Jurnal Teknik Informatika C.I.T Medicom, 18 (2) (2026) 161-170

Published by: Institute of Computer Science (IOCS)

Jurnal Teknik Informatika C.I.T Medicom

Journal homepage: www.medikom.iocspublisher.org

Design and implementation of a weapon storage access control system based on hand gesture recognition and face recognition on Raspberry Pi 5

Muhamad Daffa Abdur Rahman[1], Sunarta[2], Bagus Hendra Saputra[3]

[1,2,3] Department of Electrical Engineering, Faculty of Engineering and Defense Technology, Indonesia Defense University, Bogor, Indonesia

Article Info

Article history

Received : 28 Apr, 2026
Revised : May 28, 2026
Accepted : May 31, 2026

Keywords:

Hand Gesture Recognition;
LSTM;
Multimodal Biometric;
Raspberry Pi 5;
Sequential Fusion.

Abstract

This study presents a multimodal biometric access control system for weapon storage facilities, integrating hand gesture recognition and face recognition through a sequential fusion architecture on Raspberry Pi 5. The sequential design activates face verification only after correct gesture authentication, optimizing computational efficiency on edge hardware while establishing a dual-layer security barrier. The gesture module combines MediaPipe Hands landmark extraction with LSTM-based temporal classification, achieving near-perfect accuracy across four gesture classes. The face module employs dlib's ResNet-34 for 128-dimensional embedding comparison, with an empirically recalibrated Euclidean distance threshold of 0.34 to eliminate false acceptance risks identified during intrusion testing. Evaluation under controlled conditions yielded 0% False Reject Rate and 0% False Accept Rate across 60 trials, with reliable GPIO-controlled solenoid actuation. Results demonstrate that sequential fusion of behavioral and physiological biometrics on a single edge device provides a viable security solution for high-risk access control applications.

Corresponding Author:

Muhamad Daffa Abdur Rahman
Department of Electrical Engineering
Indonesia Defense University , Bogor, Indonesia
E-mail : mdaffaabdurhmn@gmail.coml

1. Introduction

Weapon storage facilities in military educational institutions require strict access control to ensure that only authorized personnel can access high-risk assets. Current security systems commonly rely on physical mechanisms such as padlocks and RFID cards, which authenticate objects rather than individuals. These approaches are vulnerable to duplication, theft, and loss[1]. Biometric technologies address this limitation by verifying inherent human characteristics that cannot be easily transferred or replicated[2].

However, unimodal biometric systems remain susceptible to specific attack vectors. However, unimodal biometric systems remain susceptible to specific attack vectors. Standalone face recognition can be spoofed using photographs or video recordings[3][4], while single-factor gesture systems may fail to distinguish authorized from unauthorized users who have observed the correct gestures. Multimodal

biometric systems that combine multiple modalities through fusion strategies offer improved security and resistance to falsification[5][6]. Recent advances in lightweight deep learning frameworks have enabled real-time biometric processing on edge devices. MediaPipe provides efficient hand landmark extraction[7][8], LSTM networks have proven effective for temporal gesture classification[9][10][11], and deep metric learning approaches using ResNet architectures have achieved near-human accuracy in face verification tasks[12][13]. The deployment of such models on single-board computers like Raspberry Pi enables offline, privacy-preserving biometric processing at the edge[14][15]. Several studies have explored gesture and face recognition independently: Ilhamy et al. .[9]combined LSTM and GRU for real-time sign language recognition, Mensah et al. [16] assessed FaceNet performance under multiple constraints, and Jha et al. [17]developed a desktop application combining gesture and face recognition without embedded deployment or physical actuation.

Despite these advancements, no existing study integrates gesture and face verification sequentially on a single embedded platform with physical hardware actuation for high-security applications. This study addresses this gap by proposing a sequential fusion system on Raspberry Pi 5 that employs hand gesture passwords as a behavioral gating mechanism before physiological face verification, with successful authentication triggering a solenoid door lock through GPIO-controlled relay circuitry. The key contributions of this work are: (1) a sequential fusion architecture that reduces computational load by conditionally activating face recognition only after gesture validation, (2) an empirically recalibrated face verification threshold derived from intrusion testing rather than default model parameters, and (3) a complete edge-deployed prototype demonstrating end-to-end integration from biometric inference to physical actuation.

2. Research Methodology

2.1 System Architecture

The system implements a sequential fusion approach [5] with two authentication layers. The first layer requires users to perform two correct hand gesture passwords in sequence. Upon successful gesture verification, the second layer activates face recognition through embedding comparison. This sequential design ensures the computationally intensive face recognition module only runs after confirmed intentional interaction, reducing unnecessary processing and improving system efficiency on the edge device [14].

