

Fingerprint and Identity-Marker-Based Server Room Access Security System Using YOLOv8 and Person Re-Identification

Bilqis Sofiana Muntadiroh¹, Moh. Abdullah Anshori², Chandrasena Setiadi³

^{1,2,3} Digital Telecommunication Network Study Program, Department of Electrical Engineering, State Polytechnic of Malang, 65141, Indonesia

¹bilqismuntadiroh@gmail.com, ²anshori_ma@polinema.ac.id, ³chandrasena.setiadi@polinema.ac.id

Abstract— Server rooms house vital information technology infrastructure, so unauthorized physical access can pose risks of data leakage and operational disruption. This study designs and implements an entry–exit access security system for server rooms that combines fingerprint authentication with three visual identification methods: human detection using YOLOv8n, multi-object tracking with ByteTrack, cross-frame identity continuity maintained through Person Re-Identification (OSNet-AIN), and face verification using ArcFace. When a user's fingerprint and face have both been verified, an ESP32 microcontroller activates a relay and solenoid lock to open the entry door, while at the exit door the user only needs to be re-recognized via Re-ID without repeating full authentication. All access logs are stored in a MySQL database and can be monitored in real time through a FastAPI-based website interface. Testing across variations in distance, light intensity, clothing changes, and simulated tailgating shows that YOLOv8n achieves an F1-Score of 82% at an optimal confidence threshold of 0.439, ArcFace achieves 100% recognition accuracy at an operational distance of 1–2 meters, while OSNet obtains a Rank-1 Accuracy of 31.1% and mAP of 29.1% as a supporting component for body-posture tracking. The system is proven capable of preventing tailgating attempts through validation of entry-log status in the database, demonstrating that a layered visual identification approach combined with IoT-based fingerprint authentication can serve as a more reliable server room security solution than conventional systems.

Keywords— *Face Recognition, Fingerprint Authentication, Internet of Things, Person Re-Identification, Server Room Security, YOLOv8.*

I. INTRODUCTION

The server room is a vital component of information technology infrastructure because it stores data and runs business applications essential to an organization's continuity. Physical security of the server room is crucial to prevent unauthorized access, data theft, and hardware damage. In Indonesia, Government Regulation No. 82 of 2012 on the Implementation of Electronic Systems and Transactions requires electronic system operators to maintain the confidentiality, integrity, and availability of electronic information in accordance with applicable provisions [1].

Server room security systems commonly used today still rely on CCTV cameras, conventional door locks, and periodic monitoring by staff. This approach works reasonably well under normal conditions but is not yet effective in handling unexpected situations such as access outside operating hours or break-in attempts, due to the absence of automatic early detection and real-time notification. Wicaksono et al. developed a Raspberry Pi-based server room monitoring system with motion and ultrasonic sensors that sends notifications via a website, showing that an integrated IoT approach provides a far faster response than a passive manual system [2]. Meanwhile, Suhardi et al. developed a door security

system with fingerprint authentication and IoT-based notification, but the system was not yet equipped with visual detection or identification mechanisms, so it could not provide visual evidence in the event of an illegal access attempt [3].

Based on these conditions, a research gap remains in the form of access authentication, visual identification, and real-time monitoring not yet being integrated into a single server room security platform. Prior studies also tend to focus on a single method, such as human detection with YOLO [4], Person Re-Identification in surveillance systems [5], or face recognition for door access control [6], [7], without integrating all three together with fingerprint biometric authentication. Table I summarizes five closely related studies together with the specific gap that each leaves open and that this study aims to close.

TABLE I
COMPARISON WITH CLOSELY RELATED STUDIES

Study	Method	Result	Remaining Gap
Efendi et al. [4]	YOLOv8m-seg detection	High-accuracy human presence detection	No identity verification

Suhardi et al. [3]	Fingerprint + IoT notice	Real-time notice on match	No visual identification
Manshur & Wahyono [5]	Re-ID, Bag-of-Visual-Features	88% cross-camera matching	No access control support
Mawarid et al. [6]	Face recognition (LBPH)	100% recognition rate	No cross-camera tracking
Muzakkir & Riyanto [7]	ESP32 + face recognition	Functional face-based access	No detection/Re-ID, 1-factor only

This study proposes a server room access security system that integrates fingerprint authentication with three layers of visual identification: human detection and tracking using YOLOv8n and ByteTrack, identity-continuity maintenance through Person Re-Identification (OSNet-AIN), and final verification using Face Recognition (ArcFace). All artificial intelligence inference processes run on local computation without depending on third-party cloud services, while detection results, authentication status, and access history are displayed through a real-time website interface. The contributions of this study are (1) a hybrid security architecture design that positions each identification method according to its respective failure characteristics, and (2) a quantitative evaluation of the accuracy of YOLOv8n, ArcFace, and OSNet under varying conditions of distance, lighting, and clothing change.

