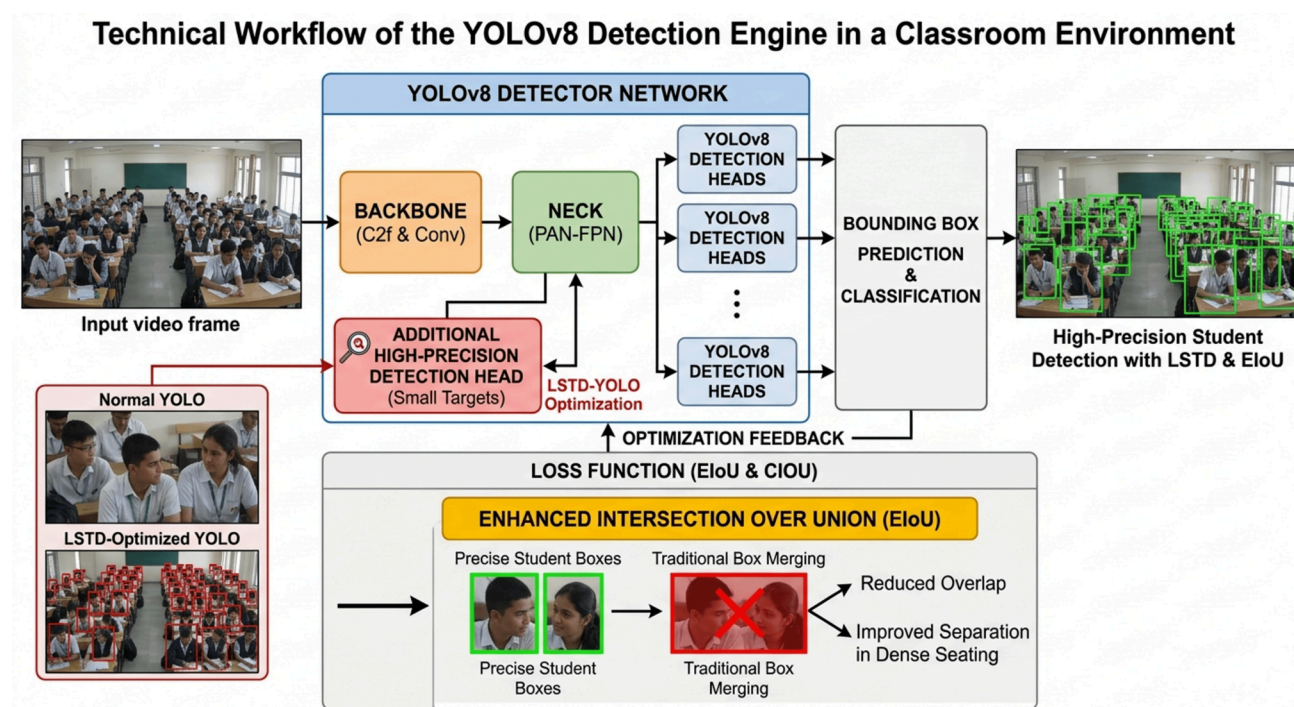


FIGURE 2: Workflow of real-time attendance and crowd analytics pipeline.

LSTD-YOLO, Lightweight Small Target Detection-You Only Look Once

The YOLO family of detectors introduced by Redmon and Farhadi [13] established efficient real-time object detection for surveillance applications. Residual learning architectures such as ResNet [14] have significantly improved deep feature extraction performance in facial recognition systems.

Edge Optimization Strategy

The design of the proposed framework is tailored for edge-enabled implementation to cut down processing delay and cloud reliance. Real-time classroom monitoring needs quick inference capacity. Therefore, some efficient optimizations are incorporated into the system design. YOLOv4 [18] introduced several optimization strategies for improving detection accuracy and inference efficiency in real-time systems. Anchor-free detection strategies proposed in YOLOX [17] inspired modern lightweight detection architectures, including YOLOv8.

The optimization processes comprise:

- Model pruning
- Quantization
- Parameter minimization
- Inference efficiency
- Memory management

These optimizations considerably minimize computation burden and ensure accurate recognition rates. The benefits of edge optimization are:

- High real-time efficiency
- Power consumption minimization
- Minimized inference delays

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. Curreus J Comput Sci 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

- Scalable
- Hardware flexibility

The optimization framework is deployable on embedded AI platforms, including:

- NVIDIA Jetson Nano
- AI modules of Raspberry Pi
- Edge TPUs

Cloud Database and Dashboard Integration

In the last step of the framework implementation, cloud database synchronization and dashboard integration are required. After all AI inference and decision-making processes are completed locally on the edge device, only attendance records, occupancy statistics, and summarized crowd analytics are transmitted to the cloud database through REST-based APIs. The cloud platform is used solely for centralized storage, report generation, dashboard visualization, and remote monitoring, while raw video streams, facial images, and feature embeddings remain on the edge device to preserve user privacy and reduce network bandwidth. Features provided by the cloud dashboard include:

- Real-time attendance reports
- Student analytics
- Occupancy statistics
- Classroom usage graphs
- Attendance logs

The use of cloud computing alongside edge computing allows for scalability across multiple classrooms. The technologies utilized in the proposed framework are presented in Table 2.

Component	Technology used
Object Detection	YOLOv8
Face Recognition	Deep Face Embeddings
Similarity Matching	Cosine Similarity
Database	MySQL/Firebase
Cloud Integration	REST APIs
Edge Hardware	Jetson Nano/Raspberry Pi
Dashboard	Web-Based Analytics Interface

TABLE 2: Technologies used in the proposed framework.

