



Real-Time Multi-Face Detection for Automated Attendance Management Using YOLOv8

Danang Arbian Sulisty¹, Lia Farokhah², Emilsa T. Bantug³, Rachel T. Alegado⁴

¹Department of informatics, Institut Teknologi dan Bisnis ASIA Malang, Malang, Indonesia

²Applied Data Science, Politeknik Artificial Intelligence Budi Mulia Dua, Yogyakarta, Indonesia

^{3,4}College of Information and Communications Technology, Nueva Ecija University of Science and Technology, Cabanatuan City, Philippines

*danangarbiana@gmail.com

Abstract

Accurate and efficient attendance management remains challenging in classroom environments because face-based systems are sensitive to illumination, pose variation, occlusion, and simultaneous student entry. This study proposes a YOLOv8n-based attendance management framework that integrates entropy-guided frame selection, automated face annotation, classroom-oriented augmentation, closed-set student identification, and temporal attendance validation. The dataset comprised 1,800 frames from six enrolled students and was expanded to 12,600 training images through augmentation applied exclusively to the training partition. The model achieved a precision of 0.620, recall of 0.692, mAP@0.5 of 0.566, and mAP@0.5:0.95 of 0.497 on the held-out test set. Model-only inference required 1.344 ms per image under the specified computational configuration, whereas the complete attendance pipeline, including frame acquisition, preprocessing, detection, rendering, and attendance logging, required an average of 0.122 s per frame, corresponding to approximately 8.2 FPS. Temporal confirmation and check-in/check-out rules prevented short-lived or repeated detections from generating duplicate attendance records. These findings demonstrate the feasibility of the framework as a proof of concept for controlled, closed-set classroom attendance involving a limited cohort. Further evaluation with more participants, independent recording sessions, and cross-environment data is required before broader deployment can be claimed.

Keywords: attendance management system; face detection; multi-face tracking; real-time detection; YOLOv8

How to Cite: D. A. Sulisty, L. Farokhah, E. T. Bantug, and R. T. Alegado, "Real-Time Multi-Face Detection for Automated Attendance Management Using YOLOv8", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 5, pp. 1381–1392, Oct. 2026, doi: [10.29207/resti.v10i5.7544](https://doi.org/10.29207/resti.v10i5.7544).

Permalink/DOI: <https://doi.org/10.29207/resti.v10i5.7544>

Received: February 2, 2026

Accepted: September 5, 2026

Available Online: October 8, 2026

This is an open-access article under the CC BY 4.0 License

Published by Ikatan Ahli Informatika Indonesia

1. Introduction

The digital information of academic administration has increased the demand for an attendance management system (AMS) that is accurate, efficient and scalable. Attendance management plays a key role in supporting the operational effectiveness of educational institutions, workplaces, and other organizations. Tracking the presence of individuals is imperative and is typically managed through roll calls, paper sheets, and ID card swipes. While these traditional methods are relatively simple, they are time-consuming, prone to errors, and vulnerable to proxy attendance, data inconsistency, and administrative inefficiency. As the educational institutions continue to operate at larger scales, the attendance process demands greater efficiency where

the use of manual records increasingly becomes inadequate [1], [2]

To overcome these challenges, various automated AMS approaches have been introduced, including Radio Frequency Identification (RFID) cards [3], [4], fingerprint biometrics [5], [6], QR code scanning [7], [8], and mobile-based check-ins [9]. The methods have been summarized in Table 1. These approaches improve automation to some extent; however, they still depend on user participation, device availability, or physical interaction during the attendance process. Nevertheless, these systems also entail some limitations. RFID and fingerprinting systems demand physical interaction that can lead to longer waiting queues and hygiene concerns. The QR codes and mobile-based systems require users to have functioning

devices and can also be vulnerable to security breaches and malfunctions. In addition, these methods can introduce various levels of intervention that can lead to potential bottlenecks during peak times [3], [8-11]. Therefore, a contactless and automated attendance mechanism that can operate passively in real classroom settings remains highly desirable.

Table 1. Existing Approaches Adopted in Attendance Management

Approach	Technique	Performance	Gaps Identified
RFID Cards	Card scanning	Reliable under supervision	Physical dependency, prone to misuse
Fingerprint Biometrics	Biometric fingerprint sensors	High accuracy under ideal conditions	Requires contact, hygiene issues
QR Code Scanning	Printed or device-displayed QR codes	Fast and low-cost	Vulnerable to duplication, needs device
Face Detection (2-Stage)	Separate detection and recognition	High recognition rate	Slow, two-step complexity
YOLO-based Face Detection	Object detection approach	Fast real-time detection	Needs separate recognition module

Face detection and recognition-based systems have emerged as promising solutions that offer contactless alternatives and overcome the shortcomings associated with the presented methods [12], [13]. These methods leverage unique facial features to identify individuals from camera streams without requiring physical interaction. This eliminates the need for physical touch or interaction with the devices. The process incorporates face detection from the video streams and matching routines from a pre-established database for the confirmation of identity and attendance marking [14]. As a result, face-based attendance systems are attractive for classroom environments because they enable simultaneous monitoring of multiple students in a non-intrusive manner.

Several methods have been used in the past for facial recognition and the associated attendance management systems. Some of the early systems made use of Haar cascades or Local Binary Pattern Histograms (LBPH) for face detection and recognition [15-17]. These systems offered lightweight solutions with acceptable performance under controlled conditions. In recent times, deep learning methods have emerged that significantly improve the robustness and accuracy of face detection. The models involving Convolutional Neural Networks (CNNs), MobileNet, and ResNet have been widely applied [18], [19]. Object detection models such as YOLO (You Only Look Once) and Single Shot MultiBox Detector (SSD) [20] have further improved face detection by formulating facial localization as a real-time object detection problem. However, despite the involvement of these advanced systems, some challenges still exist. Many systems make use of a two-stage pipeline in which face detection and identity recognition are handled separately. Such a separate approach leads to increased latency, increased

computational complexity, and negatively impacts real-time performance. In addition, many systems struggle to capture the variations in the pose, lighting, partial occlusion, and crowded environments. More importantly, prior attendance studies often emphasize detection or recognition accuracy only, while giving limited attention to attendance log integrity, such as duplicate records, repeated appearances of the same student, and the reliability of the decision logic in continuous video streams.

To address these limitations, this study proposes a YOLOv8-based real-time multi-face attendance framework that integrates face detection, student identification, and attendance validation within a single workflow. Unlike conventional systems that rely on contact-based interaction or separate detection and recognition stages, YOLOv8 is adapted as a unified detection-classification model that directly assigns detected faces to predefined student classes, improving real-time efficiency. The novelty lies in this integration strategy rather than a new detector architecture. Operational robustness is supported through informative frame filtering, dataset balancing, class-wise augmentation, and temporal attendance verification to handle pose, illumination, occlusion, repeated appearances, and duplicate records. The main contributions are: a unified framework for simultaneous face detection and student identification, a classroom-oriented data processing and augmentation strategy, check-in/check-out logic that suppresses redundant attendance entries and experimental validation in realistic multi-participant classroom scenarios demonstrating the feasibility of real-time automated attendance management.

