International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

An AI–IoT-Integrated Vision-Based Localized Energy Management System for Smart Classrooms

[1]G. Mallikharjuna Rao, [2]Tummala Aravinda Babu

[1,2]Department of ECE, Chaitanya Bharathi Institute of Technology (CBIT), Gandipet, Hyderabad, Telangana - 500075, India.
[1]mallikarjunarao_ece@cbit.ac.in, ORCID ID: 0000-0003-0180-8733
[2]aravindababu_ece@cbit.ac.in, ORCID ID: 0000-0002-0721-6918

**Corresponding Author: G. Mallikharjuna Rao*

Abstract: Conventional classroom energy management systems often treat the entire room as a single occupancy zone, leaving fans and lights on even when only a few seating areas are occupied. This paper integrates a vision-based, localized energy management framework that uses the Ultralytics YOLOv8n model for real-time, zone-based occupancy detection in a 3×2 grid. The captured frames are resized to 640 × 480 pixels and filtered using a polygonal Region of Interest to analyze the classroom seating area. People are detected within localized bounding boxes using a bottom-centre point; perspective transformation accounts for the camera's inclination and mounting angle; and each person is assigned to one of the six zones arranged in a logical 3-column, 2-row grid. The multithreaded software architecture includes a dedicated camera acquisition thread that protects the frame buffer, and inference runs on every third frame to maintain real-time responsiveness. Based on the zone-wise occupancy, states are stabilized by state persistence, and ON/OFF decisions are made independently for the fan loads in each zone. The system was evaluated and implemented in a single classroom scenario. The results demonstrate reliable person detection, perspective-corrected zone assignment, and consistent localization of load-control behavior; however, the findings are limited to the evaluated configuration and architecture and are intended to support future extensions toward multi-classroom deployment.

Keywords: Smart classroom, AIoT, vision-based occupancy detection, energy optimization, energy efficiency, sustainable buildings.

1 INTRODUCTION

Smart infrastructure is expanding, and initiatives are promoting sustainable campuses by integrating Artificial Intelligence (AI) and the Internet of Things (IoT) into building energy management and automation systems [1,2]. Smart buildings use network sensor data, edge analytics, and real-time control to dynamically adjust lighting. HVAC systems for ventilation, heating, and air conditioning are controlled based on occupant behavior and environmental conditions [3]. Traditional classrooms, however, still rely on manual switching or motion detection, such as passive infrared or ultrasonic sensors, to activate lights and ventilation. These techniques generally provide binary information and lack spatial awareness, causing redundant energy use and suboptimal comfort [4,5].

Traditional occupancy sensors, such as ultrasonic and PIR sensors, cannot provide zone-wise detection or filter out accidental triggers from conversational motion, isolating human presence [6]. However, many BMSs rely on fixed or switch-logic-based schedules. This can cause dynamic adaptation to fail to reflect occupancy variations in the class schedule accurately, and it may not account for a person's transient movements [7]. This inadequacy is intensified in classrooms and lecture halls, where a single control system manages the entire space, wasting energy in unoccupied areas. Recent research has focused on vision-based occupancy detection. These include computer vision techniques and deep learning-based methods for extracting detailed spatial occupancy patterns from camera feeds [8], YOLO (You Only Look Once)-based deep learning approaches [9], R-CNN approaches with fast responses [10], and the SSD (Single Shot MultiBox Detector) [11], with accurate and real-time occupancy localization and counting [12].

This enables reactive energy management with minimal false triggering in smart buildings. These approaches also support building energy simulations that provide realistic occupancy profiles for HVAC demand prediction and lighting control. In parallel, IoT-enabled frameworks monitor occupancy information relative to electric load activation data and support energy optimization across spatial subsystems (e.g., zones) within the indoor environment [13]. These systems deploy real-time communication protocols, edge processing, and cloud connectivity to implement demand-based control loops for lighting, ventilation, and other loads. In this instance, real-time, persistent, bidirectional communication protocols over WebSockets can be used to support low-latency telemetry and manage the building's networks, enabling the exchange of state and messages among occupants' occupancy sensors, node controls, and remote dashboards [14].

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

To address these gaps, including the lack of zone-level occupancy information, control logic, and reactive switching behavior, this paper proposes a vision-based, zone-wise occupancy detection system that integrates load automation. Unlike surveillance camera systems, which are centered on safety or security, the proposed system focuses on energy efficiency. The proposed approach divides the classroom into predefined zones and uses spatial segmentation and real-time energy optimization to map loads to the corresponding zones. The system's architecture enables low-latency, real-time communication via WebSocket and uses a switching mechanism that suppresses unnecessary toggling when no human is present. This improves the system's robustness. The major contributions include deep learning models for human detection, accurate occupancy estimation, information localization, and zone-based spatial segmentation. The system is measurable and energy-saving, enabling automation via the Internet of Things for real-time activation of electric loads. The WebSocket-based architecture enables low-latency, real-time communication and monitors switching operations. Depending on the switching frequency, the delay can reduce unnecessary toggling and improve hardware reliability. The dashboard monitors occupancy, load status, and energy consumption.

Section 2 presents related literature, and Section 3 presents the system architecture, covering the hardware architecture, vision pipeline, perception-based spatial mapping, occupancy estimation, decision logic, communication layer, and software architecture. Section 4 presents the implementation details. Section 5 presents the real-time experimental setup for the predefined classroom deployment and evaluates the switching capability. Section 6 presents the results and discusses the proposed system.

2 Related Work

Classroom occupancy detection and monitoring support energy-efficient building automation. These concepts enable demand control of classroom lighting, ventilation, HVAC, and electrical loads. In building infrastructure, energy consumption and behavioral patterns significantly influence energy demand; introducing advanced occupancy intelligence in critical areas can reduce waste and improve operational efficiency. Both traditional and advanced data-driven approaches support classroom objectives.

2.1. Traditional Sensor-Based Occupancy Detection

Traditional classroom occupancy detection methods rely on local sensors such as ultrasonic, microwave, and infrared sensors to detect human presence. These sensor technologies integrate with existing lighting systems and meet growing demand for standard energy-saving devices. However, traditional occupancy-detection technologies have several limitations. Sensors are primarily limited to spatial regions and are vulnerable to false positives, which can affect occupant counts and reduce performance in indoor environments [1,2]. Studies have investigated PIR sensors for motion detection and control. However, energy-reduction results are suboptimal, and sensors may intermittently detect occupancy in zones or operate only under certain conditions. Researchers have applied machine learning algorithms to sensing data to estimate occupant counts and adjust motion-detection levels accordingly. These approaches often offer incremental improvements but lack the spatial resolution needed for zone-level load activation and the localization required for energy optimization in classroom settings [3].

2.2. Vision-Based Occupancy Detection Using Deep Learning

A vision-based occupancy detection system provides a detailed spatial distribution of occupancy and supports movement-pattern analysis [4]. Such systems demonstrate strong real-time performance for occupancy detection in indoor environments [5-7]. Recent publications, such as those in the *Journal of Building Engineering*, show that deep learning modules can effectively detect high-density occupancy across different environments, supporting real-time assessment for building automation and sustainability. Similarly, computer vision-based occupant detection can improve energy prediction accuracy and enable precise energy management across different strategies [8]. Deep learning algorithms, including YOLO-based object detection models, SSD, and posture-detection models, are CNN-based approaches for detecting objects and postures [9]-[12].

Unlike PIR- and RF-based approaches, vision-based detection can perform spatial segmentation. It also supports spatial analysis of environmental conditions, which can be divided into zones. The system can separate spatial regions by localizing the corresponding lighting and ventilation controls. This can significantly improve energy efficiency under partial occupancy. However, despite these advantages, vision-based approaches require careful management of computational resources and privacy in educational environments.

2.3. Wireless Infrastructure and Wi-Fi-Based Sensing

Researchers have explored wireless infrastructure to address the limitations of physical sensors for occupancy detection. Wi-Fi-based sensing leverages signal variations within IoT infrastructure to estimate human presence and occupancy patterns. These approaches demonstrate that detection accuracy depends on the communication environment. However, RF-based sensing methods are sensitive to environmental changes, multipath fading, and noisy signals, which reduce stability.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

In [13], a heterogeneous wireless network is introduced that uses IoT control to automate smart buildings while coordinating communication and device management. WebSocket communication supports low-latency remote monitoring and intelligent device control in buildings [14], while Wi-Fi sensing and ML-based localization can detect occupancy without requiring a database [15]. CSI-based person counting using Wi-Fi detection has also been investigated in indoor environments. However, few effective frameworks combine precise visual occupancy detection with zone-based control. Therefore, computer vision and AIoT-based automation offer intelligent solutions for smart classroom energy optimization [16]. The WebSocket communication protocol supports bidirectional communication in IoT applications [17].

2.4. Summary and Positioning of the Proposed Work

Traditional methodologies include wireless sensing networks and vision-based deep learning algorithms. Traditional methods rely on cost-effective, high-precision sensors in wireless sensor networks; while they reduce hardware overhead, their performance can vary with environmental conditions. The vision-based system offers higher spatial resolution and can identify occupants in predefined locations, improving detection performance while reducing computational requirements. The IoT-based infrastructure can control the energy management system in a practical and feasible manner. However, a primary research gap remains. Based on the reviewed literature, the integrated combination of zone-based occupancy detection, localized relay-based load activation, temporal stabilization, and WebSocket-based communication remains insufficiently explored. The proposed framework addresses this gap by integrating vision-based occupancy detection with spatial regions through IoT-enabled intelligent automation. The integration of AI, IoT, and edge computing is emerging as a promising approach to developing energy-efficient models for sustainable smart buildings. Collectively, these studies support the proposed smart classroom framework and motivate evaluating its energy-saving potential and sustainability benefits through edge AI deployment. The proposed framework also optimizes classroom energy use.