II. METHOD

A. Research Stages

This study uses a development research approach aimed at designing and building a new system based on advances in science and technology. The research stages include identifying the physical security problem of the server room, literature study, determination of the research method, planning of hardware and software requirements, system development, testing of each component (fingerprint sensor, camera and YOLOv8 model, door control, log storage, and notification), system refinement, and finally data analysis and drawing of conclusions.

B. System Block Diagram Design

The server room access security system consists of two IP CCTV camera units [16], a laptop as a local server, an ESP32 microcontroller [15], an AS608 fingerprint sensor, a relay, a solenoid lock, an LCD, a MySQL database, and a monitoring website interface, as shown in Fig. 1. Both IP cameras are connected to the network through a POE switch and router, then send video in real time to the local server via the RTSP protocol. The local server runs a FastAPI-based backend along with an AI engine that processes the video using YOLOv8n for human detection, followed by ByteTrack for object tracking, OSNet for Person Re-Identification, and ArcFace for face verification. The fingerprint sensor connected to the ESP32 serves as an additional authentication requirement before the relay activates the solenoid lock. All identification results and access logs are stored in the MySQL database and displayed on a website accessible from outside the local network via the Ngrok tunneling service.

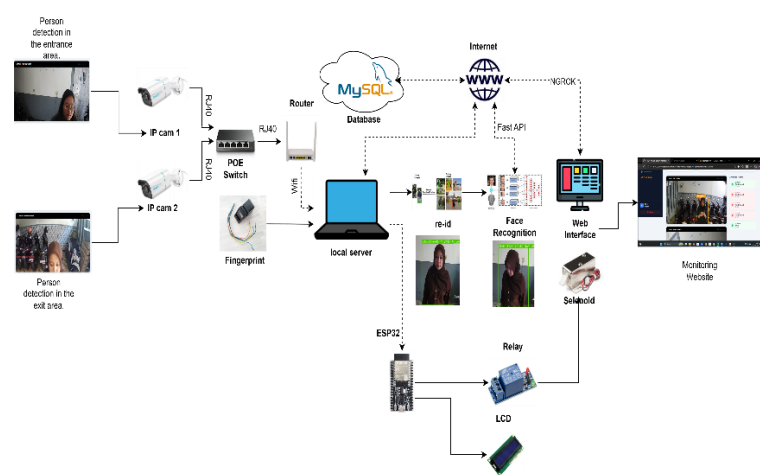


Figure 1. Block diagram of the server room access security system.

In simple terms, the five identification components form a pipeline of increasingly specific checks, each handing off to the next only after answering its own narrower question. YOLOv8n first asks “is a person present in the frame?”; ByteTrack then keeps following that same person as “one continuous object” as they move, even through brief occlusion; Person Re-Identification (OSNet-AIN) asks “does this person's overall appearance match someone the system has already seen?” to preserve identity across frames and cameras; ArcFace then asks the more precise question “does this specific face match a registered employee?”; and finally the AS608 fingerprint sensor provides an independent, non-visual physical check. Only after a user clears all of the checks required for a given door does the ESP32 open the corresponding solenoid lock. Subsections C–E below describe each of these components in turn, and Subsection F then walks through how they are chained together step by step at the entry and exit doors.

C. Object Detection and Tracking (YOLOv8n and ByteTrack)

The YOLOv8n model was chosen because it has a lightweight parameter size while still providing good detection accuracy for real-time processing [8], [9]. YOLOv8's architecture is composed of three main parts: a backbone, a neck, and a detection head, as illustrated in Fig. 2. The backbone is based on a modified CSPNet that incorporates the C2f (Concatenate-and-Copy-Forward) module to improve gradient flow and computational efficiency. The neck combines Feature Pyramid Network (FPN) and Path Aggregation Network (PANet) structures to fuse multi-scale features, allowing objects of different sizes to be detected quickly. The head adopts an anchor-free, decoupled design and integrates Distribution Focal Loss (DFL) together with CIOU/GIOU loss functions, which significantly improve detection and classification accuracy [9].

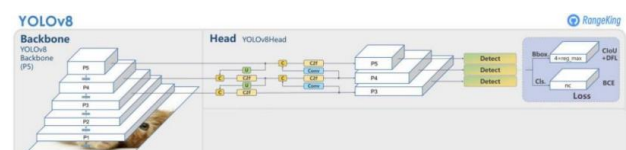


Figure 2. YOLOv8 architecture, consisting of a backbone, a neck, and a decoupled detection head [9].