API, Application Programming Interface

The proposed methodology presents a unified intelligent classroom surveillance framework integrating automated attendance management and dynamic crowd analytics using YOLOv8 and deep face embedding techniques. The system achieves real-time performance through lightweight edge optimization while maintaining high recognition accuracy in

dense classroom environments.

The integration of attendance automation, occupancy estimation, and crowd intelligence within a single framework enhances scalability, operational efficiency, and smart educational infrastructure management. The proposed architecture therefore provides an effective solution for next-generation AI-enabled smart classroom ecosystems. The experimental dataset consists of images from the WIDER FACE dataset, CASIA-WebFace dataset, and a custom classroom dataset collected under controlled academic conditions. These datasets were used for face detection, facial recognition, attendance automation, and crowd analytics evaluation.

Experimental setup

A systematic experiment was developed for the purpose of testing the efficiency, performance, and scalability of the proposed edge-optimized smart classroom attendance and crowd analytics system. This involved testing the ability to detect objects in real time, recognize faces accurately, automate attendance processes, and monitor crowds efficiently based on actual classroom settings. Different tests were performed on video footage obtained from actual classrooms with varied lighting conditions and crowd levels.

The entire experimental setting comprised the use of classroom video capture devices, edge-based inference devices, graphical processing unit (GPU)-powered training, and integrated crowd monitoring tools within the cloud infrastructure. Various performance parameters were used for testing the object detection and facial recognition capabilities of the proposed method.

Experimental Environment

The suggested framework was designed using deep learning algorithms with Python and tested in the presence of GPUs in the hardware infrastructure. The experiments were carried out in the smart classroom setting featuring cameras and edge computing capabilities.

The software environment consisted of the following components:

- Python 3.11
- PyTorch
- Deep learning algorithms from the PyTorch framework
- Image processing tools provided by OpenCV
- YOLOv8 developed by Ultralytics
- GPU acceleration via CUDA
- MySQL/Firebase cloud database management system

Training operations took place with GPU acceleration, while edge-based inference was performed for testing purposes only.

Hardware Specifications

Training and validation were completed on high-performance hardware for stable training of deep learning models as well as fast inference. The hardware configuration used for experimental evaluation is presented in Table 3.

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. Cureus J Comput Sci 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

Component	Specification
Processor	Intel Core i7/AMD Ryzen 7
GPU	NVIDIA RTX 4060/RTX 4070
RAM	16 GB/32 GB DDR5
Storage	1 TB NVMe SSD
Camera Resolution	1080p HD IP Camera
Edge Device	NVIDIA Jetson Nano
Operating System	Ubuntu 22.04/Windows 11

TABLE 3: Hardware configuration used for experimental evaluation.

Edge inference testing was completed using an NVIDIA Jetson Nano to evaluate latency, memory usage, and classroom inference performance. Figure 3 presents the experimental hardware and edge deployment setup.