3 System Architecture

As shown in Fig. 1, the system architecture of the intelligent classroom automation system is divided into three main functional sections:

i. The computer vision and artificial intelligence process section
ii. An Internet of Things control module section
iii. A user interface section

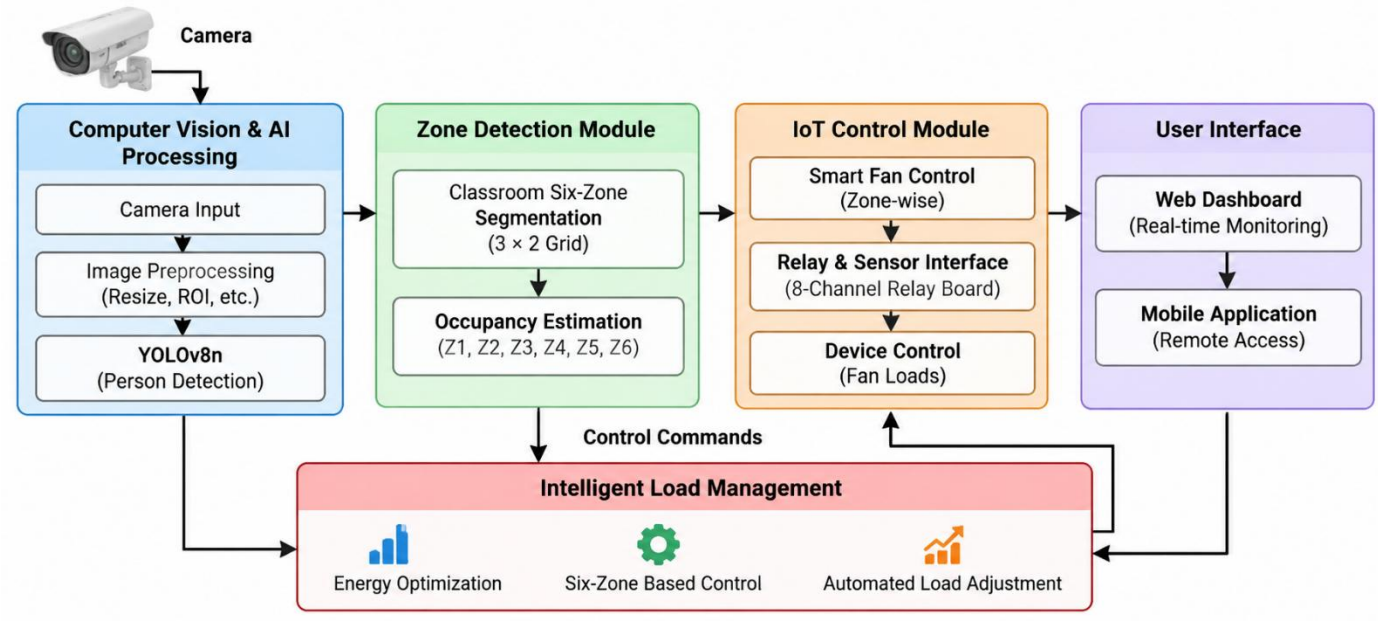


Fig. 1. System architecture of the AI-enabled smart classroom automation framework.

3.1. Hardware Architecture

The three hardware architecture sections work together to enable zone-specific, occupancy-based monitoring and control of electrical loads in indoor classroom environments.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

3.1.1. Computer Vision and Artificial Intelligence Processing Section

The computer vision and artificial intelligence processing section includes a camera module that captures visual data of the classroom. The camera input module produces video streams or real-time images to show occupancy activity within the monitored indoor environment. The object detection module uses trained detection systems to differentiate human entities from background objects with spatial precision. The artificial intelligence analysis module processes human-presence detection information from the object detection module. The artificial intelligence analysis module performs spatial mapping and classifies the classroom into predefined Zones, thereby calculating Zone-wise occupancy information.

3.1.2. IoT Control Module Section

The IoT control module receives occupancy information generated by the artificial intelligence analysis module. The IoT module section includes one or more embedded controllers interfaced with the relay and sensing units to enable electrical isolation and safeguard the activation of connected appliances. The IoT control module section manages smart lighting units, motor-driven appliances (such as fans and ventilation systems), and electrical loads installed in the classroom. The six-zone occupancy information is used to selectively switch the corresponding loads ON or OFF.

3.1.3. User Interface Section

The user interface section visualizes, monitors, and supervises the smart classroom automation system. The user interface dashboard enables remote access to the system through mobile apps and computing devices. Notifications from the IoT control module provide status updates and alerts, occupancy states, and appliance operating conditions.

3.2. Vision Pipeline

The vision pipeline continuously captures video frames from the classroom camera and resizes them to 640×480 pixels for stable imaging and resource savings. The resized frame then undergoes detection using the YOLOv8 Nano model from the Ultralytics framework for people detection. The recognition procedure detects and classifies people while ignoring irrelevant objects in the classroom zone. The detection confidence threshold is set to 0.4. In the application, however, to reduce the system's computational load, it uses only every third frame for real-time detection. The previous occupancy state is retained between inference frames to avoid unnecessary switching. The vision pipeline starts with camera frame acquisition and subsequent resizing. The resized frame is then processed using the YOLOv8n model.

3.3. Perspective-Based Spatial Mapping

The classroom camera is placed in an upper corner, and its viewing angle introduces perspective distortion. Thus, perspective distortion creates horizontal and vertical partitions in the image that do not correspond to the classroom's actual zones. To address this problem, the proposed system is based on spatial-mapping principles. First, the system selects the region of interest needed to identify the seating area's useful coverage and isolates outer spaces that do not require analysis because they do not affect classroom workload management. Four points are used to represent the classroom area in the image plane. For real-time human detection, the YOLOv8 Nano architecture is used, so detection accuracy and system interface speed are acceptable.

3.3.1. Method for Human Presence Detection and Intelligent Load Management

Each selected video frame is processed to detect individuals, obtain their coordinates, and filter detections based on confidence criteria. Each selected video frame is processed in approximately 32 milliseconds, supporting near-real-time operation. As indicated in Fig. 2, the developed method for detecting human presence in electric load monitoring begins with the first step. At this stage, the system continuously monitors the indoor environment. The system records the environment with a camera and then uses image processing to capture real-time images of activity in the monitored room.

After recording the images, the system analyzes them using computer vision and artificial intelligence. At this stage, the system processes the images to identify people and define parameters related to their presence. At this step, the system determines whether any person has been detected in the classroom. If no person is detected in a zone, the corresponding load remains OFF. The method modifies its control strategy and regulates operational states according to the loads to be controlled, from lighting apparatuses to motors and various other load types. The loads used in the specific operational context remain active during the process. If no people are present in the zone, the method moves to the next step. As a result, load-control activities are avoided, preventing unnecessary energy consumption; thus, switching loads is minimized and standby time is reduced. The method concludes with a stage that verifies the load-control actions were carried out appropriately at every moment according to the current occupancy needs.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

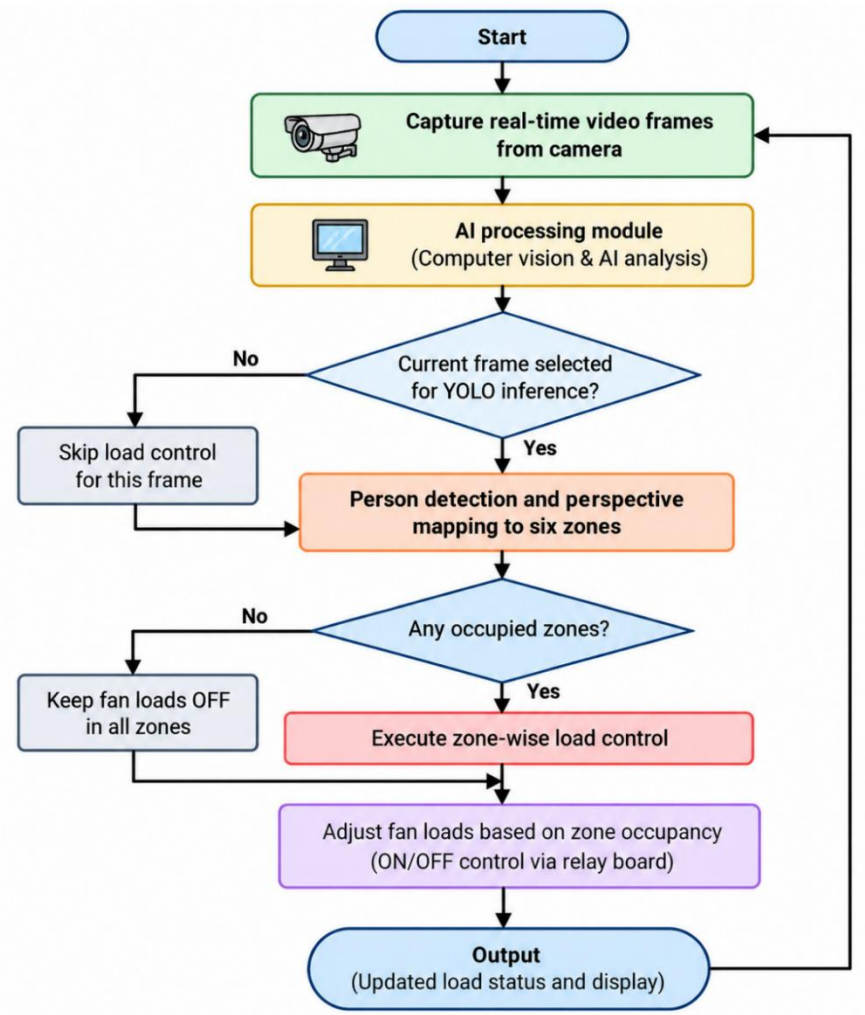


Fig. 2. Human presence detection and intelligent load management process.

3.3.2. Zone-Wise Spatial Segmentation: Classroom Spatial Recognition and Zone-based Occupancy Mapping

The classroom floor plan is predefined, with spatial zones divided into six portions. Each person is mapped to a specific Zone as shown in Fig. 3.

Perspective transformation → mapped coordinate → zone index → occupancy update

The floor plan region is selected manually in the classroom using four source points during the initial camera calibration stage for the classroom ROI:

$$P_s = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\} \quad (1)$$

where (x_1,y_1), (x_2,y_2), (x_3,y_3), and (x_4,y_4) represent the top-left, top-right, bottom-right, and bottom-left corners, respectively.