The dataset used is a Personal Protective Equipment (PPE) dataset from Roboflow Universe that provides a person class, which was then normalized and divided into 70% training data, 20% validation data, and 10% test data using a stratified split to keep the class distribution balanced. Training was carried out on Google Colaboratory with T4 GPU acceleration using the Ultralytics library. The model takes 640×640 pixel video frames as input and produces bounding-box coordinates, confidence scores, and class labels as output. These bounding-box coordinates and confidence scores then become input for the ByteTrack algorithm, which associates every high- and low-confidence detection across frames to consistently maintain the tracking identity (track ID) of human objects, including during brief occlusion [10]. Unlike conventional tracking methods that only use high-confidence detections, ByteTrack recovers weak, partially occluded detections through a two-stage association process (high-confidence matching followed by low-confidence matching using IoU and the Hungarian algorithm) and predicts interim positions with a Kalman filter, which reduces missing detections and track fragmentation while keeping the tracker independent of appearance-based Re-ID features [10]. In this study, ByteTrack maintains a consistent track ID for every user detected at the entry and exit cameras, providing the temporal continuity that the Person Re-Identification stage subsequently relies on.

D. Person Re-Identification (OSNet-AIN)

Person Re-Identification is used to re-recognize a person's identity based on overall visual body features (clothing and posture) when that identity has momentarily left the camera's field of view. The algorithm used is OSNet-AIN (Omni-Scale Network), which learns features at multiple receptive scales simultaneously to increase discriminative power between identities [11], [12]. As shown in Fig. 3, OSNet extracts visual features through four parallel convolutional streams with different receptive-field sizes ($R=3$, $R=5$, $R=7$, and $R=9$), which are then combined through a channel-wise Aggregation Gate (AG) that assigns adaptive weights to each stream according to the characteristics of the input image. This omni-scale design allows the network to jointly learn local details (e.g., clothing texture) and global body structure, which is what makes it discriminative between visually similar identities [12]. OSNet is also a lightweight network: it uses depthwise-separable convolutions and has only around 2.2 million parameters, far fewer than ResNet50-based Re-ID backbones that typically exceed 24 million parameters, without a significant loss of accuracy [12].

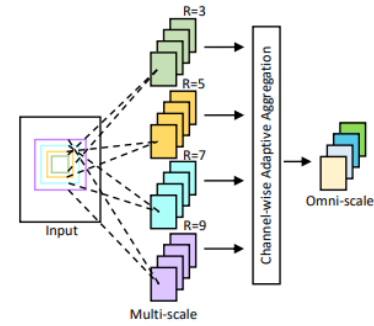


Figure 3. Omni-scale feature learning concept used by OSNet, combining four receptive-field streams through an Aggregation Gate [12].

The model was trained for 60 epochs on the Google Colaboratory platform using the PRCC dataset, which has complex variation in lighting, viewing angle, and clothing change, with 150 identities used for training and 71 different identities for final testing (query and gallery sets).

E. Face Recognition (ArcFace) and Fingerprint Authentication

Among the face-recognition approaches surveyed in the literature [14], this study uses ArcFace, which applies an additive angular margin loss function to enlarge the margin between classes in the face feature space, thereby increasing discriminative power between individuals [13]. As shown in Fig. 4, a detected face is first processed by a convolutional neural network to obtain a 512-dimensional feature embedding, which is then L2-normalized onto a hypersphere so that identity comparison depends on the angle between vectors rather than their magnitude. ArcFace then adds an angular margin to the target class before the Softmax and cross-entropy loss are computed, which pulls embeddings of the same identity closer together while pushing embeddings of different identities further apart, making the resulting representation more robust to variation in pose, lighting, and expression [13].

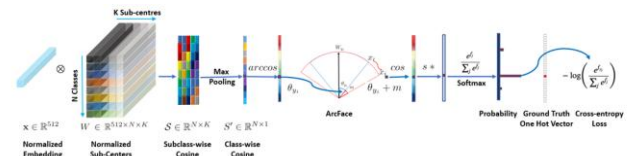


Figure 4. ArcFace training pipeline: feature extraction, L2 normalization, additive angular margin, and Softmax/cross-entropy classification [13].

The face-recognition process includes head-area detection, face cropping, image resizing, color-space conversion, pixel-value normalization, and matrix transposition from HWC to CHW format before the image is passed to the feature-vector extraction stage. The resulting feature vector is compared using cosine similarity against registered face data to determine identity match. As an additional authentication layer, similar in principle to conventional Arduino-based fingerprint door-locking systems [17], the AS608 capacitive fingerprint sensor connected to the ESP32 verifies the user's fingerprint as a final requirement before the system activates the relay and solenoid lock at the entry door.

F. Entry and Exit Access Flow

To make the interaction between the five components concrete, this subsection walks through the complete sequence of events at the entry and exit doors step by step.

1) Entry-door sequence:

Step 1 (Presence detection). YOLOv8n scans each incoming camera frame and draws a bounding box around any person it detects.