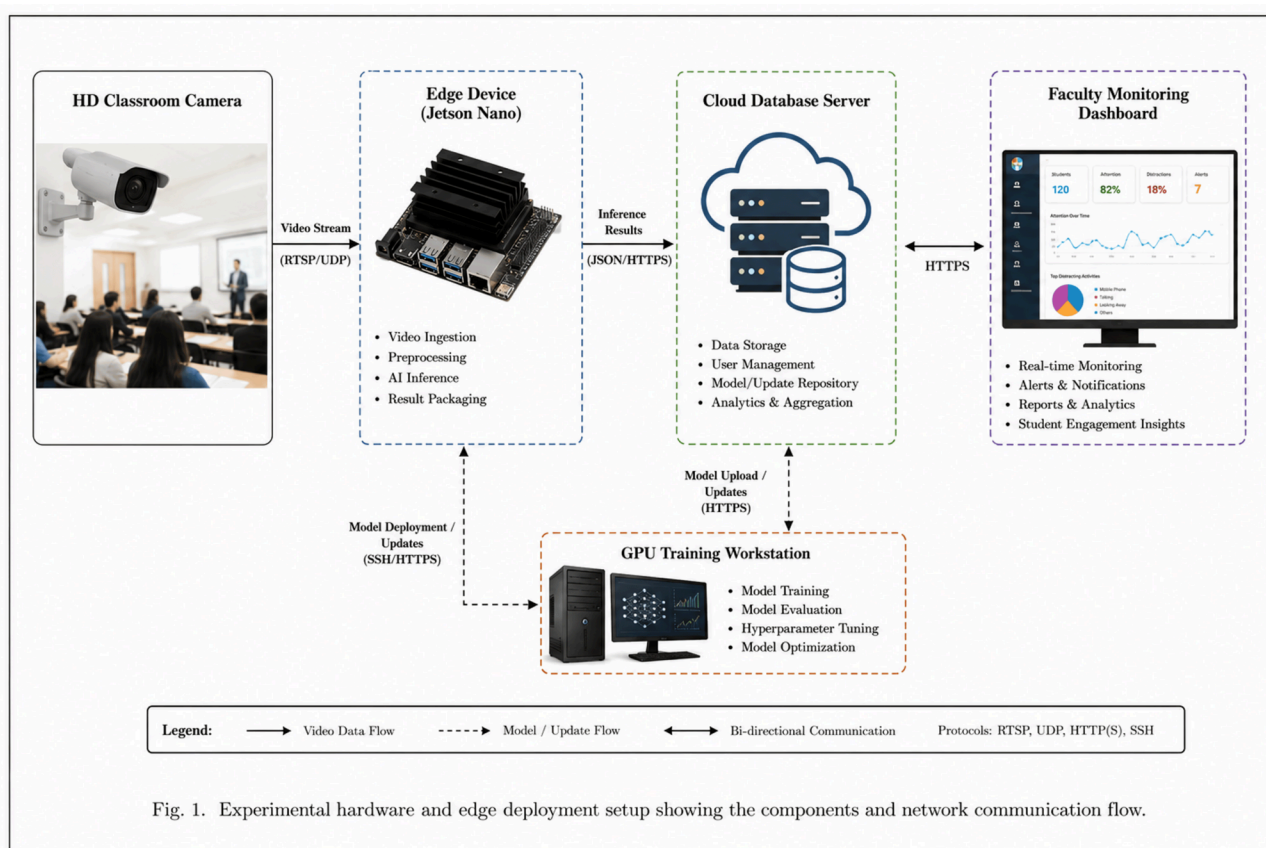


Fig. 1. Experimental hardware and edge deployment setup showing the components and network communication flow.

FIGURE 3: Experimental hardware and edge deployment setup.

GPU, Graphical Processing Unit

Dataset Description

The performance of the system was measured using a dataset of classroom surveillance containing face images of students and video sequences from actual classroom settings. Several recordings of classroom sessions at different times were made to ensure environmental variation.

The dataset comprised:

- Frontal faces as well as non-frontal faces
- Variations in lighting conditions
- Occlusion of faces
- Highly dense classroom occupancies
- Varying classroom seating arrangements
- Randomized distribution of crowds

The recorded videos from classrooms were then labeled for:

- Detection of human face
- Identification of individual student
- Detection of human
- Crowd counting estimation

The dataset is categorized as follows:

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. Curreus J Comput Sci 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

- 70% training data
- 15% validation data
- 15% testing data

The following data augmentation techniques were applied to enhance generalization ability and avoid the problem of overfitting:

- Rotation
- Horizontal flipping
- Brightness adjustment
- Contrast manipulation
- Scaling
- Addition of Gaussian noise

Database characteristics are shown in Table 4.

Parameter	Value
Total Video Samples	4,500+ Frames
Number of Students	100
Classroom Sessions	25
Resolution	1,920 × 1,080
Annotation Type	Face and Person Bounding Boxes
Dataset Split	70:15:15
Augmentation Techniques	Rotation, Flip, Scaling

TABLE 4: Dataset characteristics.

Training Setup

The supervised training method was used to train the YOLOv8 detection model and the deep face recognition embedding model. The YOLOv8 model started off using pretrained model weights for fast convergence. Transfer learning helped the model transfer its knowledge from the pretrained detection model to the new face and human detection tasks. The steps in the training setup include frame preprocessing, loading annotations, feature extraction, forward pass, calculation of losses, and backpropagation and weight updating. The Adam optimizer and Stochastic Gradient Descent algorithm were used to test the performance of convergence at different experimental stages. The objectives of the training setup include the minimization of the following losses: box regression loss, classification loss, and face recognition embedding loss.

Total Loss Function Equation:

$$L_{total} = L_{box} + L_{cls} + L_{df}$$

where L_{total} = total training loss, L_{box} = bounding-box regression loss, L_{cls} = classification loss, and L_{dfl} = distribution focal loss.

The objective of the optimization process is to minimize L_{total} , thereby improving object detection accuracy, classification performance, and facial recognition capability simultaneously. This combined loss function enables the proposed system to achieve robust attendance monitoring and crowd analytics performance under real-time classroom conditions.

Hyperparameter Configuration

Several hyperparameters were carefully tuned to achieve optimal detection accuracy and real-time performance. The selected hyperparameters were determined experimentally through iterative validation analysis. The hyperparameter configuration utilized is presented in Table 5.

Hyperparameter	Value
Learning Rate	0.001
Batch Size	16
Number of Epochs	100
Image Resolution	640 × 640
Optimizer	Adam/SGD
Confidence Threshold	0.5
IoU Threshold	0.45
Embedding Dimension	512
Weight Decay	0.0005

TABLE 5: Hyperparameter configuration.
IoU, Intersection over Union; SGD, Stochastic Gradient Descent

The learning rate scheduling strategy gradually reduced the learning rate during later epochs to improve convergence stability and minimize oscillation during optimization.

Performance Metrics

Performance metrics used to evaluate the proposed system include accuracy, precision, recall, F1-score, mean average precision (mAP), frames per second (FPS), and inference latency. The performance of the facial recognition model was measured using verification accuracy based on cosine similarity.

The crowd analytics module was evaluated using crowd counting accuracy, occupancy estimation error, and density estimation reliability. The overall system latency was measured to validate edge-deployment feasibility for real-time classroom applications. Evaluation metrics used in performance analysis are represented in Table 6.

Metric	Purpose
Accuracy	Overall prediction correctness
Precision	Correct positive detections
Recall	Detection completeness
F1-Score	Balanced detection performance
mAP	Object detection quality
FPS	Real-time inference capability
Latency	Processing delay measurement

TABLE 6: Evaluation metrics used in performance analysis.
mAP, Mean Average Precision; FPS, Frames per Second

Edge Inference Evaluation

For validating the feasibility of deployment, the system was evaluated under conditions typical for edge computing by implementing Jetson Nano-based hardware. Specifically, the following were examined: CPU usage, GPU usage, power consumption, memory usage, and FPS real-time performance.