These corner points define the visible region bounded by the classroom benches within the camera's field of view. The four destination points are mapped to a rectangular plane surface:

$$P_d = \{(0,0), (W_d, 0), (W_d, H_d), (0, H_d)\} \quad (2)$$

where W_d is the width of the transformed classroom plane, and H_d is the height of the transformed classroom plane. The 3×3 homography matrix H satisfies:

$$\lambda\,[X, Y, 1]^{\mathrm{T}} = H\,[x, y, 1]^{\mathrm{T}} \quad (3)$$

(x,y) is the original camera image point, and (X, Y) is the rectified image point; λ is the homogeneous scale factor; and H is the 3×3 homography matrix given below.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

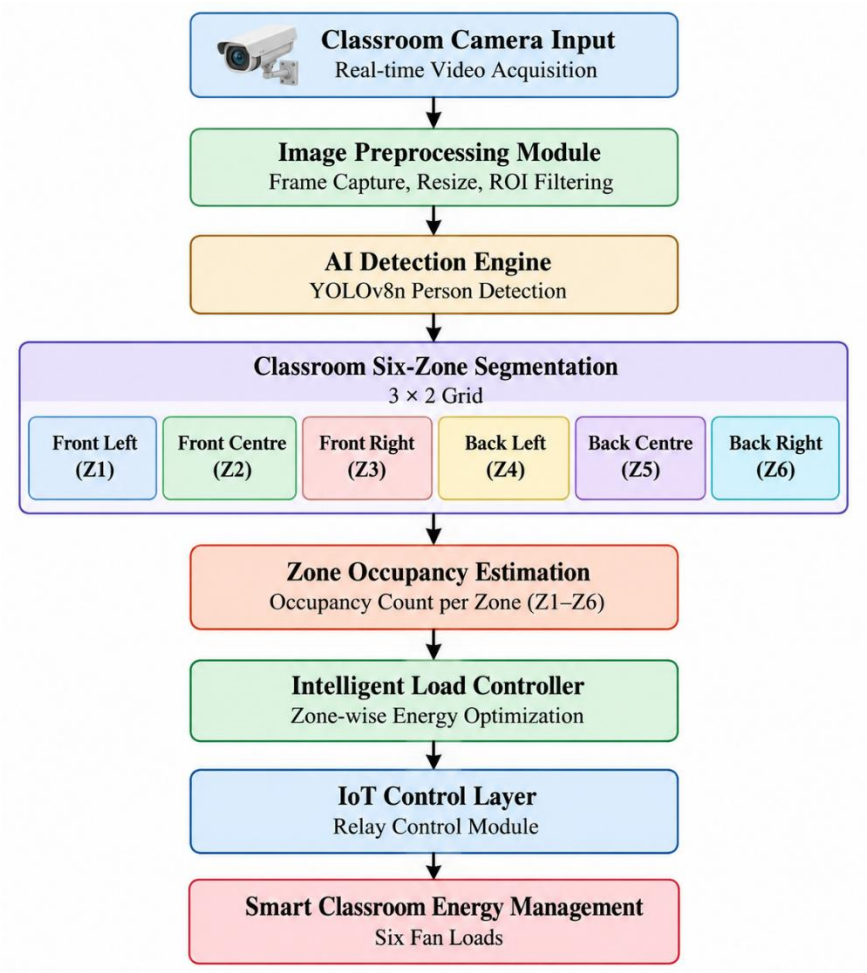


Fig. 3. Classroom spatial recognition and zone-wise occupancy segmentation.

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \tag{4}$$

After normalization by the third homogeneous coordinate, the transformed coordinates are:

$$X = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}$$
$$Y = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \tag{5}$$

The homography matrix maps image coordinates to the rectified classroom plane; the coordinates of each person's detection are transformed using homography H. The floor coordinates are assigned to 3 x 2 zone boundaries, forming six zones. The classroom has six Zones: front left, front centre, front right, back left, back centre, and back right. The zones define the spatial areas for monitoring intended load control relative to the classroom's facing orientation (located in front of the classroom), the reference zone for time spent on the teaching board or presentation area, and the spatial Zone alignment. To maintain a system's spatial orientation, the reference zone provides a constant reference during imaging analysis and occupancy mapping. The camera unit is mounted to capture a field of view that covers all defined coordinates. The camera unit continuously acquires visual data from the classroom and supplies the captured images to the computer vision system, which processes them using artificial intelligence. During human detection in the occupancy area, the system maps each detected person to one of the predefined zones using spatial coordinates. It localizes zone-wise occupancy information to support human-presence awareness and region-specific control of electric loads, such as lighting and ventilation. The detected person is represented by a bounding box $B_k = (x_k, y_k, w_k, h_k)$ where x_k, y_k denote the top-left coordinates and w_k, h_k denote the width and height of the bounding box. The bottom-centre point of the bounding box is used as the floor-contact reference point:

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

$$p_k = \begin{bmatrix} x_k + w_k/2 \\ y_k + h_k \\ 1 \end{bmatrix} \quad (6)$$

This point provides a more suitable reference for mapping a detected person to the classroom floor plane. The classroom floor plane point is transformed using the H matrix:

$$\tilde{p}'_k = H\, p_k \quad (7)$$

After normalizing, the mapped homogeneous coordinates are:

$$X_k = \frac{\tilde{x}'_k}{\tilde{w}'_k}, Y_k = \frac{\tilde{y}'_k}{\tilde{w}'_k} \quad (8)$$

where $\tilde{x}'_k$, $\tilde{y}'_k$, and $\tilde{w}'_k$ denote the three components of the transformed homogeneous coordinate $\tilde{p}'_k$.

3.4. Decision Logic

3.4.1. Camera Frame Acquisition

Initially, the camera captures the frame in the classroom at time t, which can be represented as follows:

$$I_t \in \mathbb{R}^{\wedge(H \times W \times C_{rgb})} \quad (9)$$

where H and W denote the frame height and width, respectively, and $C_{rgb} = 3$ denotes the three RGB channels.

3.4.2. Six-Zone Grid Definition

The classroom coverage region is divided into a grid of three columns and two rows (3 x 2):

$$C_z = 3,\ R_z = 2$$
$$Z = C_z \times R_z = 6 \quad (10)$$

Therefore, the total number of zones set in the classroom is considered as:

$$\mathcal{Z} = \{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6\} \quad (11)$$

3.4.3. YOLOv8n Detection

The YOLOv8n detection model is applied to every third frame of continuous images captured by the camera.

$$\gamma_t = 1, \text{ if } f_t \bmod 3 = 0$$
$$\gamma_t = 0, \text{ if } f_t \bmod 3 \neq 0 \quad (12)$$

where f_t denotes the frame index at time t, and γ_t indicates whether YOLOv8n inference is performed for that frame. When $\gamma_t = 1$, YOLOv8n inference is performed on the current frame. Therefore, the detector output is considered as:

$$\mathcal{D}_t = \{d_{1,t}, d_{2,t}, \ldots, d_{n,t}\} \quad (13)$$

Each detection is represented as a set of:

$$d_{k,t} = (x_{1,k}, y_{1,k}, x_{2,k}, y_{2,k}, s_k) \quad (14)$$

In this model, we consider only person classification, with a confidence threshold $s_k \geq 0.4$. Hence, the person-detection validation set is:

$$\mathcal{P}_t = \{d_{k,t} \in \mathcal{D}_t \mid c_k = \text{person},\ s_k \geq 0.4\}$$
$$N_t = |\mathcal{P}_t| \quad (15)$$

where c_k denotes the detected object class and s_k denotes its confidence score.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

3.4.4. Person Bottom-Centre Point and ROI Filtering

Detecting every person with a centroid bounding box is:

$$c_{x,k} = \frac{(x_{1,k} + x_{2,k})}{2}$$
$$c_{y,k} = y_{2,k} \quad (16)$$
$$p_k = [c_{x,k}, c_{y,k}]^T$$

where $c_{x,k}$ and $c_{y,k}$ denote the horizontal and vertical coordinates of the bottom-centre point of the k^{th} bounding box. A person detected from the zone, where the bottom-centre point lies within the ROI inside the polygon region R:

$$\chi_k(t) = \begin{cases} 1, if\ p_k \in \mathcal{R} \\ 0, if\ p_k \notin \mathcal{R} \end{cases} \quad (17)$$

where $\mathcal{R}$ denotes the predefined region of interest (ROI), and $\chi_k(t)$ is the ROI-membership indicator.

3.4.5. Perspective Transform for 3×2 Zone Mapping

The perspective transform region of a 3×2 grid is normalized to map the ROI polygon using a homography matrix H, which is considered as:

$$\lambda\,[u_k, v_k, 1]^T = H\,[c_{x,k}, c_{y,k}, 1]^T \quad (18)$$

where (u_k, v_k) are the mapped coordinates in the normalized grid space: $0 \le u_k < 3$, $0 \le v_k < 2$. The inside regions are clipped from the boundary region of the 3 x 2 grid, whose coordinates are indicated as:

$$u_k' = clip\,(u_k, 0, 2.99)$$
$$v_k' = clip\,(v_k, 0, 1.99) \quad (19)$$

where clip(*x*,a,b) limits *x* to the interval [a,b]. The a_k-column and b_k-row indices are:

$$a_k = floor(u_k')$$
$$b_k = floor(v_k') \quad (20)$$

where column $a_k \in \{0,1,2\}$ and row $b_k \in \{0,1\}$ as per 3 x 2 grid. Therefore, the zone index is considered as:

$$z_k = b_k C_z + a_k + 1 \quad (21)$$

Since column $C_z = 3$, it becomes: $z_k = 3b_k + a_k + 1$, $z_k \in \{1,2,3,4,5,6\}$.

3.4.6. Zone Occupancy and Binary Activity

The zone occupancy is indicated in binary form, where the zone count *i* at time t is considered as:

$$O_i(t) = \sum_{k=1}^{N_t} \chi_k(t).\mathbb{1}(z_k = i) \quad (22)$$

where N_t is the number of valid person detections at time *t*, and $\mathbb{1}(.)$ is an indicator function equal to 1 when the condition is true and 0 otherwise. The zone occupancy is then converted to a binary active or inactive state:

$$B_i(t) = \begin{cases} 1, & if\ O_i(t) > 0 \\ 0, & if\ O_i(t) = 0 \end{cases} \quad (23)$$

Hence, $B_i(t) = 1$ indicates that at least one person is detected in zone i.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

3.4.7. State Retention Between YOLO Frames

The YOLOv8n inference is applied once every three frames; therefore, the zone state is updated only on inference frames and retained on intermediate frames:

$$\widetilde{B}_{\mathrm{i}}(t) = \begin{cases} B_{\mathrm{i}}(t), & if\ \gamma_{\mathrm{t}} = 1 \\ \widetilde{B}_{\mathrm{i}}(t-1), & if\ \gamma_{\mathrm{t}} = 0 \end{cases} \tag{24}$$

Here, the intermediate frames will not be executed in the YOLOv8 model; in such cases, the previous state is retained.