Step 2 (Continuous tracking). ByteTrack assigns that bounding box a track ID and keeps following the same ID as the person moves, even during brief occlusion, so the system does not lose track of who is who from one frame to the next.

Step 3 (Appearance check). Person Re-Identification (OSNet-AIN) compares the tracked person's overall visual appearance — clothing and body posture — against previously seen identities, preserving identity continuity across frames.

Step 4 (Face verification). ArcFace crops the person's face from the frame and compares it, using cosine similarity, against the registered face database to determine which employee, if any, it belongs to.

Step 5 (Fingerprint check). Once a face match is found, the user places a finger on the AS608 sensor; the fingerprint is matched against the enrolled fingerprint database as an independent, non-visual confirmation of identity.

Step 6 (Door actuation). Only if both Step 4 and Step 5 succeed does the ESP32 activate the relay and open the entry-door solenoid lock; the event and the verified identity are then written to the MySQL access log, and a notification is sent to the website and to the employee. Three consecutive fingerprint failures instead trigger a special alert to the administrator, whose content differs depending on whether the user's face is still recognized by the camera.

2) Exit-door sequence:

Step 1 (Presence detection and tracking). YOLOv8n and ByteTrack detect and track the person at the exit camera in exactly the same way as at the entry door.

Step 2 (Re-identification only). OSNet-AIN checks whether this person's visual appearance matches an identity currently marked as “inside the room” in the users_inside_room database state. Face recognition and fingerprint verification are not repeated at this stage, since the user was already fully verified when they entered.

Step 3 (Database validation and actuation). The system cross-checks the Re-ID match against users_inside_room. If a valid entry record exists, the relay opens the exit-door solenoid lock; if not, access is denied even when the visual identity is correctly recognized. This database cross-check is what gives the system its anti-passback / anti-tailgating capability, examined further in Section III-D.

G. Hardware and Software Implementation

The hardware was implemented on a 190×110 cm access-door mock-up equipped with two IP cameras, an AS608 fingerprint sensor, a 16×2 LCD, two solenoid lock units, and an ESP32, relay, and buck converter housed in a panel box. Fig. 5 shows the wiring design connecting these components. A DC power source is stepped down by two separate buck converters to 5

V: the first supplies the ESP32, the AS608 fingerprint sensor, and the LCD, while the second supplies the two relay modules. The fingerprint sensor communicates with the ESP32 over UART, with its TX and RX pins wired to GPIO 16 and GPIO 17 respectively; the 16×2 LCD communicates over I2C, with SDA on GPIO 21 and SCL on GPIO 22. Relay 1, wired to GPIO 26, drives the entry-door solenoid lock, while relay 2, wired to GPIO 27, drives the exit-door solenoid lock, and all modules share a common ground to keep the system electrically stable.

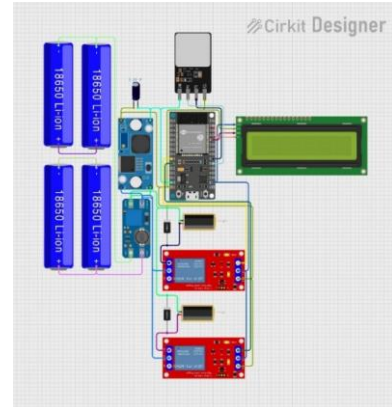


Figure 5. Wiring diagram of the ESP32, AS608 fingerprint sensor, LCD, and dual relay modules.

The system software is a website built with HTML, CSS, and JavaScript, with FastAPI as the backend and MySQL as the database, connected to the internet using Ngrok so that it can be accessed by the administrator from outside the local network. The website displays live CCTV streaming, door status, employee data management, access history, and push notifications.

H. Testing Parameters

System testing covers evaluation of YOLOv8n performance (precision, recall, F1-Score, confusion matrix), evaluation of OSNet Re-ID (train accuracy, Rank-1 Accuracy, mAP), testing of object distance from the camera, testing under varying light intensity (lux), testing of Re-ID consistency under clothing change, anti-tailgating and anti-passback testing, testing of fingerprint sensor accuracy for both registered and unregistered users, testing of door-relay actuator responsiveness, and testing of website dashboard functionality.

III. RESULTS AND DISCUSSION

A. Hardware and Website Implementation

The hardware implementation results are shown in Fig. 6. Two IP camera units are mounted above the door to capture images of users on the entry and exit sides, the fingerprint sensor is placed in the middle of the entry door for easy reach, while the LCD displays authentication status directly.



Figure 6. Hardware implementation results of the server room access door mock-up.

The developed monitoring website, shown in Fig. 7, successfully displays live camera streaming, door status, access history, and an employee management menu in real time, with all data updated automatically through communication between the FastAPI backend, the MySQL database, and the connected hardware.