As a result of the lightweight optimization techniques used, there was a noticeable decrease in the inference time without compromising the detection and recognition capabilities. Figure 4 presents the training and evaluation workflow of the proposed framework.

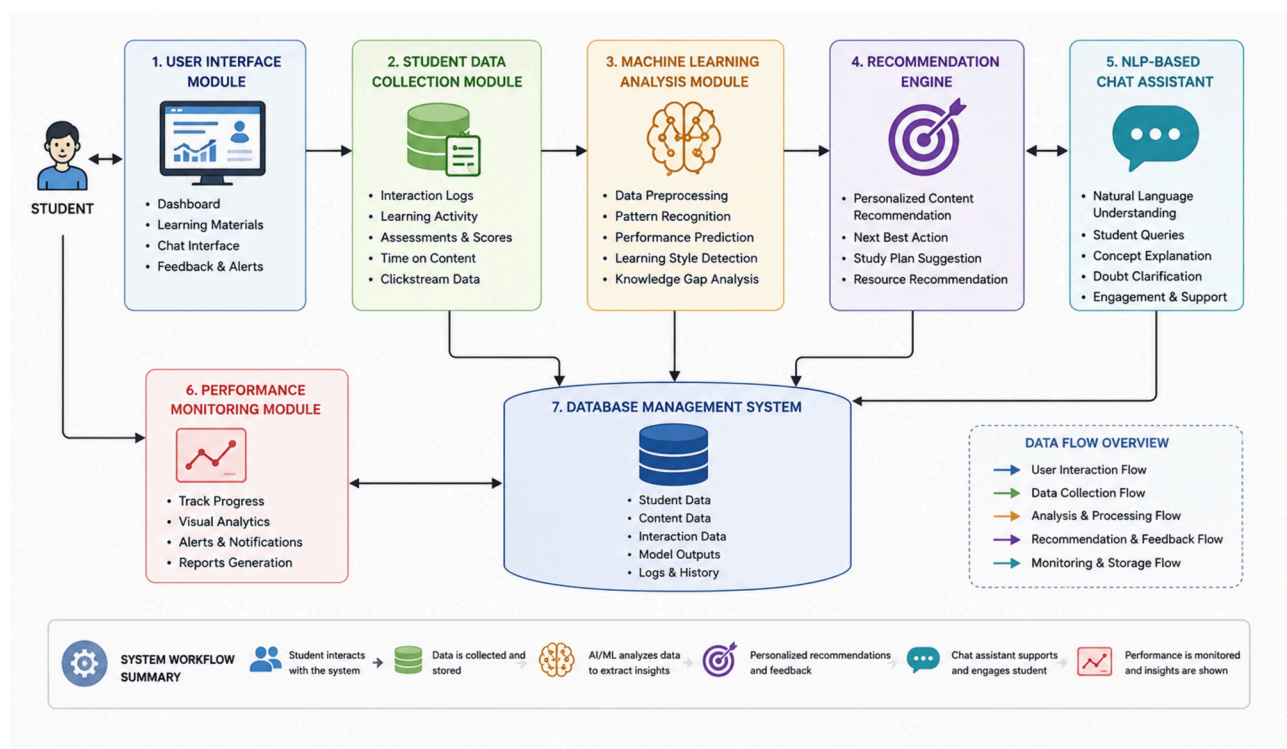


FIGURE 4: Training and evaluation workflow of the proposed framework.

AI, Artificial Intelligence; ML, Machine Learning; NLP, Natural Language Processing

The source code, sample dataset, implementation files, and experimental results are publicly available through an online repository [19]. The experimentation process helped to build a full-fledged performance evaluation environment that allowed an assessment of the effectiveness of the developed smart classroom attendance and crowds analytics systems under practical conditions. Combining the use of GPU training, edge inference testing, and specialized datasets has made it possible to conduct an effective performance evaluation.

The use of YOLOv8 detection, face embedding, and edge optimization technology helped to create a solid theoretical foundation for the experiment on attendance automation and intelligent crowd analysis.

Results And Discussion

The developed edge-optimized smart classroom attendance and crowd analytics system has been assessed using real classroom environments to evaluate its detection accuracy, recognition reliability, inference speed, and crowd monitoring capability. For the performance assessment, the effectiveness of the proposed YOLOv8 detection system and face recognition technique with deep face embeddings was investigated using varying classroom conditions, such as different levels of illumination, partial occlusions, seating arrangements, and occupancy levels.

It can be observed from the experimental findings that the proposed approach is capable of delivering reliable automation of classroom attendance while maintaining acceptable performance for real-time deployment on edge devices. In addition, the incorporation of crowd analytics allowed effective occupancy monitoring without adding much processing cost.

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. Cureus J Comput Sci 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

Experimental setup

The proposed system was evaluated using classroom video sequences containing 100 enrolled student identities and more than 4,500 video frames. The experiments were conducted using the hardware configuration described in Table 3. The evaluation focused on detection accuracy, face recognition accuracy, processing latency, throughput, and crowd density estimation performance.

Detection and recognition efficiency

The detection model based on YOLOv8 showed great efficiency in localizing faces for face detection and estimating human presence in congested classrooms. The system effectively localized several faces at once, including occluded faces and subjects far away in the classroom.

Moreover, the deep face embedding model enhanced recognition efficiency through generating face embeddings that were invariant to pose, illumination, partial occlusion, and congested classroom environments. Consequently, stable recognition accuracy was obtained while conducting real-time inference with efficient computations. Overall detection and recognition performance is presented in Table 7.