3.4.8. Fan ON/OFF Control

The on/off control command for zone i in binary form makes fan control:

$$F_{\mathrm{i}}(t) = \widetilde{B}_{\mathrm{i}}(t) \tag{25}$$

Therefore, $F_{\mathrm{i}}(t) = 1$ indicates that the fan of zone i is ON, while $F_{\mathrm{i}}(t) = 0$ indicates that the fan of zone i is OFF. The number of active zones is denoted as:

$$\mathrm{N_{active}}\ (\mathrm{t}) = \sum_{i=1}^{6} \mathrm{F_i(t)} \tag{26}$$

3.4.9. Energy Model Derived from the Implemented ON/OFF Logic

From the fan status, we derive the energy model from the binary representation indicating the fan state. If the power $\mathrm{P_i}$ of each load/ fan denotes the instantaneous power consumption, then:

$$\mathrm{P_{actual}}\ (\mathrm{t}) = \sum_{i=1}^{6} \mathrm{P_i\ F_i(t)} \tag{27}$$

where $\mathrm{P_i}$ denotes the power rating/power consumption of the fan associated with zone i, and $\mathrm{F_i(t)} \in \{0,1\}$ denotes its operating state. Over T sampling intervals, the cumulative energy consumption is given by:

$$E_{\mathrm{actual}} = \sum_{\tau=1}^{\mathrm{T}} \Delta \mathrm{t} \sum_{i=1}^{6} \mathrm{P_i\ F_i}(\tau) \tag{28}$$

where T is the number of sampling intervals and Δt is the duration of each interval in hours after appropriate unit conversion. In the conventional approach, always-ON operation is considered as:

$$\mathrm{P_{always}} = \sum_{i=1}^{6} \mathrm{P_i}$$
$$E_{\mathrm{always}} = \sum_{\tau=1}^{\mathrm{T}} \Delta \mathrm{t} \sum_{i=1}^{6} \mathrm{P_i} \tag{29}$$

The energy-saving ratio is denoted as:

$$\eta = \frac{(E_{always} - E_{actual})}{E_{always}} \times 100\%$$
$$\eta = 1 - \frac{\sum_{t=1}^{\mathrm{T}} \Delta \mathrm{t} \sum_{i=1}^{6} \mathrm{P_i F_i(t)}}{\sum_{t=1}^{\mathrm{T}} \Delta \mathrm{t} \sum_{i=1}^{6} \mathrm{P_i}} \tag{30}$$

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

3.4.10. Compact Final Model

The complete implemented mathematical pipeline is:

$I_t \rightarrow$ YOLOv8n person detection $\rightarrow$ ROI filtering $\rightarrow$ perspective 3×2 zone mapping $\rightarrow$ binary zone activity $\rightarrow$ fan ON/OFF command

Compactly: $I_t \rightarrow \mathcal{P}_t \rightarrow O_i(t) \rightarrow B_i(t) \rightarrow F_i(t)$.

3.5. Communication Layer

The IoT subsystem is built around a microcontroller platform interfaced with a multi-channel relay board. The hardware board includes a microcontroller, an 8-channel electromechanical relay board, six relay channels assigned to the six classroom zones, lighting and ventilation interfaces, and a regulated power supply. Each Zone is mapped to a dedicated relay output.

1. An AI module sends a decision regarding occupancy state.
2. An IoT controller acquires the JSON-based order.
3. Switch circuits turn on or off electrical appliances.
4. Acknowledgment of the status is sent back to the server.

3.5.1. WebSocket-enabled real-time communication framework

TCP/IP-based WebSocket technology enables smooth data transactions. There are two fundamental communication channels. The first is the Control Channel, which handles interaction between the AI and IoT. The second is the Visualization Channel, which handles interaction between the AI and the dashboard. Messages sent through the channels include the message transmission time, zone ID, number of people present, command execution status, and feedback from the device used. As shown in Fig. 4, the intelligent classroom automation system uses the WebSocket protocol as a real-time communication framework to enable bidirectional data exchange between the artificial intelligence unit and I/O control devices, supporting low-latency communication.

The framework includes a processing server. It configures the computer vision and artificial intelligence algorithms for human detection within a spatial region and for zone-wise occupancy analysis. The AI processing server has trained inference models and decision-making logic needed for intelligent load management. A real-time communication interface establishes a WebSocket connection, enabling full-duplex communication between a processing server and one or more IoT controllers. WebSocket keeps a single connection open, allowing the system to stream occupancy data, status updates, and control signals in real time without the overhead of repeatedly establishing connections. Table 1 compares MQTT and WebSocket for the project.

Table 1. Project-oriented comparison between WebSocket and MQTT.

Parameter	**WebSocket in Proposed Project**	**MQTT-Based Alternative**
Project-based proposal	Real-time classroom control	Future cloud-based expansion
Architecture	Simple architecture within the local network	Distributed architecture and an IoT-based
Communication	Bidirectional	Publish and Subscribe, Broker-based operations
Video streaming	Continuous	Publisher based
Switching Circuit	Very fast	Delay due to the broker.
Dashboard access	Web browser	A webpage to be bridged with MQTT is required
Project deployment	Less complex	Highly complex
Scalability	Suitable for fewer classrooms	Architecture is designed for future multi-classroom deployment

The return path enables closed-loop control and improves system reliability by monitoring device state, operation status, and transmission execution. Data processing and storage are handled by an AI processor server to improve models over time, enhance control strategies, and store occupancy records, history, performance metrics, and logs. Using real-time WebSocket communication, the proposed system enables scalable, responsive coordination and synchronous decision-making between decision-making modules and the physical IoT infrastructure, ensuring occupancy awareness, energy efficiency, and in-classroom control of electrical loads. Fig. 5 shows the system's implementation: an automation framework that combines multiple cameras, visual sensing, and artificial intelligence. IoT controls and supervises. This is performed according to the classroom automation Zone.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

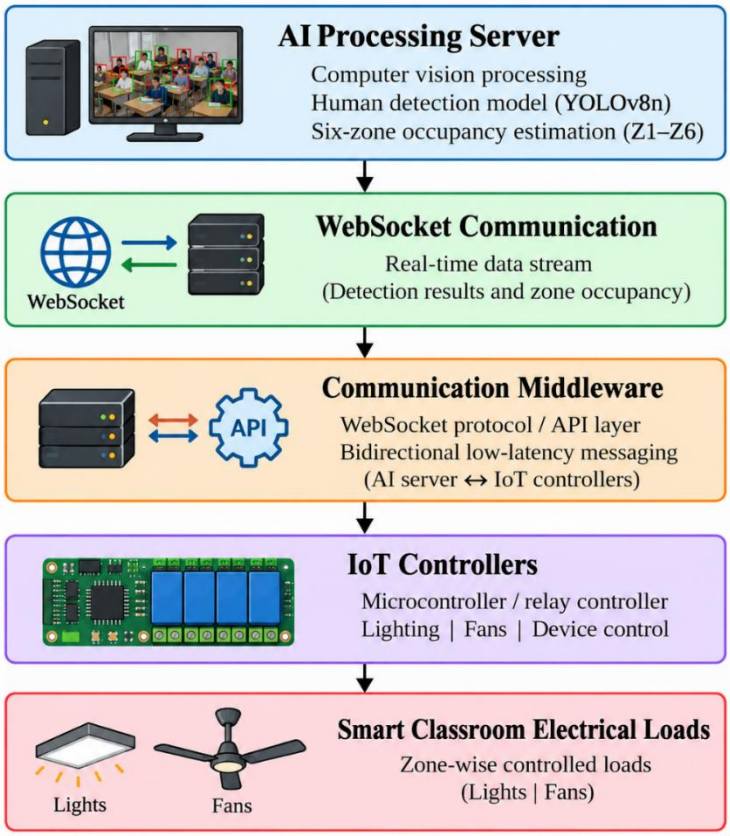


Fig. 4. WebSocket-enabled real-time communication framework.

The architecture can support multiple camera units for expanded deployment across classrooms. In the evaluated implementation, a single wide-angle camera was used to capture the classroom field of view. The camera is positioned to capture the field of view corresponding to the predefined spatial zones: front-left, front-right, back-left, and back-right. The camera streams the captured visual data to an AI processing server. The AI processing server handles image analysis, AI processing, and data processing to identify human presence and analyse spatial distribution. It also determines the zone occupancy. The server includes a computer vision executor that generates data based on real-time classroom activity. The system processes the occupancy information and communicates it to the IoT Control modules. It can serve as an automation framework for activation layers. The IoT control modules connect to relay and sensor units to monitor and ensure connected devices' safety. Signals are selectively sent to the control module. The system also includes smart light units. Additionally, the system incorporates motorized appliances, such as lights, fans, and ventilation systems, linked to specific classroom Zones. The Network Communication Interface supports wired and wireless data transfer between the AI processing server, IoT control modules, and user interface components.

The system enables real-time data transfer, updates, and synchronization across the framework layers. The User Interface Module includes a dashboard and, through the user-authorized mobile application, enables monitoring of Zone-wise occupancy status and appliance states. The system also provides alerts. The User Interface Module also allows configuration of system parameters, supervisory control, and automated decision-making, all of which can be overridden when needed. The Intelligent Processing Layer, IoT Activation Layer, Communication Infrastructure, Visual Sensing Layer, User Interface Layer, and Automation Framework enable a scalable, energy-efficient, real-time classroom management system.

3.5.2. Monitoring Dashboard and User Interaction Interface

As shown in Fig. 6, the control panel interface provides real-time zone occupancy detection, load ON/OFF status, energy consumption data, event logs, and historical trend analysis. The interface can be extended to multiple classrooms and buildings and can be accessed by any web browser or mobile device.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

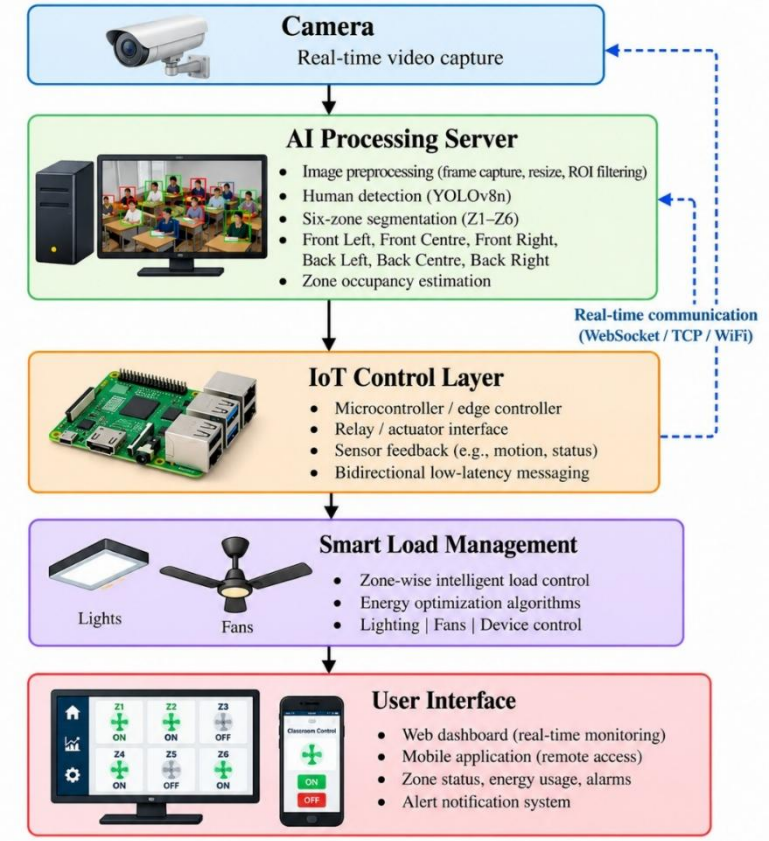


Fig. 5. End-to-end automation and communication framework for smart classroom.