Recent studies on automated attendance management systems (AMS) indicate a transition from conventional attendance technologies toward vision-based and deep learning-driven solutions. However, the literature is not only diverse in terms of algorithms, but also fragmented in terms of robustness, computational efficiency, and practical deployment. Therefore, a meaningful review should not merely list prior studies, but critically compare the technical trade-offs of the dominant approaches. A first stream of studies relies on classical computer vision pipelines, particularly combinations of Haar cascade, LBPH, Eigenfaces, Fisherfaces, and Viola-Jones [15-18], [21],

These methods remain attractive because they are relatively lightweight, easy to implement, and computationally affordable for small-scale attendance systems. However, their reported performance is highly dependent on controlled conditions. Across multiple studies, the same limitations recur: reduced robustness under non-frontal head poses, poor illumination, background clutter, facial expression changes, and partial occlusion. This suggests that the core weakness of classical pipelines is not simply lower accuracy, but limited generalization when classroom conditions become visually complex and dynamic.

A second stream adopts deep learning-based recognition or detection models to overcome the representational limitations of handcrafted features. CNN-based approaches and pre-trained deep learning detectors generally report better recognition capability than classical methods, especially under moderate visual variation [19], [20], [22]. Nevertheless, this improvement is often accompanied by higher computational cost, increased inference latency, or degraded performance when multiple faces must be processed simultaneously in real time. This issue becomes more critical in classroom scenarios, where the number of subjects per frame is larger, and face scale varies substantially. In parallel, deployment-oriented studies using IoT or web frameworks [23], [24] improve system modularity and institutional usability, but they do not fundamentally eliminate the visual and computational limitations of the underlying face-analysis modules. For example, IoT-based systems remain dependent on network stability, while web-based deployments still suffer when the recognition backbone is not robust to low-light or unconstrained classroom conditions.

A third stream focuses on portable or alternative recognition pipelines, including mobile implementations, SVM-HOG combinations, and multi-step preprocessing-based systems [21], [25], [26]. These approaches demonstrate that attendance automation can be adapted to different platforms and operational contexts. However, portability and deployment flexibility often come at the expense of lower accuracy, weaker robustness to profile faces or distance variation, and greater sensitivity to noisy backgrounds. Importantly, many of these systems still follow a sequential detect-then-recognize structure, which creates additional processing overhead and makes end-to-end real-time performance harder to sustain in crowded scenes.

Table 2 summarizes representative studies discussed in this review and highlights an important pattern: despite differences in algorithms and platforms, most existing systems still exhibit a trade-off between robustness, speed, and deployment practicality. Taken together, the literature reveals three key gaps: many attendance systems still separate face detection and identity recognition, increasing latency and error propagation; evaluations are often conducted under limited conditions rather than realistic classrooms with variations in pose, illumination, distance, occlusion, and multiple students; and most studies emphasize detection or recognition accuracy while overlooking attendance-record integrity, including repeated appearances, duplicate entries, and check-in/check-out validation. To address these limitations, this study proposes an integrated YOLOv8-based framework for real-time multi-face classroom attendance. Its contribution lies not in a new detector architecture, but in combining multi-face detection, student identification, and attendance validation within a single low-latency workflow designed to improve robustness

and generate reliable, non-redundant attendance records.

Table 2. Comparative Summary of Representative Face-Based Attendance Management Systems

Author	Method Adopted	Performance Outcomes	Identified Gaps
R. K. Shukla et al. 2024	CNN-based face detection and recognition	Good accuracy in normal conditions	Struggles with varying lighting and occlusions
Chittibomma, Surapaneni, and Maruboina 2024	Haar Cascade + LBPH face recognition	Real-time attendance marking	Sensitive to background clutter and non-frontal faces
D. Shukla, Kumari, and Bhargavi 2024	OpenCV + Haar Cascade + LBPH	Moderate performance in ideal conditions	Impacted by lighting and facial expressions
Antipona et al. 2024	Hybrid of Haar Cascade, Eigenfaces, and Fisherfaces	Improved recognition compared to single methods	Still limited under partial occlusion and angled faces
Pillai and Chellappan 2025	Deep learning with OpenCV and Dlib	Effective face detection and recognition	Requires high computational resources, struggles with multiple faces
Bavaskar 2024	IoT-integrated face recognition	Real-time recording and integration	Vulnerable to internet disruptions
Vasanthi, Vishali, and Zaiba 2025	Django-based web system with face recognition modules	Practical modular deployment	Lower performance in low-light environments
Alaftekin, Pacal, and Cicek 2024	Pre-trained deep learning face detectors + CNN recognizers	High detection rate in real-time	Latency issues in large classrooms
Rao et al. 2024	Haar Cascade + LBPH recognition	Simplicity and easy deployment	Limited robustness against pose and lighting variations
Amri 2025	Mobile network-based face recognition	Portability and flexibility	Reduced accuracy compared to server-based systems
Daly, Khemakhem, and Ltifi 2024	SVM on HOG features	Good frontal face recognition	Poor performance on profile faces
Elkari et al. 2024	Haar cascade for multi-face detection + LBPH for recognition	Works well for close faces	Poor detection in crowded and far-off conditions
Ramalingam and Jamal 2024	Histogram Normalization + Viola-Jones detection + Eigenfaces	Automatic attendance from video feeds	Challenges in noisy/complex backgrounds

2. Methods

This study proposes a proof-of-concept automated attendance pipeline for real-time multi-face detection

and student identification within a controlled classroom cohort. Videos and images are processed using entropy-guided frame selection, followed by Haar cascade-based face annotation and conversion to YOLO-compatible labels. Classroom-oriented augmentation, including zoom-out, brightness adjustment, and perspective transformation, is then applied to improve intra-class robustness. The enriched dataset is used to train a YOLOv8 model for multi-person identification, while real-time inference records enrolled students with timestamps. The overall pipeline evaluates the feasibility of integrated detection, identification, and attendance logging under classroom-like surveillance conditions.

The proposed attendance management system follows a structured and fully automated pipeline, as illustrated in Figure 1. The process is designed to simulate realistic classroom conditions and ensure robust identification through several interconnected stages. These stages include entropy-guided frame extraction, automated annotation, and a specialized attendance recording logic to handle dynamic student behavior. To improve reproducibility, each methodological stage is described together with its variables, parameter settings, and implementation role in the overall pipeline.

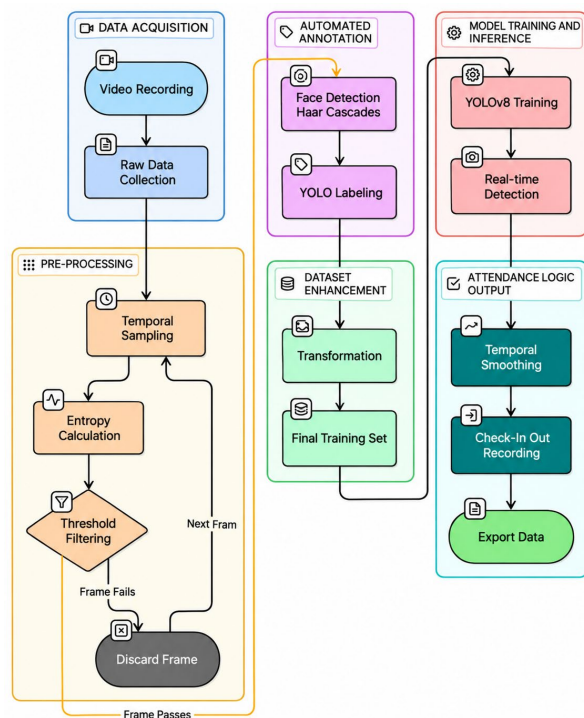


Figure 1. Research Flowchart

2.1 Dataset Description

The dataset used in this study was collected in a controlled classroom environment designed to emulate routine attendance scenarios at the entrance area. The data consist of high-resolution videos and still images acquired from a fixed surveillance camera. The camera was mounted at eye level and facing the entrance of the classroom. Each of the video sequences involves a

single or multiple individuals walking into the frame from varying distances that range between 2 to 6 meters. The subjects have exhibited natural variations in terms of gait, speed, and posture, with a deliberate inclusion of both the frontal and oblique orientations relative to the camera. This aims to simulate real-life behaviors and unpredictability.