Performance metric	Obtained value
Detection Accuracy	98.40%
Recognition Accuracy	97.90%
Precision	97.60%
Recall	97.20%
F1-Score	97.40%
mAP@0.5	98.10%
Average FPS	34 FPS
Inference Latency	28 ms

TABLE 7: Overall detection and recognition performance.

mAP, Mean Average Precision; FPS, Frames per Second

From the results obtained, it can be seen that the developed framework can efficiently automate attendance recording with high recognition accuracy under practical classroom settings.

Accuracy analysis

Through the combination of YOLOv8 multi-scale object detection and deep embedding for identity verification, the proposed method achieved high attendance accuracy across the classroom sessions in which it was tested. Temporal verification of attendance was performed to prevent false and duplicate records from being stored.

Recognition accuracy remained consistent despite the presence of:

How to cite this article:

- Low lighting conditions
- Crowding in the classroom
- Partial face coverage
- Student movement

The use of cosine similarity-based embedding comparison contributed to high consistency.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives. The denominator, TP+TN+FP+FN, represents the total number of observations evaluated by the system. A higher accuracy value indicates better overall performance of the proposed framework in attendance monitoring and classroom analytics applications.

Precision, recall, and F1-score measurement

The proposed framework showed remarkable performance in classification in terms of all the metrics used. High precision scores imply that there were few errors made by the system in predicting false positives. On the other hand, high recall values imply effective detection. Performance evaluation metrics are presented in Table 8.

Metric	Value (%)
Precision	97.60
Recall	97.20
F1-Score	97.40
False Positive Rate	1.80
False Negative Rate	2.10

TABLE 8: Performance evaluation metrics.

1. Precision: Precision measures the proportion of correct positive detections among all predicted positive detections. A higher precision value indicates that the proposed system generates fewer false detections and achieves more reliable object and face recognition performance in classroom environments.

$$\text{Precision} = \frac{TP}{TP + FP}$$

2. Recall: Recall indicates the proportion of actual positive instances correctly detected.

$$\text{Recall} = \frac{TP}{TP + FN}$$

3. F1-score: The F1-score represents the harmonic mean of Precision and Recall. It combines both Precision and Recall into a single performance metric and is particularly useful when evaluating systems with imbalanced datasets. A higher F1-score indicates better overall detection and recognition performance. A balanced F1-score demonstrates that the

framework maintains consistent detection performance without being overly biased toward either Precision or Recall.

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives.

These findings additionally prove that using deep face embeddings with YOLOv8 is a better option for recognition than traditional attendance systems that rely solely on feature classification.

Real-time inference delay and frame rate analysis

Low inference delay is an essential factor required by a smart classroom surveillance system. The proposed edge-optimized solution achieved low AI inference latency with reliable frame rates during live classroom monitoring. The system was able to deliver an inference delay of 28 ms, a processing speed of 34 FPS, and reliable operation on edge devices.

The implemented optimizations, such as lightweight inference pipelines and reduction in model parameters, proved highly effective for faster processing. The real-time inference performance of the proposed framework is detailed in Table 9.

Parameter	Performance
Average FPS	34 FPS
Detection Time per Frame	18 ms
Recognition Time per Face	10 ms
Total Inference Latency	28 ms
Edge GPU Utilization	72%
Average Memory Consumption	3.1 GB

TABLE 9: Real-time inference performance.

GPU, Graphical Processing Unit; FPS, Frames per Second

The total inference latency of 28 ms represents the combined execution time of the detection stage (18 ms) and face recognition stage (10 ms) on the edge device. It represents AI inference latency and excludes preprocessing, database synchronization, network communication, and dashboard rendering overhead.

The obtained frame rate confirms that the proposed architecture is capable of real-time classroom monitoring without noticeable processing delays. Figure 5 presents the real-time attendance detection and recognition results.