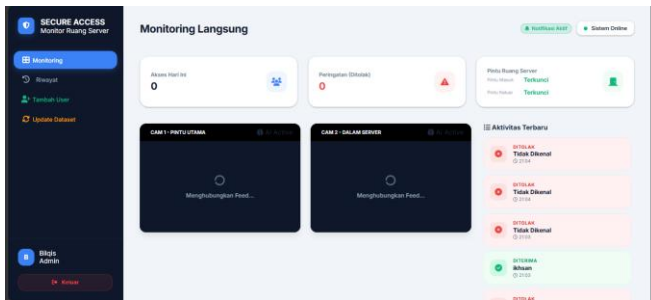


Figure 7. Real-time monitoring dashboard showing access status, camera feeds, and access history.

B. YOLOv8n Model Training Results

The YOLOv8n model was trained using the PPE dataset for 100 epochs. Fig. 8 shows box_loss, cls_loss, and dfl_loss values decreasing consistently on both the training and validation data until the final epoch, while precision, recall, mAP@0.5, and mAP@0.5:0.95 values increased and stabilized at around 0.80, 0.85, 0.84, and 0.54 respectively, indicating that the training process had converged.

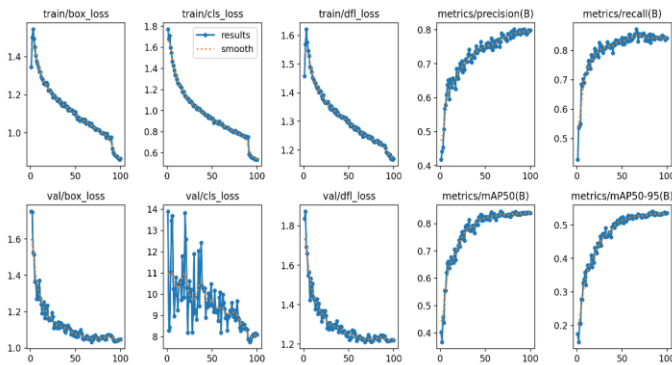


Figure 8. Training result graph of the YOLOv8n model.

Further evaluation using a confusion matrix at an operational threshold of 0.25 (Fig. 9) shows that the model correctly detected 1,799 person objects (true positives), 730 detections were false positives, and 173 person objects failed to be detected (false negatives), as summarized in Table II.

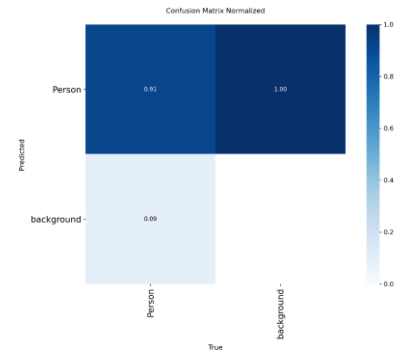


Figure 9. Normalized confusion matrix of person object detection results.

TABLE II
SUMMARY OF YOLOv8N DETECTION CONFUSION MATRIX

Class	TP	FP	FN
Person	1799	730	173

Based on these values, precision, recall, and F1-Score were calculated as follows:

$$\text{Precision} = TP / (TP + FP) = 1799 / 2529 \approx 0.711$$

$$\text{Recall} = TP / (TP + FN) = 1799 / 1972 \approx 0.912$$

$$F1 = 2 \cdot (P \cdot R) / (P + R) = 2 \cdot (0.711 \times 0.912) / (0.711 + 0.912) \approx 0.799$$

This F1-Score of 79.9% represents performance at the default confidence threshold (0.25). The F1-Confidence curve in Fig. 10 shows that the most optimal precision–recall balance is actually achieved at a confidence threshold of 0.439 with an F1-Score of 82%, so adjusting the threshold from 0.25 to 0.439 could increase the precision–recall balance by 2.1 percentage points without changing the model weights.

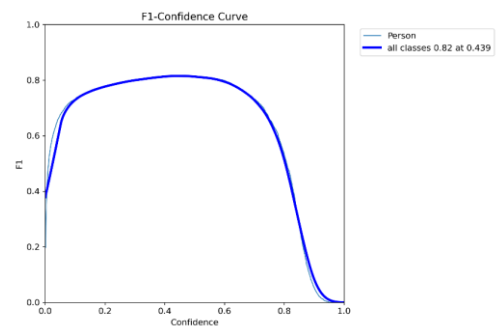


Figure 10. F1-Confidence curve of the YOLOv8n model.

C. Person Re-Identification (OSNet-AIN) Training Results and Accuracy

Person Re-Identification is one of the main identification methods in this system, as stated in the research title, so its accuracy needs to be evaluated specifically in addition to its

role within the overall architecture. Fig. 11 shows the OSNet-AIN training results. On the training data, total loss decreased steadily to below 1.0 and train accuracy increased to 99.49% across the 150 identities trained. However, in the final evaluation using a query set and gallery set containing 71 new identities not seen during training, performance dropped to a mAP of 29.1% and a Rank-1 Accuracy of 31.1%, although it increased to 54.0% at Rank-20, as summarized in Table III.