3.5.3. System Scalability and Deployment Considerations

The architecture supports potential deployment across multiple classrooms, with scope for edge AI, cloud data aggregation, and integration with building management systems. The modular hardware design keeps costs low and requires little to no structural changes.

3.6. Software Architecture

The software architecture is designed to acquire and process real-time video. A separate camera acquisition thread continuously captures frames. The latest frames, stored in the shared frame buffer, are synchronized to ensure safe access during reading and updating. The processing threads then read the most recent frame from the shared buffer and run the vision pipeline. This separation between camera acquisition and frame-capture inference prevents delays during detection model execution. Camera buffer optimization is also used to reduce latency and avoid accumulating old frames.

The software architecture comprises camera acquisition threads, a shared frame buffer, a resizing module, YOLOv8n inference, ROI filtering, bottom-centre point extraction, perspective transformation, zone assignment, occupancy state updates, load-control decision-making, and real-time visualization modules. These multithreaded designs improve the framework's responsiveness and enable independent inference. Therefore, skipping inference on intermediate frames reduces the processing load while maintaining real-time classroom monitoring. Finally, the software architecture supports reliable occupancy estimation, zone-wise mapping and localization, and load-control decision-making in practice.

4 Implementation Details

The proposed classroom energy management system was implemented as a real-time, computer vision-based application using OpenCV, Python, and the Ultralytics YOLOv8 Nano framework. This implementation performs person detection, spatial zone mapping, occupancy estimation, and load-control decision-making on an edge computing platform using the classroom's live video stream. It aims to convert occupancy information into switching decisions with minimal delay. The object detection part uses the YOLOv8 Nano model, chosen to balance detection accuracy and processing resource usage.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

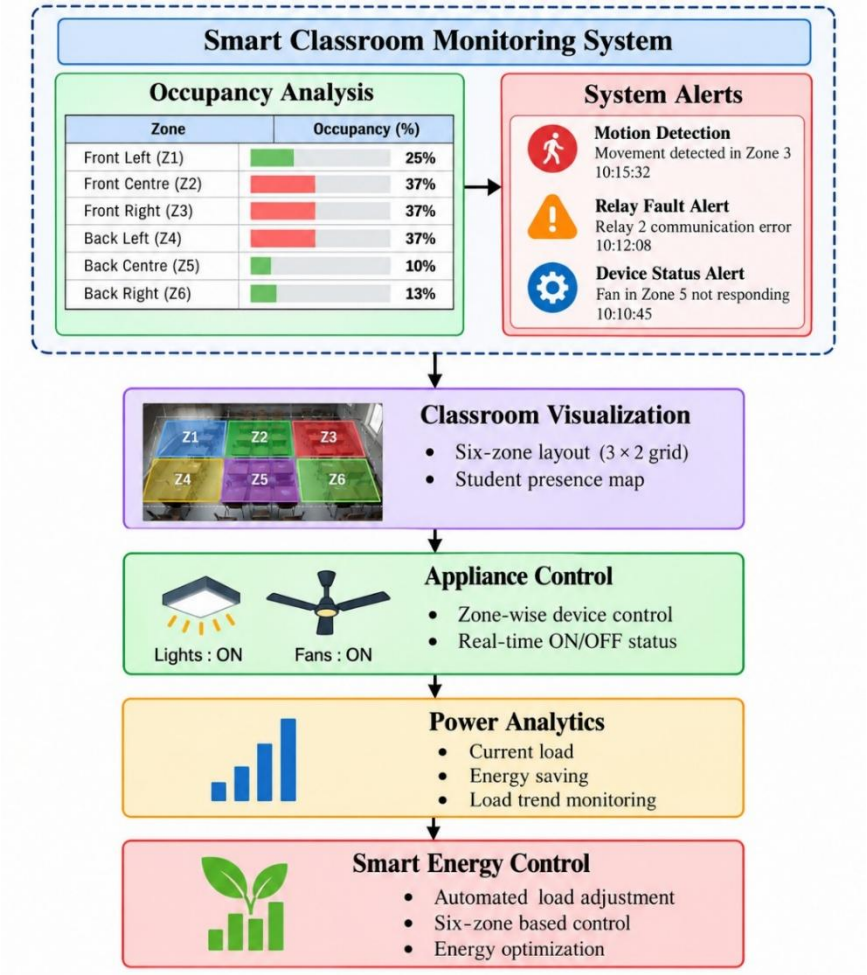


Fig. 6. Monitoring dashboard for real-time occupancy, appliance status, and energy consumption.

Thus, the model is configured to detect only people in the classroom and ignore irrelevant objects such as benches, boards, bags, walls, and furniture, reducing unnecessary processing and focusing on person detection. The input data is resized to 640 × 480 pixels for all processing steps to preserve spatial mapping and reduce computational requirements. The confidence threshold is set to 0.4 to filter out low-confidence detections. The detector considers only person detections with confidence scores that meet the threshold.

After detecting a person, the system finds the bounding box's bottom-centre point, checks it against the polygon, and discards detections outside the region of interest. The system assesses the validity of the bottom-centre point on a normalized classroom plan and uses homography to map it to one of the six zones of a 3 × 2 grid. The software implementation uses a multithreaded frame acquisition architecture. The camera thread continuously reads frames from the video source and stores the latest frame in the shared buffer, while the processing thread accesses the buffer for resizing, mapping, decision-making, and detection using a thread-safe access mechanism. This ensures that the pipeline processes the most recent frame rather than a delayed frame remaining in the camera queue.

The implementation pipeline is summarized as capturing the latest camera frame using the acquisition thread, resizing the frame to 640 × 480 pixels, performing YOLOv8 person detection on every third frame, applying confidence filtering with a threshold of 0.4, selecting only person detections, determining the bottom-centre point for each detected person, applying polygon-based region-of-interest filtering, transforming the valid points using perspective mapping, assigning each point to one of the six zones in the 3 × 2 grid, updating the zone occupancy vector, and applying temporal stabilization and load-control logic. Finally, the live frame is displayed with the boundaries, occupancy labels, and zone load status. This implementation ensures that the system follows the defined model configuration, preprocessing steps, spatial mapping techniques, and control logic. Table 2 clearly shows the model configuration, data preprocessing steps, spatial mapping techniques, control logic, and software structure used in the implementation.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

5 Experimental Setup

5.1. Experimental Environment and Deployment Setup

The artificial intelligence-based IoT smart classroom management system was tested in a limited experimental environment within an educational establishment. The classroom was divided into six zones based on the control system requirements. The study used a wide-angle ceiling-mounted video camera. The AI-based algorithms ran on a GPU-equipped computer, with an IoT system for supervision and an 8-channel relay board controlling six ceiling fans, each rated at 70 W; Wi-Fi provided wireless communication, as shown in Fig. 7.

5.2. Test Scenarios

The system was evaluated under several occupancy patterns using three operational scenarios:

Scenario 1: Full Occupancy. Occupants are present in all six zones for the duration of the session. This scenario represents maximum classroom use and serves as the basis for validating the system's performance.

Scenario 2: Partial Occupancy. In this case, occupants are present in only two of the six zones during the session. This scenario assesses the system's effectiveness in minimizing electricity use and recording information on vacant zones.

Scenario 3: Dynamic/Intermittent Occupancy reflects changing occupant behavior, meaning that occupants are absent from zones at times due to movement.

Table 2. Hardware and software configurations used for the implementation of the experimental setup.

Parameter	Configuration
Processor	Lenovo IdeaPad Slim 5 14IAH8 (laptop deployment)
Central Processing Unit	12th Gen Intel Core i5-12450H, 2.00 GHz
Graphics Processing Unit	Intel UHD Graphics, integrated
RAM	16.0 GB (15.7 GB usable)
Operating system	Windows 11
Python version	Python 3.12
OpenCV version	OpenCV version 4.10.0
Ultralytics version	Ultralytics v8.3.121
Detection model	YOLOv8 Nano
Object class detected	Only Person, COCO class ID 0
Camera source	Hikvision DS-2CD2132-I 3MP IR Mini IP Dome Camera
Input/capture resolution	640 × 480 pixels
Camera height (installation)	3.2 m
Camera viewing angle (installation)	30 degrees downward
Camera position (installation)	Upper corner of the classroom
Region of Interest (ROI) coordinates	(0,125), (315,55), (609,88), (51,415)
Region of Interest (ROI) type	Perspective zone of classroom region
Detection confidence threshold	0.4
Capturing frames	YOLO inference for every 3 captured frames
Total number of zones	6
Zone arrangement	3 columns × 2 rows
Person position reference	Based on the bottom-centre point of the person bounding box
Fan mapping	Zone 1→F_1, Zone 2→F_2, Zone 3→F_3, Zone 4→F_4, Zone 5→F_5, Zone 6→F_6
Fan condition (ON state)	The fan corresponding to the zone is ON if at least one person is detected in that zone.
Fan condition (OFF state)	The fan corresponding to a zone is OFF when no person is detected in that zone.
Current fan control implementation	Physical relay control for six fan zones during the reported experiments
Computer vision library	OpenCV
Detection framework	Ultralytics YOLO
Programming language	Python
Threading	Continuous camera frame acquisition thread
Frame buffer setting	cv2.CAP_PROP_BUFFERSIZE = 1
Display resolution	640 × 480 pixels

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

5.3. Performance Metrics

The following quantitative indicators were used: human detection accuracy (%), average inference time (ms), communication latency (ms), energy consumption (kWh), and switching frequency (toggles per hour).