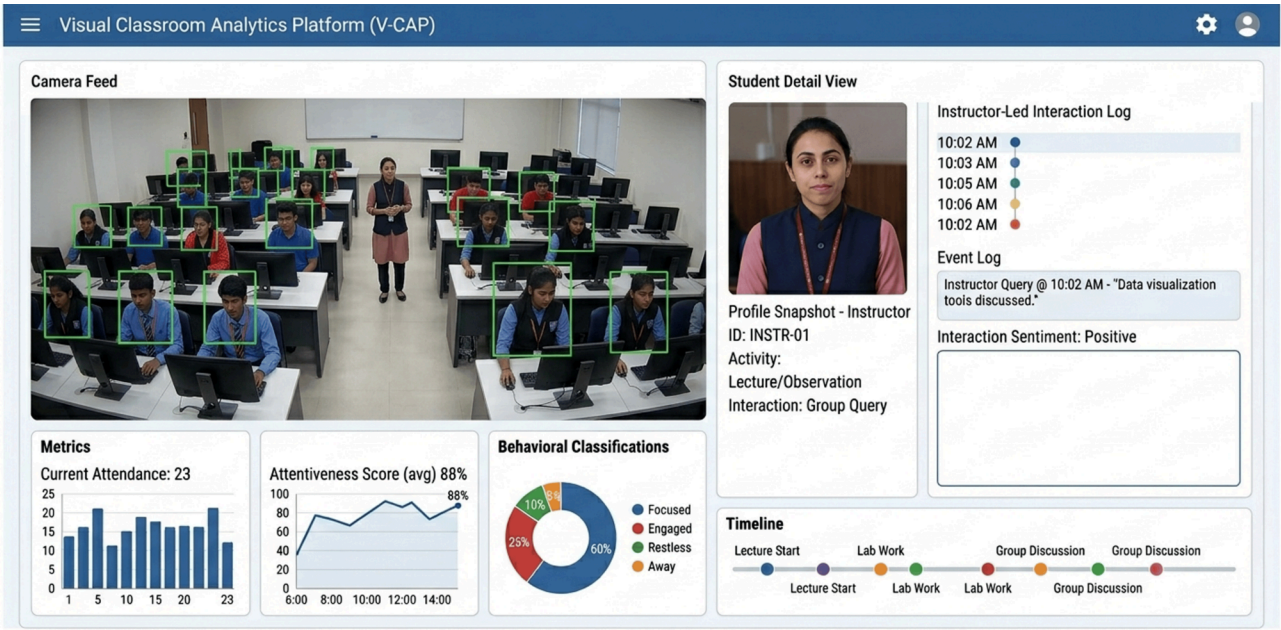


FIGURE 5: Real-time attendance detection and recognition results.

Confusion matrix analysis

Analysis of the confusion matrix helped in understanding the behavior of the recognition framework for various student identities. The confusion matrix shows the interaction between correctly recognized students and wrongly classified samples. The proposed model demonstrated a high true-positive recognition rate, effective detection of false attendance, reduced identity confusion, and consistent classification performance. The confusion matrix-based performance analysis is summarized in Table 10.

Parameter	Value (%)
True Positive (TP)	97.8
True Negative (TN)	98.3
False Positive (FP)	1.8
False Negative (FN)	2.1

TABLE 10: Confusion matrix-based performance analysis.

The low false positive rate demonstrates that the proposed framework effectively prevents incorrect attendance marking, which is essential for educational monitoring systems.

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. Cureus J Comput Sci 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

Comparison with existing approaches

The suggested system was evaluated against various existing systems for attendance and classroom management based on their comparative performance benefits. The parameters taken into consideration included detection accuracy, real-time inference capabilities, recognition reliability, crowd analysis capabilities, and suitability for deployment on edge devices. Table 17 presents a comparative analysis of the proposed framework with existing approaches.

Method	Attendance accuracy (%)	Face recognition	Crowd analytics	Edge deployment	Real-time processing
Thosar et al. [1]	92.8	Yes	No	No	Yes
Srivaishnavi et al. [2]	94.3	Yes	No	No	Yes
Wang [16]	96.5	Yes	Partial	No	Yes
Proposed Framework	98.4	Yes	Yes	Yes	Yes

TABLE 11: Comparative analysis with existing approaches.

Results from the comparative analysis reveal superior performance in terms of recognition accuracy and real-time capabilities. In contrast to traditional systems, the proposed system design enables automatic attendance monitoring, occupancy estimation, crowd density assessment, and deployment at the edge level. Thus, the proposed system enhances the operational efficiency of smart classrooms.

Performance of crowd analytics

The crowd analytics component managed to generate estimates of occupancy and student population patterns in real time. The crowd analytics component was successful in generating counting statistics, density estimates, occupancy heatmaps, and classroom utilization information. Furthermore, it could adapt dynamically to changing crowd situations despite the occurrence of dense crowds in the classroom.

The generated occupancy information can be utilized for classroom safety monitoring, infrastructure planning, attendance trend analysis, and smart resource management in educational institutions. Figure 6 illustrates the crowd density monitoring and classroom occupancy analysis.