Figure 11. Training and evaluation results of the OSNet-AIN model.

TABLE III
SUMMARY OF PERSON RE-IDENTIFICATION (OSNET-AIN) MODEL ACCURACY

Evaluation Metric	Value
Train Accuracy (150 identities)	99.49%
Rank-1 Accuracy (71 test identities)	31.1%
Rank-20 Accuracy	54.0%
mAP	29.1%

The gap of more than 68 points between train accuracy and Rank-1 Accuracy in Table III indicates overfitting due to the limited number of training identities relative to the complexity of visual variation, and high intra-class variation caused by the use of identical work uniforms among employees, which further reduces the discriminative power of visual features between individuals. This finding forms the basis for the architectural decision not to position Re-ID as the sole determinant of identity, but rather as a body-tracking filter working alongside Face Recognition. Cosine-similarity testing on six identity-pair combinations shows Re-ID cosine distances ranging from 0.147–0.352 (average 0.2465), far narrower than ArcFace's separation margin of 0.4652, so the final verification decision is still handed over to the Face Recognition module, which has a much sharper discriminative power between individuals. Nevertheless, Re-ID accuracy remains an important indicator because this module is what maintains cross-frame identity-tracking continuity before final verification is performed.

D. Distance, Lighting, Clothing, and Tailgating Tests

Distance testing was carried out on four subjects over a range of 0.5–2 meters to determine the effect of distance on Re-ID accuracy for both cameras (CCTV 1 at the entry door and CCTV 2 at the exit door) compared with Face Recognition, as summarized in Table IV.

TABLE IV
RE-ID AND FACE RECOGNITION ACCURACY BY DISTANCE

Distance	Face Recognition	Re-ID CCTV 1	Re-ID CCTV 2
0.5 m	0%	25%	25%
1.0 m	100%	75%	100%
1.5 m	100%	50%	75%
2.0 m	100%	100%	75%

Face Recognition shows a binary failure pattern with respect to distance, failing completely (0%) at 0.5 meters but succeeding perfectly (100%) from 1.0 meter onward, indicating that optical focus limitation is the main cause. In contrast, Re-ID accuracy on both cameras does not show a linear trend with distance: Re-ID CCTV 1 increased from 25% (0.5 m) to 75% (1.0 m), dropped to 50% (1.5 m), then rose again to 100% (2.0 m), while Re-ID CCTV 2 peaked at 100% at 1.0 m before dropping to 75% at 1.5–2.0 m. This non-monotonic pattern indicates that Re-ID accuracy is also influenced by situational factors such as bounding-box contamination from shadows or light reflection, as well as similarity in body posture between subjects that narrows the cosine distance between feature vectors.

Light-intensity testing was carried out at three illuminance levels (309 lux, 226 lux, and 54 lux), measured with a lux meter, across the same four subjects and four distances used in the distance test. Table V recapitulates the combined results across both cameras (48 trials per illuminance level).

TABLE V
RECAPITULATION OF LIGHT-INTENSITY (LUX) TEST RESULTS

Lux	Trials	Face Rec. Cam 1	Re-ID Cam 1	Re-ID Cam 2
309 lux	48	75%	50%	75%
226 lux	48	68.75%	81.25%	62.5%
54 lux	48	75%	81.25%	81.25%

Face Recognition accuracy on Camera 1 is fairly stable at the highest and lowest illuminance levels (75% at both 309 lux and 54 lux) but dips to 68.75% at 226 lux; this non-monotonic pattern suggests the dip is more likely explained by subject-specific factors during that test session (e.g., face angle or expression) than by illuminance itself. Re-ID accuracy on Camera 1 shows a clearer trend, rising from 50% at 309 lux to 81.25% at both 226 lux and 54 lux, consistent with high illuminance causing glare that contaminates the bounding box and degrades the visual features OSNet extracts; lower illuminance reduces this glare effect. Re-ID on Camera 2 shows the opposite trend, peaking at 75% under 309 lux and falling to 62.5% and 81.25% at 226 lux and 54 lux respectively, suggesting that Camera 2's placement and viewing angle make it more sensitive to sensor noise under low light than to glare under bright light. Taken together, the two cameras exhibit different sensitivities to illuminance depending on their position and angle, indicating that camera-specific calibration should account for the lighting conditions of each installation site.

Re-ID consistency under clothing change was tested on four subjects with different scenarios — changing clothes, putting on a jacket, and removing a jacket — as summarized in Table VI.