The videos have been recorded in Full HD (1920x1080 pixels) at 30 frames per second. This enables the system to capture the fine-grained temporal and spatial variations in the subject motion. The individuals start approaching the camera from various angles ranging between 0° (directly facing the camera) to approximately $\pm 45^\circ$. Occasionally, the individual's glance sideways, downward, or away from the lens. The walking styles vary in terms of stride length, head pose, and pace (slow, normal, brisk). The inter-subject distances within the multi-person frames range from 0.5 meters to 1.5 meters, thereby creating the overlap scenarios and occlusion effects that can add complexity to the detection and identification. Still images were sampled from the video sequences and annotated with diverse poses, lighting variations and mild occlusion patterns.

These variations were intentionally introduced to increase intra-subject appearance diversity; however, they should not be interpreted as a substitute for broader subject diversity. The raw dataset consisted of 1,800 extracted frames collected from classroom recordings of six enrolled subjects. Through classroom-oriented augmentation, the number of training images increased to 12,600. It should be noted that augmentation increased appearance variability but did not increase the number of unique identities. The summary of the dataset is presented in Table 3.

Table 3. Dataset Characteristics

Attribute	Description
Video Resolution	1920 × 1080 (Full HD)
Frame Rate	30 FPS
Raw Image/frame count	1,800
Augmented Image/frame count	12,600+
Number of Enrolled Subjects	6 distinct individuals
Camera Placement	Eye-level, frontal view (indoor fixed setup)
Subject Distance Range	2 – 6 meters from camera
Facing Angles	0° to $\pm 45^\circ$ (frontal to oblique)
Walking Styles	Slow, Normal, Brisk (natural gait variation)
Inter-Subject Distance	0.5 – 1.5 meters (for multi-person scenes)
Lighting Conditions	Indoor ambient lighting with minor variability
Image Augmentation	Zoom, Brightness, and Perspective shifts
Annotation Method	Haar-based automated YOLO labeling
Study Scope	Closed-set identification for classroom attendance
Claimed Generalization	Limited to controlled indoor classroom conditions

A key methodological consideration is that the dataset contains only six distinct individuals. Therefore, the

present task should be interpreted as a closed-set identification problem involving a limited classroom cohort. Although augmentation increased the number of training samples, it did not increase the number of unique identities. Accordingly, this study does not claim broad demographic generalization; rather, it evaluates whether a YOLOv8-based pipeline can operate reliably for real-time attendance within a small, predefined group under controlled indoor conditions. This limitation is explicitly acknowledged because subject diversity remains an important factor for deployment to larger populations.

2.2 Automated Frame Extraction

To reduce manual sampling effort while preserving informative training content, an automated frame extraction procedure was applied to each recorded video. Instead of retaining all frames uniformly, the method selects candidate frames based on temporal stride and image entropy so that redundant or low-information frames are removed before annotation. This helps retain frames containing useful visual cues such as facial orientation, movement phase, occlusion presence and mild occlusion. Each of the videos has been sampled at a fixed temporal stride, denoted as Δf . Extra checks have been incorporated to guarantee that no two frames are near duplicates in structure or spatial entropy. Let $V = \{v_t\}_{t=1}^T$ denote a video sequence consisting of T total frames. The subset of extracted frames, $F = \{f_i\}_{i=1}^N$, is derived according to Equation 1.

$$f_i = v_{i \cdot \Delta f}, \quad \text{where } i \cdot \Delta f \leq T, \quad i \in \mathbb{N} \quad (1)$$

We additionally compute the Shannon entropy $H(f)$ of each frame f_t to estimate its information content.

$$H(f) = -\sum_{x \in \mathcal{X}} P(x) \log P(x) \quad (2)$$

In Equation 2, $P(x)$ denotes the pixel intensity distribution of the frame, and x is the probability of observing intensity level. The frames with entropy values that are below a predefined threshold were discarded to avoid blurred, overly dark or visually uninformative samples. The entropy threshold was determined empirically through exploratory inspection of the training videos and then fixed for all experiments. The exact value used in this study was 0.5. Similarly, the temporal stride was set to 1 frame to balance sample diversity and redundancy reduction. At a frame rate of 30 FPS, this corresponds to sampling one frame every 0.33 seconds.

To reduce near-duplicate sampling, redundancy was controlled through the joint use of stride-based sampling and entropy filtering, with consecutive candidate frames exhibiting very limited structural change treated as redundant and discarded. These parameter settings were fixed across all experiments to ensure reproducibility, forming the core of the automated extraction procedure. The algorithm processes a video sequence using a specified sampling stride and an entropy threshold, beginning by initializing an empty set to store the selected frames. As it systematically scans through the video advancing by

the given stride, each sampled frame is first converted to grayscale to compute its intensity histogram and corresponding Shannon entropy. This entropy value acts as a measure of spatial detail; a frame is retained and appended to the extracted set only if its entropy meets or exceeds the predetermined threshold, while low-information frames are entirely skipped. Once the entire video sequence has been processed through this evaluation loop, the algorithm concludes by outputting the final collection of highly informative frames, which is then forwarded to the annotation stage.

2.3 Face Detection and Annotation

After frame extraction, each selected frame was processed using Haar cascade face detectors to automatically localize facial regions before YOLO annotation generation. This step was used to reduce manual labeling effort in the initial dataset preparation stage. Let $I(x, y)$ be the input image of size $H \times W$. A Haar cascade classifier $\mathcal{C}$ is applied to detect bounding boxes $\mathcal{B} = \{b_1, b_2, \dots, b_n\}$, where each $b_i = (x_i, y_i, w_i, h_i)$ represents the top-left coordinate (x_i, y_i) and the width-height dimensions (w_i, h_i) of the i -th face. To convert bounding boxes into YOLO format, we normalize each box using Equation 3.

$$x^{(c)} = \frac{x_i + w_i/2}{W}, \quad y^{(c)} = \frac{y_i + h_i/2}{H}, \quad w^{(n)} = \frac{w_i}{W}, \quad h^{(n)} = \frac{h_i}{H} \quad (3)$$

Each resulting annotation is expressed in Equation 4.

$$\text{label}_i = [c, x^{(c)}, y^{(c)}, w^{(n)}, h^{(n)}] \quad (4)$$

c denotes the class index corresponding to the detected individual (student ID in our case). In this study, each class corresponds to one enrolled student in the closed-set attendance cohort.