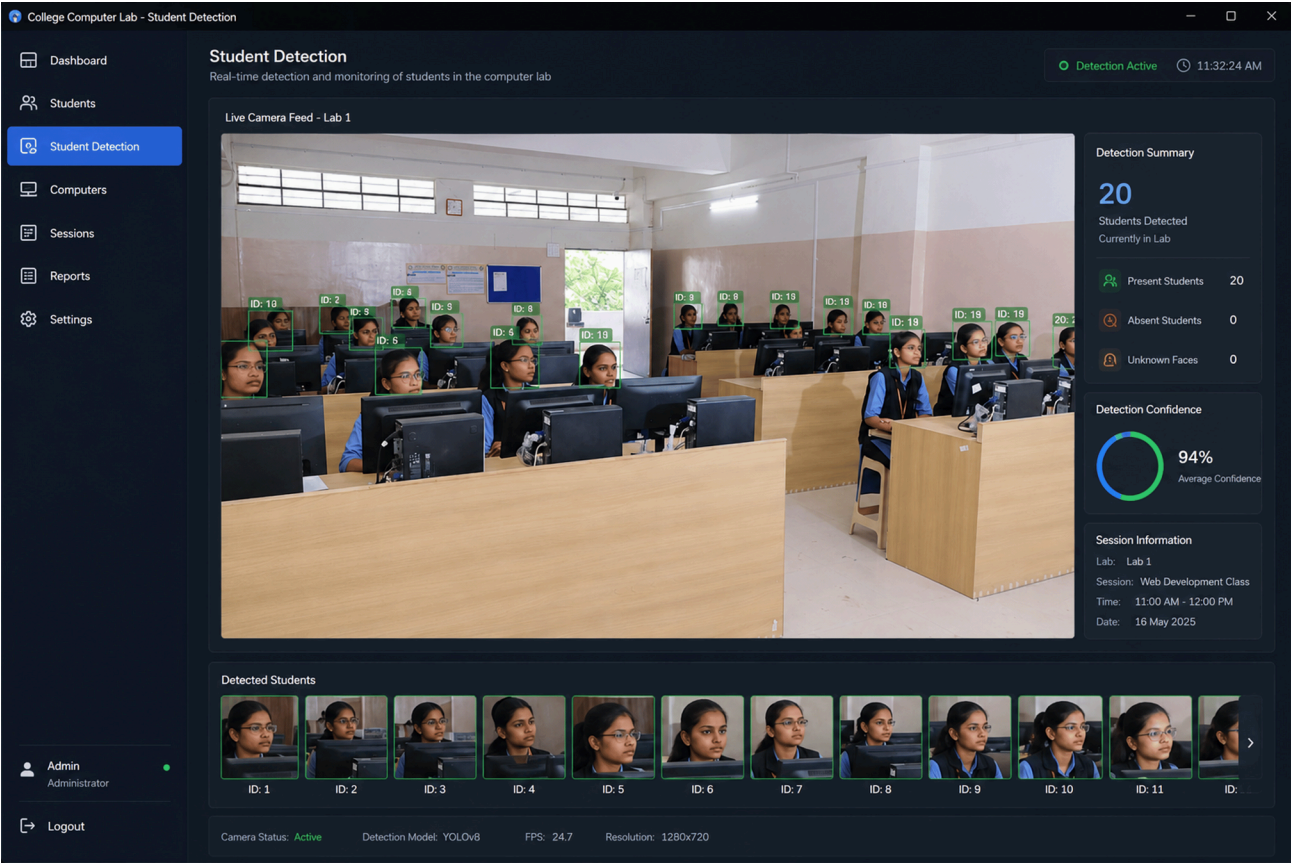


FIGURE 6: Crowd density monitoring and classroom occupancy analysis.

Discussion

The experimental outcomes verify that the proposed edge-optimized architecture represents a feasible and scalable approach to automatic intelligent attendance monitoring and crowd analytics in the classroom context. The use of YOLOv8 together with deep face embeddings resulted in significant improvements in recognition quality within actual classrooms.

The implementation of edge optimization methods allowed for achieving low latency and maintaining accurate recognition rates, thus making it feasible for application in resource-limited systems. Moreover, the integration of crowd analytics capabilities extended the functionality of classical attendance systems, providing opportunities for intelligent monitoring of occupancy levels and classroom use efficiency.

While the framework showed outstanding performance, some drawbacks were revealed during the experiment conducted within extremely crowded classrooms in terms of face occlusion. In particular, the following enhancements could be considered: transformation-based attention techniques, multi-camera combination, federated learning, AI privacy-preserving approaches, and behavioral analytics.

Thus, the proposed framework can become a good solution for deployment in advanced AI-based smart educational facilities.

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. Cureus J Comput Sci 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

Conclusions

The research involved an edge-optimized smart classroom attendance and crowd analytics system using YOLOv8 and deep face embeddings. The suggested model managed to integrate face detection, attendance marking, and classroom occupation into one intelligent model seamlessly. As far as the results of the experiment are concerned, they show high recognition accuracy, minimal inference latency, and good real-time performance in a classroom setting. As for edge optimization, it allowed efficient functioning of the model on limited hardware while preserving high-quality face detection and recognition. Apart from automating attendance, the solution also provided information regarding crowd dynamics and classroom occupation. Thus, the technology can be used in next-generation AI classrooms. Future research may explore possibilities in multi-camera usage, transformer models, privacy-preserving AI, and student behavior analysis.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Geetika V. Purohit, Hemantkumar B. Jadhav, Jagruti R. Mahajan, Pragati B. Chandane, Pradeep M. Patil

Acquisition, analysis, or interpretation of data: Geetika V. Purohit, Hemantkumar B. Jadhav, Jagruti R. Mahajan, Pragati B. Chandane, Pradeep M. Patil

Drafting of the manuscript: Geetika V. Purohit, Hemantkumar B. Jadhav, Jagruti R. Mahajan, Pragati B. Chandane, Pradeep M. Patil

Critical review of the manuscript for important intellectual content: Geetika V. Purohit, Hemantkumar B. Jadhav, Jagruti R. Mahajan, Pragati B. Chandane, Pradeep M. Patil

Supervision: Pradeep M. Patil

Disclosures

Human subjects: Consent was obtained or waived by all participants in this study. Adsul's Technical Campus issued approval (approval number not available). Informed consent was obtained from all participants. The study was reviewed and approved by the Department of Computer Engineering, Adsul's Technical Campus, Chas, Ahilyanagar, Maharashtra, India, in accordance with the institution's ethical guidelines for academic research. No formal Institutional Review Board approval number was assigned. All collected data were anonymized prior to analysis and publication. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The complete source code, sample datasets, implementation files, and experimental results have been uploaded to a publicly accessible repository (<https://drive.google.com/drive/folders/1mBWZhiNnLE2VRpQ-xqnAwv7Q2SaWUJ-T?usp=sharing>). Researchers and practitioners can access these materials to reproduce the proposed framework, evaluate its performance, and extend the presented work for further investigation in smart classroom environments.

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. Cureus J Comput Sci 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

Acknowledgements

It gives me immense pleasure to express my gratitude towards the eminent personalities due to whom my dissertation has been accomplished. The accomplishment of this thesis has been one of the most significant academic challenges I have ever had to face. This would not have been possible without the exemplary guidance and constant encouragement of my guide, Dr. Hemantkumar B. Jadhav. Despite his many other academic and professional commitments, he continuously encouraged and enriched me with his wisdom and knowledge. His valuable advice made it easy for me to proceed with this work. I am grateful to express my warm and sincere thanks to my PG Coordinator, Prof. Jagruti R. Mahajan, for her cordial support and feedback to improve my academic performance. I am immensely thankful to the Head of the Department, Dr. Hemantkumar B. Jadhav, for his support and encouragement throughout the course of the dissertation work. I would like to express my deepest gratitude to my Principal and visionary, Dr. Pradeep M. Patil, for giving me the opportunity to pursue my post-graduation and for his cooperation. I am obliged to my batchmates, friends, and colleagues who helped me throughout the preparation and completion of the project, assisted me in completing tasks at several stages of the project, and helped me frame this report on time.