TABLE VI
RE-ID CONSISTENCY TESTING UNDER CLOTHING CHANGE

Subject	Scenario	Result
Bilqis	Changed clothes	Detected
Cikita	Changed clothes	Detected
Fariz	Put on jacket	Detected
Ikhsan	Removed jacket	Detected

All subjects (4 of 4) were still successfully recognized despite visual changes in clothing, with tracking confidence scores dropping to a range of 64–71% but still remaining above the system threshold. This proves that OSNet is able to find a match on local-scale features that remain unchanged, such as the color of lower-body clothing, body-posture proportions, and head-area characteristics, so the system still maintains dynamic visual identity continuity across frames even when part of the clothing features changes.

As a representative example of this test, Fig. 12 shows the actual Re-ID detection output for subject Bilqis after changing clothes, captured at both the entry camera (CCTV 1) and the exit camera (CCTV 2); the system correctly retained her tracked identity label at both viewpoints despite the change in visual appearance.



Figure 12. Re-ID detection results for subject Bilqis after a clothing change: entry camera CCTV 1 (left) and exit camera CCTV 2 (right).

For completeness, Fig. 13 presents the corresponding Re-ID detection captures for the remaining three subjects tested under clothing change — Cikita (changed clothes), Fariz (put on a jacket), and Ikhsan (removed a jacket) — each shown from the entry camera (CCTV 1, left) and the exit camera (CCTV 2, right); all three were likewise correctly retained under their original tracked identity.

Cikita — Changed clothes



Figure 13. Re-ID detection results for Cikita, Fariz, and Ikhsan during the clothing-change test (CCTV 1 left, CCTV 2 right).

Anti-tailgating and anti-passback testing shows that security validation does not rely solely on visual identification accuracy, but also on validation of the user's entry-log status in the database (`users_inside_room`). The simulated scenario proceeded as follows. At entry, two subjects, Fariz and Bilqis, walked through the entry door together; the system detected and authenticated only Fariz, recording him in the `users_inside_room` database state, while Bilqis, tailgating behind him, had no authentication record and was therefore not logged as inside the room. At exit, Fariz approached the exit camera and was recognized by Re-ID with 65.5% confidence; because his identity was validated against a matching entry log, the system triggered the relay and the exit door opened. Bilqis then approached the same exit camera and was also recognized by Re-ID, with 70.9% confidence, showing that the visual match itself was successful, but the cross-check against `users_inside_room` found no corresponding entry record for her; the system consequently withheld the relay signal, the exit door remained locked, and the room continued to be monitored. This confirms that the anti-passback mechanism depends on database-state validation rather than on visual recognition.

accuracy alone, closing a structural false-acceptance gap that Re-ID by itself could not.

TABLE VII
FINGERPRINT SENSOR ACCURACY TEST RESULTS

Subject	No. of Trials	Successful	Accuracy
Bilqis	10	7	70%
Riska	10	8	80%
Chitta	10	9	90%
Average	30	24	80%

Table VII shows a fingerprint recognition success rate of 80% (24 of 30 attempts) across three registered users, with all failures classified as false rejections rather than false acceptances. In testing against five unregistered fingerprint samples, the sensor rejected all attempts (100%) without a single case of false acceptance, showing that the sensor operates conservatively in line with a security principle that prioritizes preventing unauthorized access.

E. Actuator Responsiveness and Dashboard Functionality

Testing of the door relay's responsiveness across three legitimate access scenarios yielded an average response time of 3,388.7 ms with a coefficient of variation of 76.6% on a limited sample ($n=3$), so this average value cannot yet be generalized as a quantitative representation of system performance and requires further testing based on digital timestamps with a larger sample. Nevertheless, all four test scenarios (including the access-denial scenario without an entry log) showed 100% agreement between the access-control logic and system actuation. Testing of the website dashboard's functionality also achieved 100% functional validity across all tested features, although this testing is binary in nature and lacks quantitative latency parameters.

F. Discussion

A synthesis of the test results shows that the hybrid architecture designed for this system is empirically validated through the complementary failure patterns of each module. Re-ID is vulnerable to false acceptance due to visual similarity between identically uniformed individuals, fingerprint recognition is vulnerable to false rejection but never accepts an unregistered fingerprint, while Face Recognition has sharp inter-class separation but is sensitive to image-capture distance. These three distinct failure profiles form the rational basis for the decision to position Face Recognition as the primary identity determinant, Re-ID as a supporting tracking filter, and both fingerprint verification and database validation as additional security layers. As a result, the individual weaknesses of each module do not lead to aggregate system failure, since they are compensated for by other verification layers with different failure characteristics. This validity is qualitative-functional in nature at a limited testing scale, so quantitative proof at a larger operational scale — covering a greater number of samples and a wider variety of Re-ID datasets — remains necessary as a further validation stage.