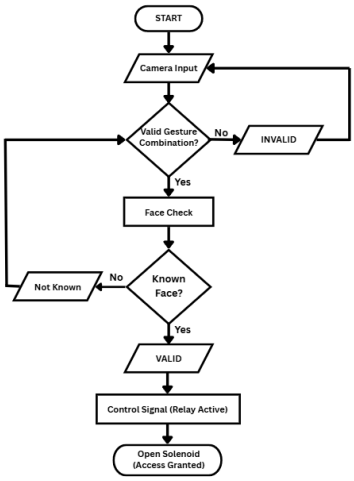

Figure. 1 System workflow of sequential fusion access control

Figure 1 illustrates the complete workflow of the proposed system. The process begins with the camera capturing the user's hand gesture. The system then classifies the first gesture using the LSTM model. If the first gesture matches the assigned password, the system proceeds to capture and verify the second gesture. Upon successful verification of both gestures, the system transitions to the face recognition stage, where the captured face embedding is compared against stored references. If the Euclidean distance falls below the 0.34 threshold, access is granted and the solenoid door lock is activated. If any stage fails, access is denied and the system returns to the initial state.

The system is powered by a Raspberry Pi 5 device, featuring a Broadcom BCM2712 chip with an ARM Cortex-A76 processor operating at 2.4 GHz) [10] with a Logitech C270 webcam for visual input. System output is actuated through a GPIO-connected relay module controlling a 12V solenoid door lock.

2.2 Hand Gesture Recognition

The gesture recognition pipeline consists of spatial feature extraction using MediaPipe Hands [7][8], which extracts 21 three-dimensional landmarks (63 values per frame).

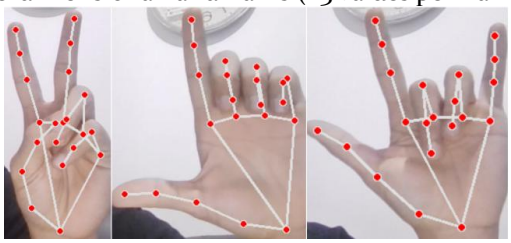

Figure. 2 MediaPipe hand landmark extraction results on gesture classes

As illustrated in Figure 2, MediaPipe successfully detects and maps 21 hand landmarks on each gesture class used in this study, including neutral hand position, peace, metal sign, and L. The red dots indicate the detected landmark positions, while the connecting lines represent the structural relationships between joints on the palm and fingers. These extracted landmark coordinates are converted into numerical data sequences that serve as input for the LSTM classification model. Temporal classification is performed using an LSTM network.[19], which captures dynamic gesture patterns through its gating mechanism (forget, input, and output gates), classifying gestures into four classes: peace, metal sign, L, and unknown. The training dataset comprises 69,600 samples constructed using a sliding window of 20 consecutive frames from recorded gesture sessions across multiple subjects performing each gesture class. The dataset size follows recommendations for temporal sequence classification, where sufficient samples per class are required to capture gesture variability and prevent overfitting [21]. The dataset was split into 80% training (55,680 samples) and 20% testing (13,920 samples) using stratified random sampling to maintain class distribution proportionality. Normalization (wrist-centered translation and palm-size scaling) and spatial augmentation (random rotation ±15°, scale variation 0.85x–1.15x, Gaussian noise σ=0.01, and X-axis mirroring) were applied to improve generalization[20].

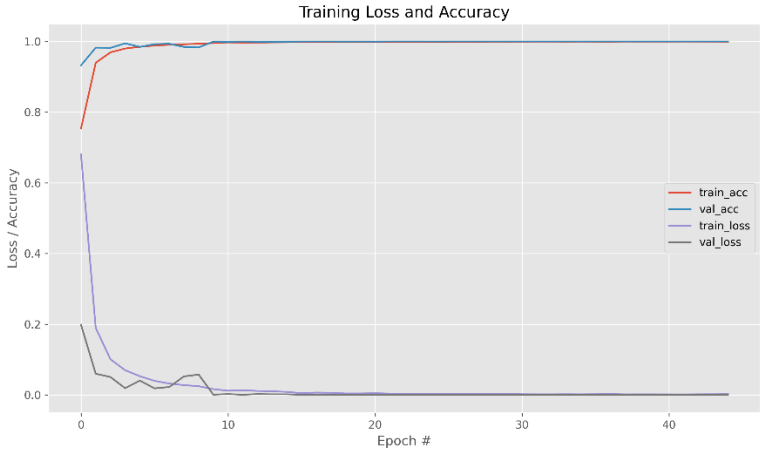

Figure. 3 Training loss and accuracy curves across epochs

Figure 3 shows the training and validation metrics across epochs. The training accuracy (train_acc) along with validation accuracy (val_acc) metrics climb progressively toward a perfect score of 1.0, whereas both the training and validation errors drop steadily approaching zero. The strong correspondence across the training and validation trajectories demonstrates the network's ability to grasp the temporal hand movements while successfully avoiding both overfitting and underfitting issues. The final evaluation on the test set yielded a Test Accuracy of 99.97% and a Test Loss of 0.0011.