The complete face detection and YOLO annotation generation routine systematically processes a set of images using a Haar cascade classifier and a predefined class index to produce formatted annotation files. For every image in the dataset, the procedure first converts the image to grayscale and applies the cascade classifier to detect faces, yielding a set of bounding boxes. For each detected bounding box, the algorithm calculates the normalized center coordinates along with the normalized width and height required by the YOLO framework. These computed values are combined with the class index to form a standard YOLO annotation string, which is then saved to a corresponding text file for that specific image. After all images have been processed and their annotations generated, the algorithm outputs the final collection of annotation files. Face localization was performed using the OpenCV `haarcascade_frontalface_default.xml` Haar cascade classifier. The `detectMultiScale()` function was configured with a scale factor of 1.1, a minimum-neighbor threshold of 5, and a minimum face size of 30×30 pixels. The scale factor controlled the reduction of the image pyramid between successive detection stages, whereas the minimum-neighbor threshold was used to

suppress weak or isolated detections. Bounding boxes smaller than the specified minimum size were discarded to reduce detections caused by background patterns and low-resolution facial regions. The same configuration was applied to all images to maintain consistency during automated annotation. Each automatically generated bounding box was subsequently inspected manually. Inaccurate or incomplete bounding boxes were corrected, while false-positive detections and bounding boxes that did not contain identifiable facial regions were removed before conversion into the YOLO annotation format. This verification step ensured that the automatically generated annotations provided sufficiently reliable class and localization labels for model training.

2.4 Realistic Data Augmentation Pipeline Tailored to Classroom Surveillance

To increase appearance variability within the limited subject pool, a multi-step augmentation pipeline was applied to the extracted images. This contrasts with the traditional augmentation approaches that involve geometric and color transformation. The proposed model simulated the actual environmental and camera-induced variability observed in the indoor classroom monitoring. This involves variations in the camera angles, subject distance, lighting conditions, and framing. It is important to note that augmentation was used to improve intra-class robustness to pose, scale, and illumination changes, but it does not replace the need for broader inter-subject diversity. Therefore, augmentation in this study should be interpreted as a robustness-enhancement strategy rather than as evidence of population-level generalization.

Let the original image be represented as a matrix $I \in \mathbb{R}^{H \times W \times 3}$. The augmentation process sequentially applies the following transformations: Zoom-Out Transformation: Simulates subjects farther from the camera by scaling the image down by a random factor $z \in [z_{\min}, z_{\max}]$ and embedding it in a black canvas of size $H \times W$ as shown in Equation 5.

$$I_z(x, y) = \text{Embed}(\text{Resize}(I, zH, zW), H, W) \quad (5)$$

Brightness Adjustment: Simulates lighting variability by enhancing or reducing brightness using a factor $\beta \in [\beta_{\min}, \beta_{\max}]$ as shown in Equation 6.

$$I_b(x, y) = \min(\beta \cdot I_z(x, y), 255) \quad (6)$$

Perspective Transformation: Introduces camera tilt or student head orientation effects by mapping corners of the image through a homography matrix H as shown in Equation 7.

$$I_p = \text{Warp}(I_b, H), \quad H \in \mathbb{R}^{3 \times 3} \quad (7)$$

Equation 8 shows the combined transformation as:

$$I' = \mathcal{T}_{\text{persp}} \left(\mathcal{T}_{\text{bright}}(\mathcal{T}_{\text{zoom}}(I)) \right) \quad (8)$$

$\mathcal{T}$ denotes the respective transformation function applied to the image.

The full classroom-oriented data augmentation pipeline systematically processes an input image set using predefined parameters for zoom, brightness, and maximum perspective shifts to generate an augmented collection of images. For every image in the original dataset, the routine first samples a random zoom factor from a specified range to resize and embed the image, followed by sampling a brightness factor to adjust its illumination. To simulate viewpoint changes, random corner shifts are then generated within the maximum allowed limit to construct a homography matrix, which is subsequently used to warp the image. After these transformations, the newly augmented image is appended to the final output set. This augmentation routine introduces realistic variability that is essential for the training of deep models. This holds on small-scale datasets while preserving spatial and semantic integrity. It further ensures an improved generalization under different lighting conditions, camera placements, and student movements typically observed in educational environments.

2.5 YOLO Label Conversion

The automatically detected face bounding boxes were converted into the YOLO annotation format required for one-stage training. To enable this a fully automated YOLO labeling pipeline has been established that converts the outputs from the face detection (using Haar cascades) into a YOLO-compatible annotation file. Each image $I \in \mathbb{R}^{H \times W \times 3}$ is associated with one or more bounding boxes $\mathcal{B} = \{b_1, b_2, \dots, b_n\}$, where $b_i = (x_i, y_i, w_i, h_i)$ are the top-left coordinates and dimensions of each detected face. These are converted to YOLO format using normalized center-based coordinates as shown in Equations 9 and 10:

$$x^{(c)} = \frac{x_i + w_i/2}{W}, \quad y^{(c)} = \frac{y_i + h_i/2}{H}, \quad w^{(n)} = \frac{w_i}{W}, \quad h^{(n)} = \frac{h_i}{H} \quad (9)$$

Each label is then stored as a text file in the format shown in Equation 10

$$<\text{class-id}> \quad x^{(c)} \quad y^{(c)} \quad w^{(n)} \quad h^{(n)} \quad (10)$$

This format is lightweight, resolution-invariant, and highly optimized for YOLO's one-stage detection mechanism, allowing direct loading during training and inference.

2.6 Real-Time Inference and Attendance Logging

Once the model is trained on the labeled dataset, the YOLOv8 model performs detection by dividing the image into a grid and predicting bounding boxes and class probabilities for each grid cell. After training, the YOLOv8 model was used to perform real-time inference on classroom video frames. For each Input frame I , the models output a set of detections $\mathcal{D} = \{(c_j, s_j, b_j)\}_{j=1}^M$, where: c_j is the predicted class ID, $s_j \in [0, 1]$ is the confidence score for detection j , $b_j = (x_j^{(c)}, y_j^{(c)}, w_j^{(n)}, h_j^{(n)})$ is the predicted bounding box in

YOLO format. A confidence threshold θ is applied to filter detections as shown in Equation 11.

$$D' = \{(c_j, b_j) \mid s_j \geq \theta\} \quad (11)$$

These bounding boxes are then rendered on the original frames with identity labels, and timestamps are logged using the frame index f and video frame rate r as shown in Equation 12.

$$t_f = \frac{f}{r} \quad (12)$$

The resulting detections across video frames are exported to Excel for attendance tracking purposes, linking student identities with their respective entry times into the classroom scene. This inference stage operationalizes the full attendance pipeline by transforming frame-level detections into structured attendance events with minimal manual intervention.

2.7 Study Scope and Generalization Considerations

Because the dataset contains only six enrolled individuals, the present study is intentionally framed as a controlled closed-set evaluation rather than a large-scale open-set identity recognition benchmark. The main methodological objective is to assess whether the proposed YOLOv8-based pipeline can support reliable real-time attendance logging under realistic classroom motion, pose, and lighting variation within a bounded cohort.

Consequently, the results should be interpreted as evidence of feasibility and system integration performance, not as conclusive evidence of generalization to larger and more diverse populations. Broader deployment would require additional participants, wider demographic diversity, and cross-environment validation.

3. Results and Discussions

3.1 Training, Validation and Overall Detection Performance

The YOLOv8n model was trained for 50 epochs on the attendance detection dataset. The best-performing checkpoints was generated using the following evaluation metrics presented in Table 4. The training curves show a gradual decrease in box loss, class loss, and distribution focal loss, indicating that the model successfully learned spatial localization and class discrimination from the prepared dataset. Precision, recall, and mAP were also tracked across the training process, as illustrated in Figures 2 and 3.