6 Results and Discussions

The dataset was generated by recording videos under different classroom conditions, including an empty classroom, a fully occupied classroom, a partially occupied classroom, and students at different locations. We extracted frames periodically to avoid duplication. We then removed all blurred and low-quality frames. As a result, the dataset includes students sitting on different benches, students standing or moving, empty classroom images, corner camera views, and different illumination conditions. Based on this, the dataset was divided into training (70%), validation (20%), and test (10%) sets. The training images were used to train the model, and accuracy was assessed on unseen test images. The annotation strategy is to draw a bounding box for each visible student and not annotate empty benches. If a student is partially visible, annotate only the visible portion to avoid large boxes that cover multiple students and to maintain the class name throughout the dataset. The values are normalized to the range [0, 1] for YOLOv8 Nano-formatted annotations. The YOLOv8 Nano training configuration was selected for its fast object detection, which is suitable for real-time classroom occupancy monitoring. The YOLOv8n model was trained using an image size of 640 × 640 pixels, while the live inference pipeline used frames resized to 640 × 480 pixels.

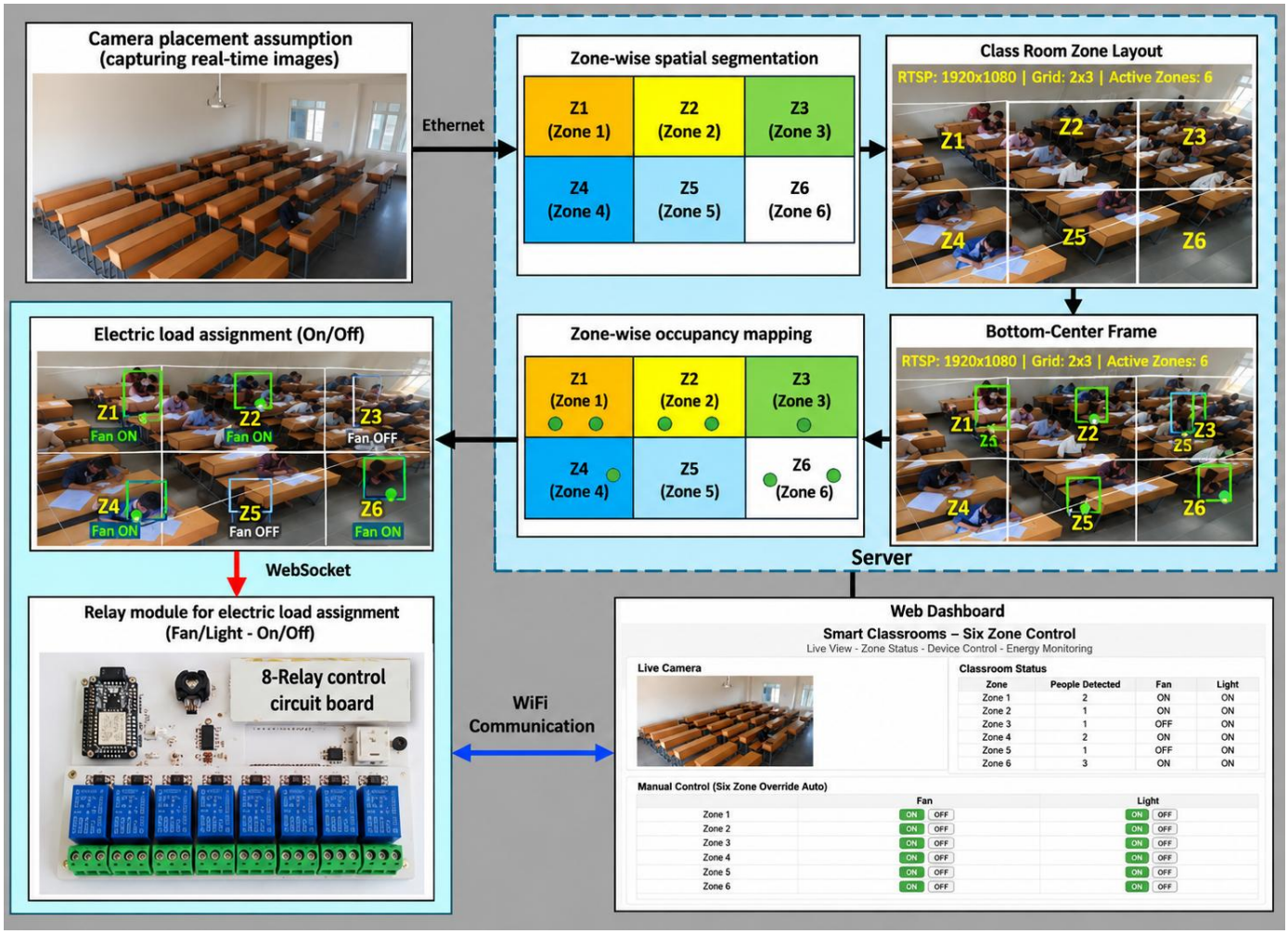


Fig. 7. The proposed AI–IoT framework: a classroom can detect occupancy using vision and control a smart fan in each zone.

In this context, high precision means fewer false alarms, which matters for classroom automation because false detections can unnecessarily switch lights and fans. The reported precision indicates reliable student detection. Recall measures the model's ability to detect students present in the classroom. This means fewer missed students in critical occupancy monitoring, as missed detections may cause classroom appliances to be switched off. These metrics were used to evaluate the effectiveness of the proposed YOLOv8 Nano-based classroom detection system. The performance was evaluated using precision, recall, F1-score, and mean average precision.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

The main components are true positives, false positives, false negatives, and true negatives. A true negative represents a background instance that is correctly identified as not containing a student. A false positive detects a non-student as a student. A false negative occurs when a student is present, but is not detected. The zone-wise occupancy detection and corresponding fan status obtained during the classroom experiment are illustrated in Fig. 8. As shown in Fig. 9, the confusion matrix was generated from an evaluation set comprising 60 student-present instances and 60 student-absent instances using the proposed YOLOv8 classroom occupancy detection method. The results show 56 true positives, 2 false positives, 4 false negatives, and 58 true negatives. These values yield a precision of 96.55%, recall of 93.33%, an F1-score of 94.92%, and an overall accuracy of 95%, indicating its efficiency for real-time energy management in smart classroom automation. Precision measures the proportion of correctly detected students among all detected students. The F1 score balances Precision and recall. It is mainly used to address both false positives and false negatives and to provide a single performance measure for comparison.

Intersection over union (IoU) measures the overlap between the predicted and ground-truth bounding boxes. So, an IoU of 1 indicates perfect overlap. In YOLOv8 Nano, it is typically considered correct when IoU is greater than or equal to 0.5. An IoU value of 0.5 is commonly used as a reference threshold for object-detection evaluation. Higher IoU values indicate greater overlap between the predicted and ground-truth bounding boxes. The reported mAP values were computed using bounding-box IoU matching rather than simple person-count accuracy. Average Precision is a standard metric for evaluating object detection performance. In this study, mAP@50 and mAP@50–95 were used to evaluate bounding-box detection performance.

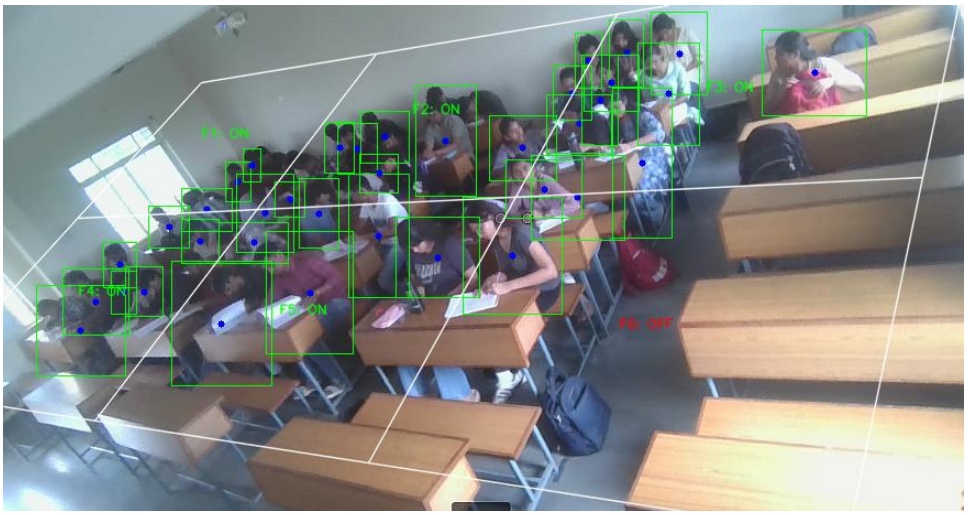


Fig. 8. Zone-wise occupancy detection and fan status in a classroom using YOLO-based person detection.

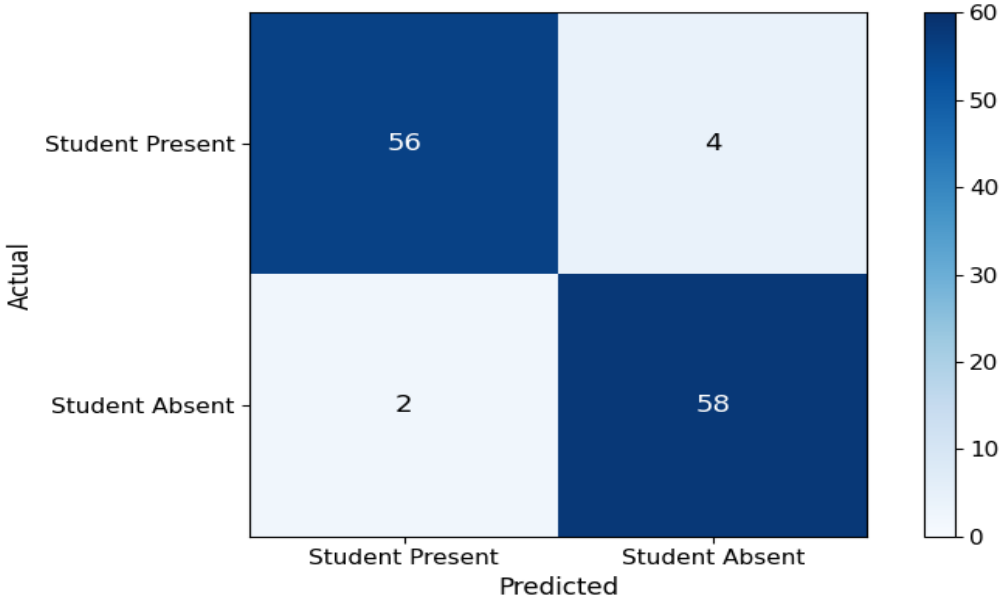


Fig. 9. Confusion matrix of YOLOv8 nano-based student detection model.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

Table 3. Result analysis through hardware setup.