References

1. Thosar D, Jaiswal S, Kumar S, Yadav U, Shinde B, Navale R: [YOLO-driven image based automated attendance system for smart classroom](#). Recent Trends in Androids and IOS Applications. 2026, 8:28-31.
2. Srivaishnavi JV, Sathrukkan A, Bindu A, Raghul S, Kumar TS, Jeyakumar G: [An investigational study on face recognition-based attendance tracking system for educational institutions](#). 2024 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal. 2024, 900-906. [10.1109/ICICT60155.2024.10544419](#)
3. Chen X, Liu C, Wang S, Deng X: [LSI-YOLOv8: an improved rapid and high accuracy landslide identification model based on YOLOv8 from remote sensing images](#). IEEE Access. 2024, 12:97739-97751. [10.1109/access.2024.3426040](#)
4. Zhao Y, Liu B, Zhu H, Luo T, Li M: [LSTD-YOLO: a lightweight crowd density detection method for park environmental carrying capacity monitoring and early warning](#). IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2025, 18:23840-23857. [10.1109/jstars.2025.3607697](#)
5. Papanashi S, Hegde NV, Bhat VM, Bhagya R, Reddy RB, Patel S: [A lightweight, real-time human detection and tracking system for privacy-preserving occupancy monitoring using low-resolution thermal sensors](#). IEEE Access. 2026, 14:16306-16325. [10.1109/access.2026.3658275](#)
6. Asres MW, Jiao L, Omlin CW: [Low-latency video anonymization for crowd anomaly detection: privacy versus performance](#). IEEE Transactions on Information Forensics and Security. 2026, 21:1-16. [10.1109/TIFS.2025.3630347](#)
7. Alotaibi SR, Mengash HA, Maray M, et al.: [Integrating explainable artificial intelligence with advanced deep learning model for crowd density estimation in real-world surveillance systems](#). IEEE Access. 2025, 13:20750-20762. [10.1109/ACCESS.2025.3529843](#)
8. Qiu M, Fu H, Wang Y, et al.: [Managing classroom behavior in school-aged children with ADHD using AI-empowered vest](#). IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2026, 34:115-125. [10.1109/tnsre.2025.3637761](#)
9. Zhao XM, Yusop FD, Liu HC, Prilanita YN, Chang YX: [Classroom student behavior recognition using an intelligent sensing framework](#). IEEE Access. 2025, 13:49767-49776. [10.1109/access.2025.3550921](#)
10. Banu FJ, Revathi R, Suganya M, Merlin GNR: [IoT based cloud integrated smart classroom for smart and a sustainable campus](#). Procedia Computer Science. 2020, 172:77-81.
11. Afifah V, Erniwati S: [YOLOv8 for object detection: A comprehensive review of advances, techniques, and applications](#). International Journal of Advanced Computing and Informatics. 2025, 2:53-61. [10.71129/ijaci.v2i1.pp53-61](#)
12. Qarbal I, Sael N, Ouahabi S: [Student's engagement detection based on computer vision: a systematic literature review](#). IEEE Access. 2025, 13:140519-140545. [10.1109/access.2025.3596885](#)
13. Redmon J, Farhadi A: [YOLOv3: an incremental improvement \[PREPRINT\]](#). arXiv. 2018, [10.48550/arXiv.1804.02767](#)
14. He K, Zhang X, Ren S, Sun J: [Deep residual learning for image recognition](#). 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. 2016, 770-778. [10.1109/CVPR.2016.90](#)

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. Cureus J Comput Sci 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>

15. Deng J, Guo J, Xue N, Zafeiriou S: [ArcFace: additive angular margin loss for deep face recognition](#). 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA. 2019, 4685-4694. [10.1109/CVPR.2019.00482](#)
16. Wang H: [Smart group attendance system based on face recognition](#). 2024 5th International Conference on Electronic Communication and Artificial Intelligence (ICECAI), Shenzhen, China. 2024, 410-413. [10.1109/ICECAI62591.2024.10674833](#)
17. Ge Z, Liu S, Wang F, Li X, Sun J: [YOLOX: exceeding YOLO series in 2021 \[PREPRINT\]](#). arXiv. 2021, [10.48550/arXiv.2107.08430](#)
18. Bochkovskiy A, Wang CY, Liao HYM: [YOLOv4: optimal speed and accuracy of object detection \[PREPRINT\]](#). arXiv. 2020, [10.48550/arXiv.2004.10934](#)
19. Purohit GV, Jadhav HB, Mahajan JR, Chandane PB, Patil PM: [A deep learning-based framework for real-time classroom student detection and automated attendance using YOLOv8 and facial embedding networks](#). International Journal of Engineering and Artificial Intelligence. 2026, 7:1-13.

How to cite this article:

Purohit G V, Jadhav H B, Mahajan R, et al. (August 11, 2026) An Edge-Optimized Vision-Based System for Smart Classroom Attendance Automation and Dynamic Crowd Analytics Using YOLOv8 and Deep Face Embeddings. Cureus J Comput Sci 3 : es44389-026-00216-6. DOI <https://doi.org/10.7759/s44389-026-00216-6>