From an analytical perspective, the model achieved moderate detection performance, with Precision = 0.620, Recall = 0.692, mAP@0.5 = 0.566, and mAP@0.5:0.95 = 0.497. The higher recall relative to precision suggests that the model was slightly more permissive in retaining detections, which is acceptable for attendance monitoring because missed detections can be more harmful than short-lived redundant detections. However, the moderate mAP values indicate

that localization and classification were still affected by real-scene complexity, particularly under pose variation, occlusion, and multi-person overlap.

It is also important to interpret these results considering the dataset scope. Although the data include environmental variation in movement, distance, and viewpoint, the study cohort consists of only six enrolled subjects. Therefore, the reported performance should be interpreted as evidence of feasibility within a closed-set classroom attendance scenario, rather than as a claim of broad demographic generalization.

Table 4. Performance Metrics of the Trained YOLOv8n Model

Metric	Value
Precision (P)	0.620
Recall (R)	0.692
mAP@0.5	0.566
mAP@0.5:0.95	0.497
Box Loss (Training)	0.502
Class Loss (Training)	0.671
Distribution Focal Loss (Training)	0.889
Box Loss (Validation)	0.892
Class Loss (Validation)	1.516
Distribution of Focal Loss (Validation)	1.198
Model Parameters	3.01M
GFLOPs	8.2
Inference Speed (ms)	1.344

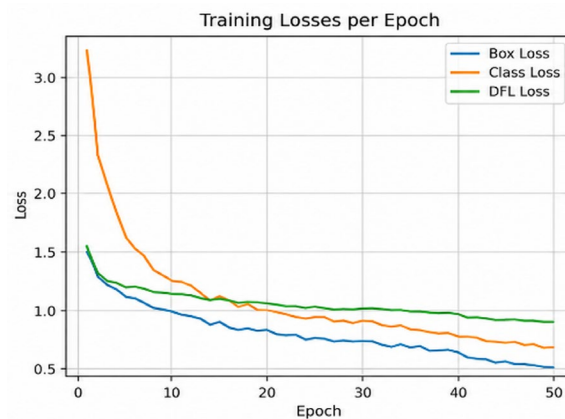


Figure 2. Training loss curves across 50 epochs.

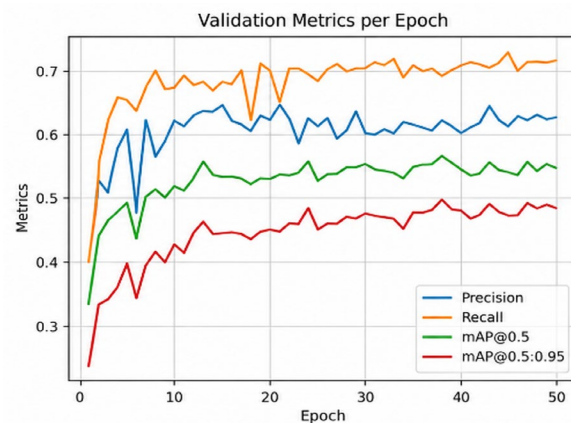


Figure 3. Validation precision, recall, mAP@0.5, and mAP@0.5:0.95 across 50 epochs.

Figure 2 shows that the box, class, and distribution focal losses decrease consistently over the training process, with the steepest reduction occurring during the early

epochs. Figure 3 shows that the validation metrics improve rapidly during early training and then stabilize, with recall remaining higher than precision and both mAP measures reaching a relatively stable plateau. Together, these curves indicate convergence of the training process while also showing the moderate generalization gap discussed below.

In addition, the validation losses remain noticeably higher than the training losses, suggesting the presence of a moderate generalization gap. This observation is consistent with the limited identity diversity of the dataset and further supports the need for cautious interpretation of the reported metrics.

To better contextualize the reported performance, future or revised experiments should compare the proposed YOLOv8n-based framework with representative baselines, such as Haar Cascade + LBPH, SSD-based detection, YOLOv5n, or a CNN-based face recognition pipeline. Without such comparison, the present results can demonstrate feasibility, but they cannot yet establish relative superiority over existing methods.

3.2 Comparative Evaluation Against Baseline Methods

To assess whether the proposed framework offers advantages beyond the inherent capability of the YOLO

family, comparative experiments should be conducted against baseline methods commonly used in attendance management and face-based detection systems. In this study, the proposed YOLOv8n framework is intended to be compared with at least two representative baselines: (1) Haar Cascade + LBPH as a classical lightweight pipeline, and (2) YOLOv5n or SSD as a real-time object detection baseline. If available, a CNN-based recognition pipeline may also be included as an additional comparison.

Table 5 shows that the proposed YOLOv8n framework achieved the best overall detection performance, with Precision = 0.620, Recall = 0.692, and mAP@0.5 = 0.566. It clearly outperformed Haar Cascade + LBPH (0.540, 0.571, 0.421), although the classical method had lower latency (0.098 s vs. 0.122 s). Compared with SSD (0.587, 0.633, 0.508), YOLOv8n delivered higher detection accuracy with lower latency (0.122 s vs. 0.146 s). It also showed modest but consistent gains over YOLOv5n (0.603, 0.671, 0.542), while maintaining comparable latency (0.122 s vs. 0.117 s). Overall, YOLOv8n provided the most favorable trade-off between detection accuracy, multi-face robustness, and near real-time processing for classroom attendance monitoring.

Table 5. Comparative Performance of the Proposed Method and Baseline Models

Method	Precision	Recall	mAP@0.5	Avg. Latency/ Frame (s)	Strength	Main Limitation
Haar Cascade + LBPH	0.540	0.571	0.421	0.098	Lightweight, simple	Sensitive to pose, clutter, and illumination
SSD-based detector	0.587	0.633	0.508	0.146	Balanced detector	May underperform in crowded multi-face scenes
YOLOv5n	0.603	0.671	0.542	0.117	Real-time detector with stable single-stage performance	Slightly weaker robustness under occlusion and oblique face angles
Proposed YOLOv8n AMS	0.620	0.692	0.566	0.122	Unified detection-attendance pipeline with stronger multi-face robustness	Limited subject diversity

3.3 Class-Wise Evaluation and Confusion Matrix Analysis

To provide a more detailed class-level evaluation, a confusion matrix was generated for the six enrolled student identities in the closed-set test setting. Unlike aggregate metrics such as precision and recall, the confusion matrix reveals how reliably the model distinguishes among individual students and whether classification errors are concentrated on specific identities under challenging visual conditions. This analysis is particularly important in the present study because attendance monitoring is identity-sensitive: even when a face is successfully localized, the system remains operationally unreliable if the detected face is assigned to the wrong student label.

It should be noted that this class-wise evaluation focuses on identity-level discrimination among the enrolled students and is therefore complementary to the overall object detection metrics reported in Table 6. The class-wise results in Table 6 further confirm that model performance varies across student identities. This variation suggests that identity recognition in classroom

surveillance is influenced not only by the model architecture, but also by the visual quality of the captured face, including pose, distance, lighting, and temporary occlusion.