Measure variable	Equation / Calculation	Result
Actual student-present instances	$TP + FN = 56 + 4$	60
True Positive (TP)	Correctly detected students' presence	56
False Negative (FN)	Present but predicted absent	4
False Positive (FP)	Absent but predicted present	2
True Negative (TN)	Absent and predicted absent	58
Total predicted present	$TP + FP = 56 + 2$	58
Precision	$\frac{TP}{TP + FP} \times 100 = \frac{56}{58} \times 100$	96.55%
Recall / Sensitivity	$\frac{TP}{TP + FN} \times 100 = \frac{56}{60} \times 100$	93.33%
F1-score	$\frac{2TP}{2TP + FP + FN} \times 100 = \frac{112}{118} \times 100$	94.92%
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN} \times 100 = \frac{56 + 58}{120} \times 100$	95.00%
Miss Rate / FNR	$\frac{FN}{TP + FN} \times 100 = \frac{4}{60} \times 100$	6.67%
False Discovery Rate	$\frac{FP}{TP + FP} \times 100 = \frac{2}{58} \times 100$	3.45%
Specificity	$\frac{TN}{TN + FP} \times 100 = \frac{58}{60} \times 100$	96.67%
Negative Predictive Value	$\frac{TN}{TN + FN} \times 100 = \frac{58}{62} \times 100$	93.55%
mAP@50	YOLO validation on annotated test set at IoU = 0.50	96.40%
mAP@50–95	YOLO COCO-style validation over IoU = 0.50–0.95	82.75%

Real-time performance metrics, as part of deployment performance, include frames per second, inference time, CPU utilization, memory usage, and power consumption. Frames per second indicate the number of frames processed per second. Inference time represents the processing time for a single frame. CPU utilization represents processor usage during detection, while memory usage refers to the RAM consumed by the model. Power consumption represents the energy required for continuous operation. For YOLOv8n, frames per second range from 10 to 20. The interfacing time is 50 to 100 milliseconds per frame. CPU usage is 40-70%, and memory usage is 500-900 MB.

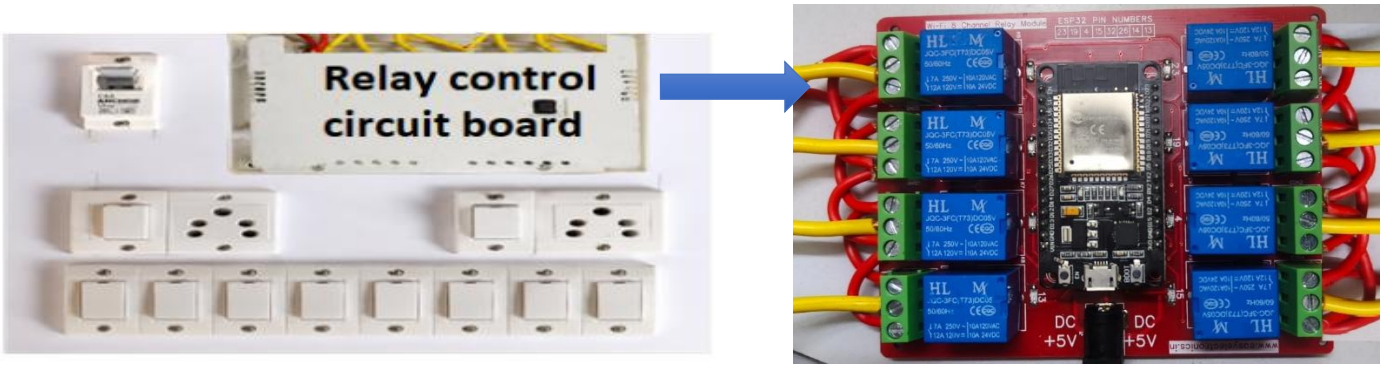
6.1. Communication Latency Analysis

The persistent WebSocket-based communication architecture supports real-time control and responsive load activation. Timestamp-based decision logging was used to measure the end-to-end latency between AI inference and IoT actuation. The measured average latency was approximately 18 ms.

6.2. Energy Consumption Evaluation

The six ceiling fans have a combined rated power of 420 W. The experiment compares the baseline all-ON configuration with the actual physical ON/OFF control of the fans through the relay hardware. The relay hardware used for fan-load control is shown in Fig. 10.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026



(a) (b)

Fig. 10. (a) Relay control circuit board for testing purposes on the switchboard, (b) 8-Relay control circuit board inside the case.

The experiment evaluates zone-wise occupancy using IoT-based load control and AI analytics in a classroom, with the resulting load control used to improve energy efficiency. The experiment compares 30 classroom sessions of 50–60 minutes each using six 70-W ceiling fans against a baseline in which all six fans operate throughout each session. As shown in Fig. 11, under zone-wise control, each session uses less energy than the baseline. This means the baseline energy was fixed at 0.35 kilowatt-hour for 50-minute sessions and 0.42 kilowatt-hour for 60-minute sessions; with this consumption control, it ranges from 0.25 to 0.39 kilowatt-hour depending on how many zones are empty during the sessions. As shown in Fig. 12, the energy savings range from 3% to 30%, with a mean of 15.7%. This indicates that energy savings depend on classroom occupancy conditions and not solely on the controller.

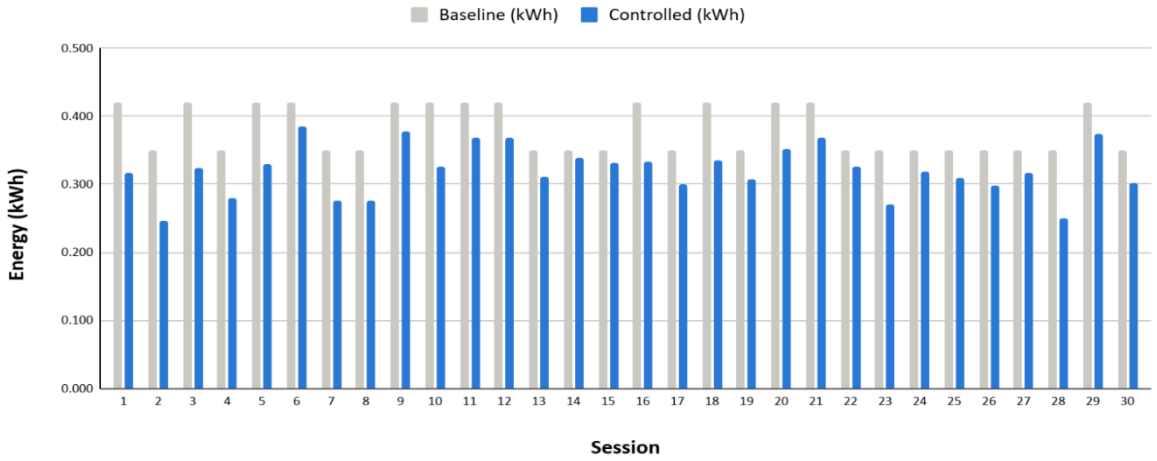


Fig. 11. Session-wise fan energy consumption: baseline (all fans ON) vs occupancy-based zone control (6 × 70 W fans).

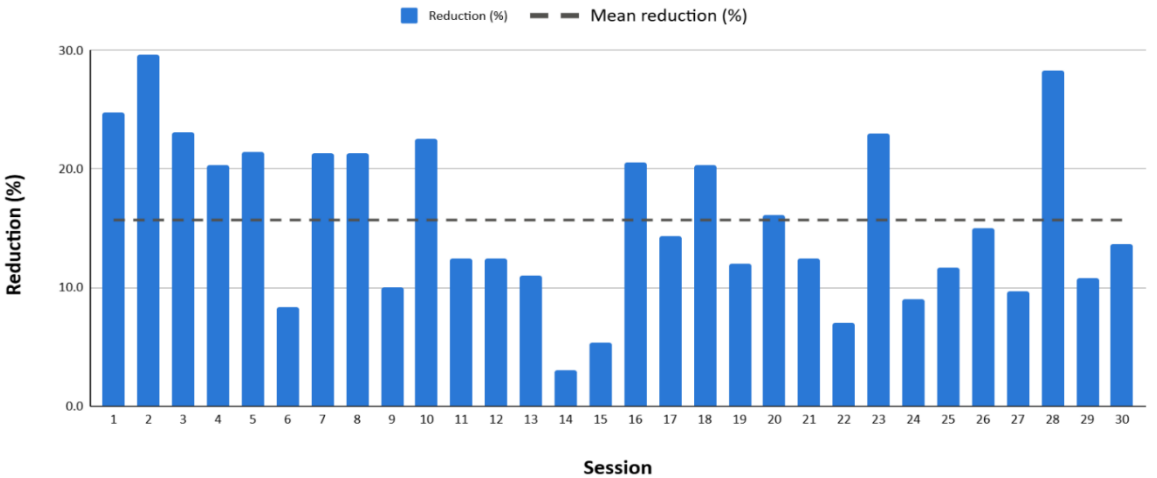


Fig. 12. Session-wise percentage energy reduction achieved by occupancy-based zone fan control

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

As shown in Fig. 13, cumulative energy over 30 sessions shows that total fan energy decreased from 11.41 kWh to 9.61 kWh, resulting in a saving of 1.80 kWh (15.8%). The curves move steadily apart, indicating that the savings accumulate consistently rather than resulting from a few unusual sessions. As shown in Fig. 14, energy reduction is strongly related to attendance, with a correlation coefficient of r = −0.82. Sessions with approximately 25 students achieve 25%–30% savings, whereas near-full classes with 55–60 students achieve only 3%–8% savings. As attendance increases, more zones are occupied, resulting in fewer fans being switched off.

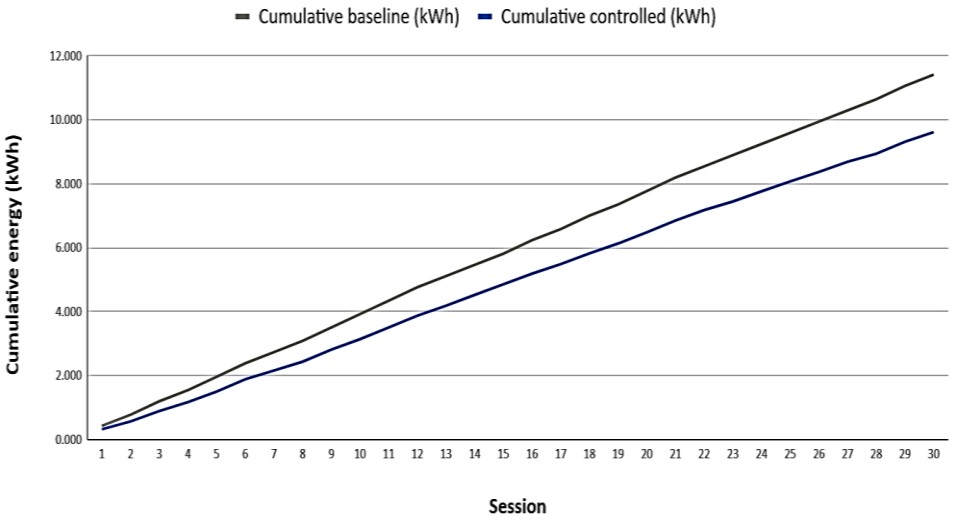


Fig. 13. Cumulative fan energy consumption over 30 sessions: baseline vs occupancy-based zone control.