IV. CONCLUSION

This study successfully designed and implemented a server room access security system that integrates AS608 fingerprint authentication, real-time human detection using YOLOv8n and ByteTrack, Person Re-Identification using OSNet, and face verification using ArcFace, with an ESP32 as the actuator controller, FastAPI and MySQL as the backend and database, and a website interface for monitoring, all running on local computation without dependence on third-party cloud services. The system succeeded in maintaining user identity continuity despite clothing changes, with a Re-ID success rate of 100% in that test. The test results show that YOLOv8n achieved an F1-Score of 79.9% at the default threshold and could potentially reach an optimum of up to 82% at a threshold of 0.439, ArcFace achieved 100% accuracy at an operational distance of 1–2 meters, while OSNet achieved a mAP of 29.1% and a Rank-1 Accuracy of 31.1% as a supporting component for posture tracking. The limited performance of OSNet under field conditions is the main justification for the success of the hybrid biometric architecture approach, in which the final identity-verification decision is handed over to ArcFace, which is far more precise. The system also proved robust against tailgating attempts through a second security layer based on database status validation, which is able to close the gap left by fluctuations in visual accuracy through strict enforcement of access logic. Further development is recommended in the form of adding more image-capture angle variations to enrich the Re-ID dataset, calibrating cameras according to the lighting characteristics of each location, migrating AI inference to edge computing devices, and adding a feature to limit the maximum room-access capacity.

ACKNOWLEDGMENT

The authors would like to thank the Digital Telecommunication Networks Study Program, Department of Electrical Engineering, State Polytechnic of Malang, and the supervising lecturers who provided guidance and support throughout this research.

REFERENCES

- [1] K. Rahayu, N. S. Marikar Ssahib, and S. Idayanti, "Problematics of Electronic System Operator (PSE) Regulations in Indonesia," *Pancasakti Law J.*, vol. 1, no. 1, p. 69, 2023.
- [2] M. F. Wicaksono, M. D. Rahmatya, and Ilham, "IoT Implementation for Server Room Security Monitoring Using Telegram API," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 12, no. 5, pp. 1931–1937, 2022.
- [3] E. R. Suhardi, S. Z. Gunawan, and A. A. Hidayat, "Development of a Door Security System Using Fingerprint with an Internet of Things-Based Notification System," *J. Media Elektr.*, vol. 20, no. 1, 2022.
- [4] M. Y. Efendi, R. Wulanningrum, and B. Setiawan, "Design of a Human Detection System with YOLO on CCTV Video," *Inotek*, vol. 8, pp. 1149–1154, 2024.

- [5] M. Y. Manshur and Wahyono, "Person Reidentification in Intelligent Surveillance Systems Using the Bag of Visual Feature Method," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 9, no. 3, pp. 248–253, 2020.
- [6] M. Mawarid, M. Djaohar, and N. H. Yuninda, "Automatic Door Access Management System Based on Face Recognition at SMK Taruna Bhakti Depok," vol. 8, pp. 45–52, 2025.
- [7] T. Muzakkir and A. Riyanto, "Design of Room Access Using a Face Recognition System," vol. 02, no. 02, pp. 2–7, 2023.
- [8] N. R. Muntari, I. C. Nisa, A. Srikaningih, A. Y. Adyatma, and M. Yusril, "Application of the YOLOv8 Algorithm for Real-Time Face Identification Using CCTV for Student Attendance," vol. 4, no. 3, pp. 1155–1165, 2024.
- [9] J. Terven, D. M. Córdova-Esparza, and J. A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Mach. Learn. Knowl. Extr.*, vol. 5, no. 4, pp. 1680–1716, 2023.
- [10] Y. Zhang et al., "ByteTrack: Multi-Object Tracking by Associating Every Detection Box," in *Proc. ECCV*, vol. 13682, pp. 1–21, 2022.
- [11] T. He et al., "Cloth-Changing Person Re-identification from A Single Image with Gait Prediction and Regularization," *arXiv preprint*, 2016.
- [12] K. Zhou and Y. Yang, "Omni-Scale Feature Learning for Person Re-Identification," in *Proc. ICCV*, 2019.
- [13] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," in *Proc. CVPR*, pp. 1–17, 2019.
- [14] H. L. Gururaj and B. C. Soundarya, "A Comprehensive Review of Face Recognition Techniques, Trends, and Challenges," *IEEE Access*, vol. 12, pp. 107903–107926, 2024.
- [15] K. Y. Maulana, "What Is ESP32, One of the Popular Wi-Fi Modules," *AnakTeknik.co.id*, 2025.
- [16] "IPC-D120HA-LUC 2 MP Smart Hybrid Light Fixed Dome Network Camera," *Hikvision Product Datasheet*.
- [17] S. Sahril, R. Suppa, and M. Muhallim, "Fingerprint-Based Door Locking System Using Arduino," *Jurasik*, vol. 7, no. 1, p. 19, 2022.