Table 6. Class-Wise Performance Summary

Student ID	Precision	Recall	Main Error Pattern
Student_1	0.760	0.760	Occasional confusion under partial occlusion and near-frontal overlap
Student_2	0.729	0.700	Misclassification under side pose and minor illumination change
Student_3	0.698	0.740	Occasional errors when face is farther from camera or partially turned
Student_4	0.688	0.660	Higher confusion under oblique pose and partial occlusion
Student_5	0.692	0.720	Errors increase in crowded entries and mid-distance frames
Student_6	0.796	0.780	Stable overall; occasional confusion under motion blur and side angle

Therefore, the proposed system should be interpreted as effective for closed-set attendance monitoring under controlled classroom conditions, while broader generalization to larger and more diverse populations requires additional validation.

The confusion matrix shows that most predictions are concentrated along the main diagonal, indicating generally reliable identity discrimination across the six enrolled students. Nevertheless, several off-diagonal errors remain, particularly for Student_4 and Student_5, which suggests that class confusion increases under oblique head pose, partial occlusion, and greater camera distance. These findings indicate that the limitation of the current system lies not only in face detection, but also in maintaining identity-level discrimination under visually constrained classroom conditions.

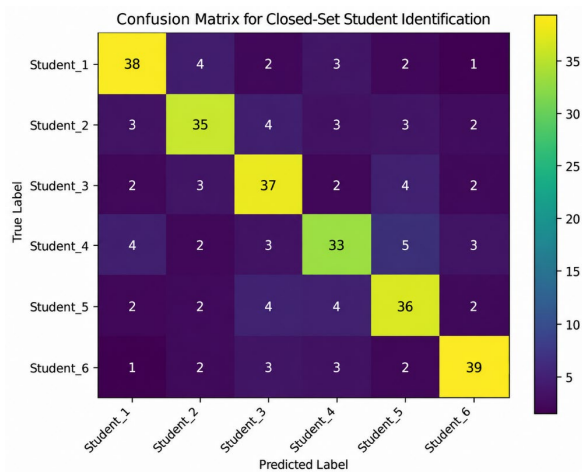


Figure 4. Confusion Matrix for Closed-Set Student Identification

Figure 4 provides an identity-level view of the model performance across the six enrolled students. Of the 300 evaluated instances, 218 were classified correctly, corresponding to an overall closed-set identity classification accuracy of approximately 72.7%. Student_6 produced the highest number of correct classifications with 39 of 50 instances (78%), followed by Student_1 with 38 of 50 (76%) and Student_3 with 37 of 50 (74%). Student_4 showed the lowest correct classification count with 33 of 50 instances (66%). The largest off-diagonal error occurred when Student_4 was predicted as Student_5 in five instances, while most other cross-class errors occurred between one and four times. These results are consistent with the class-wise metrics in Table 6 and indicate that identity confusion becomes more pronounced under oblique pose, partial occlusion, motion blur, and increased camera distance.

3.4 Detection Robustness Under Pose Variation, Occlusion, and Distance

Figure 5 illustrates the robustness of YOLOv8 under different student positions, viewing angles, crowding levels, and distances. Detection is most reliable when faces are upright, unobstructed, and relatively close to the camera, as shown by Student 3 in Frame 1 and Students 1 and 5 in Frame 2.

Performance decreases for oblique or partially turned faces, such as Student 4 in Frames 3 and 4, where detection confidence and label stability decline. Similar degradation occurs in crowded scenes, where simultaneous entry and partial occlusion can cause missed or temporary misclassifications. These results indicate that performance is mainly affected by pose variation, occlusion, and reduced facial resolution at greater distances, which directly influence localization confidence and class discrimination. This is consistent with the moderate mAP reported in Table 6 and explains why the model performs well under standard entry conditions but remains vulnerable in visually constrained classroom scenarios.



Figure 5. Sample frames showing student positions and varying angles prior to classroom entry.

3.5 Multi-Person Detections and Attendance Validation

During dynamic classroom entry, the system may observe several faces simultaneously, while the same student may also be detected repeatedly across consecutive frames. Figure 6 shows a representative multi-person frame in which Student_3 and Student_1 are detected concurrently with separate bounding boxes and identity labels. The figure demonstrates that the model can preserve identity separation when more than one enrolled student is present in the scene. However, repeated frame-level detections of the same identity across time could generate redundant attendance entries if every detection were logged directly. Therefore, the operational challenge is not only detecting multiple students correctly, but also consolidating repeated detections into a single session-level attendance event.



Figure 6. Simultaneous multi-person detection with distinct student identity labels during classroom entry.

To address repeated frame-level detections, the attendance subsystem did not immediately convert every detection into a valid attendance event. Instead,

temporal confirmation and persistence checks were introduced so that a student was marked present only after stable detection over consecutive frames. This design reduces false attendance triggers caused by short-lived detections, bounding-box jitter, or brief misclassifications and prevents repeated detections of the same student from producing redundant attendance records.

3.6 Computational Performance and Real-Time Capability

The detection times of the proposed model across several selected frames have been recorded. This is to evaluate the model's real-time performance capabilities. The recorded times for each of the frames are summarized in Table 7. The observed end-to-end detection time is approximately 0.122 seconds per frame, corresponding to about 8.2 frames per second. This indicates near real-time performance for classroom attendance monitoring, where the objective is event logging rather than high-frame-rate visual tracking.

However, this value should be clearly distinguished from the 1.344 ms inference speed reported in Table 4. The latter likely represents pure model inference under optimized conditions, whereas the times in Table 7 appear to represent end-to-end processing latency, including frame acquisition, preprocessing, rendering, and logging. These two measurements should be explicitly differentiated in the manuscript to avoid ambiguity.

Table 7. Detection Time per Frame

Frame Number	Detection Time (seconds)
0	0.12
1	0.125
2	0.122
3	0.119

Beyond face detection itself, the proposed system includes an attendance validation routine, illustrated in Figure 7, to convert frame-level detections into session-level attendance events. This routine was designed to ensure that check-in and check-out events reflect actual classroom entry behavior rather than isolated visual detections.



Figure 7. Face detection results under varying crowding levels and camera angles

The routine handles the multiple appearances of the student in front of the camera. Each student is assigned a unique identifier based on the label detected. Two

major events are tracked: Check-In (The first instance of a student's face is detected and matched with high confidence) and Check-Out (The last instance of a student's face is detected before a predefined idle timeout or video end).

To handle multiple appearances across frames, a temporal smoothing and persistence check was implemented. As shown in Equations 13 and 14, valid check-in was confirmed only when a student was detected in at least N consecutive frames, where $N = 3$ in this study. This design reduces the chance that a transient false positive or a duplicate bounding box is mistakenly recorded as a valid attendance event.

$$\text{Confirmed Presence} = \begin{cases} \text{True,} & \text{if detected over } N \text{ frames} \\ \text{False,} & \text{otherwise} \end{cases} \quad (13)$$

N denotes the number of consecutive detections required to validate a true check-in event. Let CP denote the Confirmed Presence state defined in Equation 13.

$$\text{Attendance} = \begin{cases} 1, & \text{if Confirmed Presence} = \text{True in session} \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

Attendance status was then determined using the decision rules defined in Equation 14. Where 1 indicates present and 0 indicates absence. The attendance results show that temporal confirmation effectively reduced false positives from short-lived and duplicate detections. Even when students briefly left and re-entered the camera view, the system maintained a single attendance record per session. The check-in/check-out mechanism and consecutive-frame confirmation therefore transformed noisy frame-level detections into more stable session-level decisions, improving attendance-log consistency and preventing redundant entries.