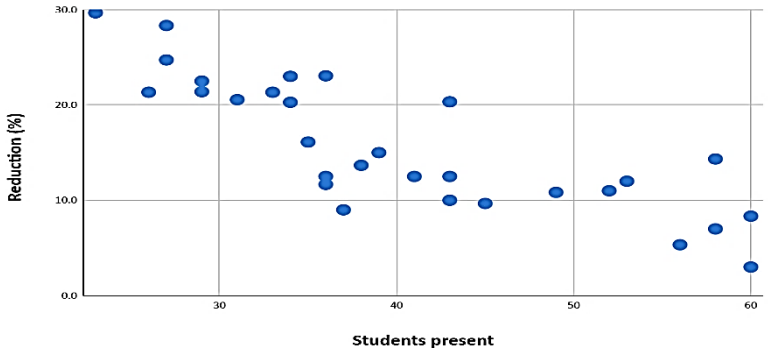


Fig. 14. Relationship between classroom attendance and energy reduction.

Table 4 shows the relevant relationships and energy-consumption details.

Table 4. Terms and relationships used for energy-consumption calculations.

Quantity	Equation	Unit
Total fan ON time	$T_{ON} = T_1 + T_2 + T_3 + T_4 + T_5 + T_6$	min
Baseline energy	$E_{base} = \frac{NPTs}{60 \times 1000}$	kWh
Controlled (system) energy	$E_{sys} = \frac{PT_{ON}}{60 \times 1000}$	kWh
Energy saved	$E_{saved} = E_{base} - E_{sys}$	kWh
Reduction per session	$R = \frac{E_{saved}}{E_{base}} \times 100$	%
Mean reduction	$\bar{R} = \frac{(R_1 + R_2 + \ldots + R_{30})}{30}$	%
Overall reduction	$R_{overall} = \frac{\sum E_{saved}}{\sum E_{base}} \times 100$	%
Zone ON share	$S_i = \frac{T_i}{T_s} \times 100$	%

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

Here,

- N is the number of fans/zones = 6.
- P is the rated power of one fan = 70 W.
- T_s is the total duration of the session, in minutes.
- T_i is the ON duration of the fan/zone i, in minutes.
- T_{ON} is the total fan ON time.

6.3. Summary of Experimental Findings

The proposed AIoT system was evaluated at the classroom scale by verifying zone-wise mapping and relay activation for the controlled loads. Energy consumption was calculated for 30 repeated sessions from the cumulative fan ON durations and the rated fan power, comparing an always-on configuration with the proposed zone-wise configuration. We assessed detection and control performance using inference latency, classification metrics, and end-to-end activation latency, as shown in Table 5 and Table 6.

Table 5. Live frame processing for person detection and zone assignment in a classroom.

Time	Frame	People inside classroom	Persons Detected	Actual zones	Predicted zones	Inference (ms)
10:05:12.103	3	3	3	1,2	1,2	31.4
10:05:12.207	6	4	4	1,2,5	1,2,5	32.6
10:05:12.310	9	2	2	3,6	3,6	30.9
10:05:12.414	12	5	4	1,2,4,5	1,2,5	33.2
10:05:12.518	15	3	3	2,3	2,3	31.8
10:05:12.621	18	1	2	5	5,6	32.1
10:05:12.725	21	0	0	None	None	34.0
10:05:12.828	24	4	4	1,4,5	1,4,5	30.7
10:05:12.932	27	2	2	2,6	2,6	31.9
10:05:13.036	30	3	3	3,5	3,5	31.4

Table 6. Experimental performance metrics of the proposed YOLOv8 nano-based smart classroom energy optimization system.

Metric	Result
Detection Accuracy	95%
Average Inference Time	32 ms
Communication Latency	18 ms
Energy Reduction	3%–30% (depending on classroom occupancy and zone-wise student distribution)

7 CONCLUSION

This paper presents a hardware-based experimental setup for an electrical load management system that integrates IoT and AI technologies in a classroom. The predefined zone-wise classroom spatial detection is configured to detect people in each zone and control the fan loads in the corresponding region. Testing indicates that, under partial occupancy, the proposed zone-wise control achieves energy savings of 3% to 30%. These results demonstrate the setup's potential for improving energy efficiency. Finally, the deep learning-based classroom spatial awareness results support further development of IoT-based human occupancy detection and analysis. These results demonstrate the system's potential for robustness, sustainability, and scalability; support energy optimization; and contribute to reduced electricity consumption.

FUNDING INFORMATION

This research was funded by Chaitanya Bharathi Institute of Technology (CBIT), Hyderabad, under the major research project titled “Design and Implementation of IoT-Enabled Smart Building Solutions for Enhanced Energy Efficiency,” as per Sanction Order No. CBIT/PROJ-IH/1018/ECE/D003/2024, dated 26 March 2024.

USE OF GENERATIVE ARTIFICIAL INTELLIGENCE

Generative artificial intelligence (GenAI) tools were used to assist with grammatical and language corrections and to improve the presentation of the manuscript, including the redrawing and visual enhancement of figures originally prepared by the authors. The original ideas, concepts, technical content, and interpretations are those of the authors. The authors reviewed and verified the AI-assisted content and remain fully responsible for the accuracy, originality, and integrity of the manuscript and its figures.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 4 | pp. 27-47 | Oct-Dec 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 23 Mar 2026 | Revised: 04 Oct 2026 | Accepted: 12 Oct 2026 | Published: 29 Oct 2026

ETHICS STATEMENT

The study used the existing CCTV camera infrastructure of the institution for classroom video acquisition. Permission to use the classroom CCTV footage for this study was obtained from the Head of the Institute. The collected data were used solely for the stated research purposes, and the authors take responsibility for the ethical use, handling, and reporting of the data.

STATEMENT OF CONFLICT OF INTERESTS

The authors declare no conflicts of interest related to this study.

LICENSING

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

REFERENCES

[1] P. Chaudhari, Y. Xiao, M. M.-C. Cheng, and T. Li, "Fundamentals, Algorithms, and Technologies of Occupancy Detection for Smart Buildings Using IoT Sensors," *Sensors*, vol. 24, no. 7, p. 2123, Mar. 2024, doi: 10.3390/s24072123.

[2] I. Khan, O. Zedadra, A. Guerrieri, and G. Spezzano, "Occupancy Prediction in IoT-Enabled Smart Buildings: Technologies, methods, and future directions," *Sensors*, vol. 24, no. 11, p. 3276, May 2024, doi: 10.3390/s24113276.

[3] K. M. Al-Obaidi, M. Hossain, N. a. M. Alduais, H. S. Al-Duais, H. Omrany, and A. Ghaffarianhoseini, "A review of Using IoT for Energy Efficient Buildings and Cities: A Built Environment Perspective," *Energies*, vol. 15, no. 16, p. 5991, Aug. 2022, doi: 10.3390/en15165991.

[4] W. Zhang, J. Li, P. W. Tien, and J. K. Calautit, "Vision-based occupancy detection in indoor environments: A comparison of standard RGB and thermal cameras," *Journal of Building Engineering*, vol. 114, p. 114215, Oct. 2025, doi: 10.1016/j.jobe.2025.114215.

[5] P. W. Tien, S. Wei, and J. Calautit, "A Computer Vision-Based Occupancy and Equipment Usage Detection Approach for reducing building energy demand," *Energies*, vol. 14, no. 1, p. 156, Dec. 2020, doi: 10.3390/en14010156.

[6] M. Jia, A. Komeily, Y. Wang, and R. S. Srinivasan, "Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications," *Automation in Construction*, vol. 101, pp. 111–126, Feb. 2019, doi: 10.1016/j.autcon.2019.01.023.

[7] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng, "Occupancy-driven energy management for smart building automation," *in Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, 2010, pp. 1–6, doi: 10.1145/1878431.1878433.

[8] W. Zhang, J. Calautit, P. W. Tien, Y. Wu, and S. Wei, "Deep learning models for vision-based occupancy detection in high occupancy buildings," *Journal of Building Engineering*, vol. 98, p. 111355, Nov. 2024, doi: 10.1016/j.jobe.2024.111355.

[9] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[10] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

[11] W. Liu et al., "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016*, Lecture Notes in Computer Science, vol. 9905, pp. 21–37, 2016, doi: 10.1007/978-3-319-46448-0_2.

[12] H. H. Nguyen, T. N. Ta, N. C. Nguyen, V. T. Bui, H. M. Pham and D. M. Nguyen, "YOLO Based Real-Time Human Detection for Smart Video Surveillance at the Edge," *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)*, Phu Quoc Island, Vietnam, 2021, pp. 439-444, doi: 10.1109/ICCE48956.2021.9352144.

[13] R. Djehaiche, S. Aidel, A. Sawalmeh, N. Saeed and A. H. Alenezi, "Adaptive Control of IoT/M2M Devices in Smart Buildings Using Heterogeneous Wireless Networks," *in IEEE Sensors Journal*, vol. 23, no. 7, pp. 7836-7849, 1 April 2023, doi: 10.1109/JSEN.2023.3247007.

[14] K. Ma and R. Sun, "Introducing WebSocket-Based Real-Time Monitoring System for remote intelligent buildings," *International Journal of Distributed Sensor Networks*, vol. 9, no. 12, p. 867693, Dec. 2013, doi: 10.1155/2013/867693.

[15] H. Choi, M. Fujimoto, T. Matsui, S. Misaki and K. Yasumoto, "Wi-CaL: WiFi Sensing and Machine Learning Based Device-Free Crowd Counting and Localization," *in IEEE Access*, vol. 10, pp. 24395-24410, 2022, doi: 10.1109/ACCESS.2022.3155812.

[16] A. I. Golam Mowla, R. Hossain and Asaduzzaman, "CSI-Based People Counting in WiFi Networks: Leveraging Occupancy Detection," *2024 IEEE International Conference on Computing, Applications and Systems (COMPAS)*, Cox's Bazar, Bangladesh, 2024, pp. 1-6, doi: 10.1109/COMPAS60761.2024.10796139.

[17] I. Fette and A. Melnikov, "The WebSocket Protocol," RFC 6455, Dec. 2011, doi: 10.17487/RFC6455.