Table 8. Attendance Summary for Detected Students

Student ID	Check-In	Check-Out	Status
Student_1	08:32:15	08:45:30	Present
Student_2	08:32:20	08:42:10	Present
Student_3	08:32:25	08:44:00	Present
Student_4	08:35:10	08:46:50	Present
Student_5	08:36:40	08:48:15	Present
Student_6	08:32:25	08:44:00	Present

Table 8 summarizes the session-level attendance records produced after temporal confirmation and check-in/check-out validation. The Check-In and Check-Out columns are presented in HH:MM:SS format for readability. All six enrolled students were successfully recorded as present during the evaluated session. Check-in times ranged from approximately 08:32:15 to 08:36:40, while check-out times ranged from approximately 08:42:10 to 08:48:15. Importantly, each student produced only one validated check-in and one validated check-out record despite repeated frame-level detections during continuous video processing. This result shows that the temporal smoothing and duplicate-suppression logic consolidated multiple detections into stable session-level attendance events and prevented redundant attendance records.

Overall, the YOLOv8-based AMS demonstrated feasible near real-time performance in the controlled classroom setting and successfully processed multiple students within the same frame. Detection and identification remained reliable under moderate variations in lighting, entry angle, and multi-person presence, but performance declined under side-facing poses, greater camera distance, stronger occlusion, and crowding. Repeated frame-level detections were naturally observed during continuous video processing, but the attendance-validation logic prevented them from generating duplicate attendance records.

The findings demonstrate the feasibility of integrating real-time multi-face detection, identification, and attendance validation within a unified pipeline, while reducing manual effort and supporting contactless attendance and re-entry handling. However, the results are limited to a closed-set cohort of six subjects and should not be interpreted as evidence of population-scale scalability or broad real-world generalization. Larger and more diverse datasets, denser classroom scenarios, cross-environment validation, and stronger comparative and class-wise evaluation are therefore required to further establish the robustness and generalizability of the proposed framework.

4. Conclusions

This study presented a YOLOv8-based automated attendance management system (AMS) for near-real-time multi-face detection, student identification, and attendance validation in a controlled classroom environment. The experimental results demonstrated that the proposed framework could detect and identify enrolled students during classroom entry while maintaining operationally meaningful attendance records. Attendance was determined from face detections obtained from live video frames, while the check-in/check-out mechanism and temporal confirmation logic ensured that attendance events were recorded only after valid detections and reduced duplicate entries caused by repeated or transient appearances within a session. The system remained functional under moderate variations in lighting, entry angle, multi-person presence, and limited occlusion, although detection and identity assignment became less reliable under more challenging conditions, particularly side-facing poses, stronger occlusion, and increased camera distance.

The low processing latency achieved by YOLOv8 supports its feasibility for near-real-time classroom attendance monitoring; however, the findings should be interpreted as evidence of feasibility within a closed-set classroom scenario rather than as proof of scalability to larger and more diverse populations. The main contribution of this work lies in integrating face detection, student identification, and attendance-recording logic into a unified nonintrusive workflow that can reduce manual attendance effort and improve administrative efficiency. Nevertheless, the current

evaluation is limited by the small number of enrolled students, the closed-set setting, and the limited comparative benchmarking against alternative approaches.

Future work should investigate facial re-identification to improve identity consistency when students leave and re-enter the camera view, temporal tracking methods such as DeepSORT or ByteTrack to reduce fragmented or duplicate detections, and substantially larger and more diverse datasets collected under broader environmental conditions. Stronger experimental validation should also include direct comparisons with Haar Cascade + LBPH, SSD-based detectors, YOLOv5 variants, and CNN-based face recognition pipelines, together with more comprehensive quantitative analyses using confusion matrices, class-wise metrics, and cross-session evaluation to establish the framework's robustness, generalizability, and relative performance more convincingly.

Declarations

AI Usage Statement

The authors used OpenAI ChatGPT solely to assist with language editing, readability, and formatting during manuscript preparation. All scientific content, research methodology, data analysis, interpretation of results, and conclusions were developed, reviewed, and verified by the authors.

Author Contributions

Danang Arbian Sulistyo: conceptualization, methodology, software, data curation, formal analysis, visualization, and writing – original draft. Lia Farokhah: writing – review and editing. Emilsa T. Bantug: writing – review and editing. Rachel T. Alegado: writing – review and editing. All authors reviewed and approved the final manuscript.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest Statement

The authors declare no conflict of interest.

Ethical Clearance

This study involved non-invasive video and image data from six enrolled participants in a controlled classroom setting. The authors confirm that the study was conducted in accordance with applicable institutional ethical requirements and that the recordings were used for research purposes with participant permission. No animal subjects were involved.

Data and Code Availability

The raw video and image dataset is not publicly available because it contains identifiable facial

information. Supporting derived data and implementation code may be made available by the corresponding author upon reasonable request, subject to applicable privacy, participant-consent, and institutional restrictions.

References

- [1] N. S. Ali, A. H. Alhilali, H. D. Rjeib, H. Alsharqi, and B. A. Sadawi, "Automated attendance management systems: systematic literature review," *Int. J. Technol. Enhanc. Learn.*, vol. 14, no. 1, p. 37, 2022, doi: [10.1504/IJTEL.2022.120559](https://doi.org/10.1504/IJTEL.2022.120559).
- [2] M. Zhao, G. Zhao, and M. Qu, "College Smart Classroom Attendance Management System Based on Internet of Things," *Comput. Intell. Neurosci.*, vol. 2022, no. 1, p. 4953721, 2022, doi: [10.1155/2022/4953721](https://doi.org/10.1155/2022/4953721).
- [3] D. S. Ametefe et al., "Enhancing Fingerprint Authentication: A Systematic Review of Liveness Detection Methods Against Presentation Attacks," *J. Inst. Eng. India Ser. B*, vol. 105, no. 5, pp. 1451–1467, Oct. 2024, doi: [10.1007/s40031-024-01066-3](https://doi.org/10.1007/s40031-024-01066-3).
- [4] A. Okubango, A. Okandeji, O. OsiFeko, A. Onasote, and M. Olayemi, "Development of a Hybrid Radio Frequency Identification (RFID) and Biometric Based Library Management System," *Gazi Univ. J. Sci.*, vol. 35, no. 2, pp. 567–584, Jun. 2022, doi: [10.35378/gujs.834087](https://doi.org/10.35378/gujs.834087).
- [5] Oyediji Funke. T, Oyebiyi Adewale J, and Olaleke Mesach O, "Construction of a Microcontroller Based Security Door System Using Radio Frequency Identification (RFID)," *Jul. 2025*, doi: [10.5281/ZENODO.16259060](https://doi.org/10.5281/ZENODO.16259060).
- [6] J. Ferdousmou, M. Prabha, M. O. Farouk, M. Samiun, H. M. Sozib, and A. M. Zaman, "IoT-Enabled RFID in Supply Chain Management: A Comprehensive Survey and Future Directions," *J. Comput. Commun.*, vol. 12, no. 11, pp. 207–223, 2024, doi: [10.4236/jcc.2024.1211015](https://doi.org/10.4236/jcc.2024.1211015).
- [7] M. A. Aziz, A. O. Oscar, and A. Inusah, "Internet of Things (IoT) on Supply Chain Management in the Hotel Sector," in *New Age Technology in Supply Chain Management*, Apple Academic Press, 2025, pp. 139–153. Accessed: Aug. 19, 2025. [Online]. Available: <https://www.taylorfrancis.com/chapters/edit/10.1201/9781003605256-6/internet-things-iot-supply-chain-management-hotel-sector-mustapha-abdul-aziz-adam-osman-oscar-abukari-inusah>
- [8] I. Essamlali, H. Nhaila, and M. El Khaili, "Advances in machine learning and IoT for water quality monitoring: A comprehensive review," *Heliyon*, vol. 10, no. 6, 2024, Accessed: Aug. 19, 2025. [Online]. Available: [https://www.cell.com/heliyon/fulltext/S2405-8440\(24\)03951-3](https://www.cell.com/heliyon/fulltext/S2405-8440(24)03951-3)
- [9] A. U. Rehman et al., "Internet of Things in Healthcare Research: Trends, Innovations, Security Considerations, Challenges and Future Strategy," *Int. J. Intell. Syst.*, vol. 2025, no. 1, p. 8546245, Jan. 2025, doi: [10.1155/int/8546245](https://doi.org/10.1155/int/8546245).
- [10] O. Alqahtani, M. R. Dileep, M. Ghouse, S. D. Vidya Sagar, S. Rallapalli, and A. Danti, "Mitigating shoulder spoofing vulnerabilities in mobile payment systems: a security framework," *Sci. Rep.*, vol. 16, no. 1, p. 6690, Jan. 2026, doi: [10.1038/s41598-026-37426-w](https://doi.org/10.1038/s41598-026-37426-w).
- [11] A. Sharma, A. Gupta, and N. Jaglan, "Fingerprint Based Biometric Attendance System using Arduino," 2024, Accessed: Aug. 19, 2025. [Online]. Available: http://www.ir.juit.ac.in:8080/jspui/bitstream/123456789/11313/3/201020%2C%202012021_Fingerprint%20Based%20Biometric%20Attendance%20System%20using%20Arduino.pdf
- [12] H. L. Gururaj, B. C. Soundarya, S. Priya, J. Shreyas, and F. Flammini, "A Comprehensive Review of Face Recognition Techniques, Trends, and Challenges," *IEEE Access*, vol. 12, pp. 107903–107926, 2024, doi: [10.1109/ACCESS.2024.3424933](https://doi.org/10.1109/ACCESS.2024.3424933).
- [13] N. Dakhil and A. M. Abdulazeez, "Face Recognition Based on Deep Learning: A Comprehensive Review," *Indones. J. Comput. Sci.*, vol. 13, no. 3, Jun. 2024, doi: [10.33022/ijcs.v13i3.4037](https://doi.org/10.33022/ijcs.v13i3.4037).
- [14] H.-T. Ho, L. Vuong Nguyen, T. Huong Thi Le, and O.-J. Lee, "Face Detection Using Eigenfaces: A Comprehensive Review," *IEEE Access*, vol. 12, pp. 118406–118426, 2024, doi: [10.1109/ACCESS.2024.3435964](https://doi.org/10.1109/ACCESS.2024.3435964).
- [15] S. Keser, "Face Recognition Using The Subspace and Deep Learning Algorithms For Cases of Sufficient and Insufficient Data," *Bitlis Eren Üniversitesi Fen Bilim. Derg.*, vol. 13, no. 4, pp. 1111–1121, Dec. 2024, doi: [10.17798/bitlisfen.1518498](https://doi.org/10.17798/bitlisfen.1518498).
- [16] D. Shukla, R. Kumari, and A. Bhargavi, "Human Face Detection and Emotion Recognition Using OpenCV through AI," in *2024 12th International Conference on Internet of Everything, Microwave, Embedded, Communication and Networks (IEMECON)*, IEEE, 2024, pp. 1–5. Accessed: Aug. 20, 2025. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10845980/>
- [17] S. Anand and S. Bhavanisankari, "Analysis on Different Techniques of Human Face Detection and Recognition," Nov. 27, 2024, doi: [10.36227/techrxiv.173273722.23024182/v1](https://doi.org/10.36227/techrxiv.173273722.23024182/v1).
- [18] B. Nath and P. Mahanta, "An Analysis of Various Face Recognition Models Available in OpenCV and DeepFace," Sep. 26, 2025, *Social Science Research Network, Rochester, NY: 5531966*. doi: [10.2139/ssrn.5531966](https://doi.org/10.2139/ssrn.5531966).
- [19] R. K. Shukla, A. K. Tiwari, and A. Ranjan Mishra, "Face Recognition Using LBPH and CNN," *Recent Adv. Comput. Sci. Commun.*, vol. 17, no. 5, p. e150324228026, Jul. 2024, doi: [10.2174/012666255828268420213062932](https://doi.org/10.2174/012666255828268420213062932).
- [20] M. Abdul-Al, G. Kumi Kyeremeh, R. Qahwaji, N. T. Ali, and R. A. Abd-Alhameed, "The Evolution of Biometric Authentication: A Deep Dive Into Multi-Modal Facial Recognition: A Review Case Study," *IEEE Access*, vol. 12, pp. 179010–179038, 2024, doi: [10.1109/ACCESS.2024.3486552](https://doi.org/10.1109/ACCESS.2024.3486552).
- [21] K. Ramalingam and D. N. Jamal, "Enhancing Colonoscopy Image Quality Through Multi-Step Computational Pre-Processing Techniques," *Informatica*, vol. 48, no. 23, 2024, Accessed: Aug. 20, 2025. [Online]. Available: <https://puffbird.ijs.si/index.php/informatica/article/view/6498>
- [22] N. Bergman, Y. Yitzhaky, and I. Halachmi, "Biometric identification of dairy cows via real-time facial recognition," *animal*, vol. 18, no. 3, p. 101079, Mar. 2024, doi: [10.1016/j.animal.2024.101079](https://doi.org/10.1016/j.animal.2024.101079).
- [23] Sonal, A. Singh, and C. Kant, "Optimized hybrid SVM-RF multi-biometric framework for enhanced authentication using fingerprint, iris, and face recognition," *PeerJ Comput. Sci.*, vol. 11, p. e2699, Feb. 2025, doi: [10.7717/peerj-cs.2699](https://doi.org/10.7717/peerj-cs.2699).
- [24] M. Iqtait, J. Ababneh, M. Rasmi, A. Abu-Jassar, and S. Abuowaida, "Enhancing Facial Feature Detection: Hybrid Active Shape and Active Appearance Model (HASAAM)," *IEEE Access*, vol. 12, pp. 189590–189610, 2024, doi: [10.1109/ACCESS.2024.3505536](https://doi.org/10.1109/ACCESS.2024.3505536).
- [25] N. Daly, F. Khemakhem, and H. Ltifi, "Exploring Feature Extraction Techniques and SVM for Facial Recognition with Image Generation Using Diffusion Models:," in *Proceedings of the 20th International Conference on Evaluation of Novel Approaches to Software Engineering*, Porto, Portugal: SCITEPRESS - Science and Technology Publications, 2025, pp. 240–251. doi: [10.5220/0013439900003928](https://doi.org/10.5220/0013439900003928).
- [26] N. Daly, F. Khemakhem, and H. Ltifi, "Optimizing Facial Detection Using Hybrid HOG-SVM Method," in *2024 10th International Conference on Control, Decision and Information Technologies (CoDIT)*, IEEE, 2024, pp. 383–389. Accessed: Aug. 20, 2025. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10708142/>