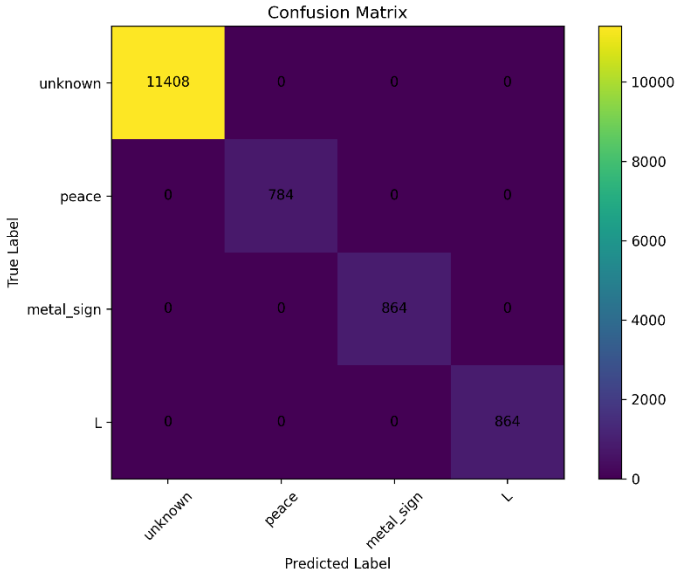

Figure. 4 Training loss and accuracy curves across epochs

Figure 4 presents a confusion matrix that validates the categorization efficacy for every one of the four distinct hand signals. From a total of 13,920 test samples, the model achieved 100% accuracy with a macro F1-score of 1.00 [21]. Each class was predicted correctly without any misclassification: peace (784 samples), metal sign (864 samples), L (864 samples), and unknown (11,408 samples). These results demonstrate that the LSTM architecture with normalized and augmented MediaPipe landmarks produces a reliable gesture classifier suitable for the first verification layer.

2.3 Face Recognition

The face recognition module uses dlib's pre-trained ResNet-34 model [12] to extract 128-

dimensional face embeddings. This architecture employs deep residual learning [13] to generate discriminative feature vectors from facial images. Verification is performed by computing the Euclidean distance between input and stored reference embeddings[16][22].

The face enrollment dataset consists of 148 facial images across 3 registered subjects, from which 101 valid face encodings were extracted (68.2% extraction rate). The participant pool comprises 6 individuals (3 registered, 3 unregistered) of mixed gender, selected to represent a realistic small-team operational scenario typical of military weapon storage facilities where access is restricted to a limited number of authorized personnel. The threshold selection procedure follows a two-stage approach. First, during training, the system computes intra-class distances (variation within the same person's encodings) and inter-class distances (separation between different persons). The optimal threshold is initially calculated as the midpoint between the maximum intra-class distance and the minimum inter-class distance. The training phase produced a recommended threshold of 0.42. Second, during validation testing with unregistered subjects, the threshold is empirically verified and recalibrated if security violations are identified. As detailed in Section 3.3, intrusion testing revealed that the 0.42 threshold was insufficient, prompting recalibration to 0.34. This two-stage approach—automated calculation followed by empirical validation—ensures the threshold is both mathematically grounded and operationally verified [23].

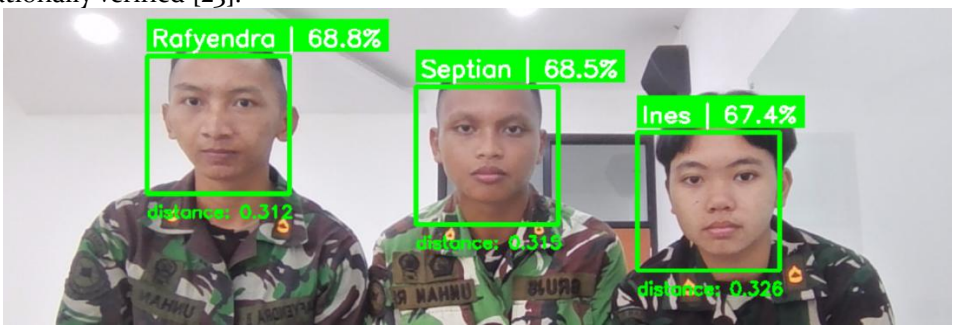

Figure. 5 Face recognition results on three registered users

Figure 5 shows the face recognition system successfully identifying three registered users (Rafyendra, Septian, and Ines) simultaneously in a single frame. The system displays each user's identity label along with the corresponding Euclidean distance values, all below the 0.34 threshold. This result demonstrates that the face recognition module is capable of real-time multi-face identification as the final verification layer.

2.4 Evaluation Metrics

The effectiveness of the security architecture is assessed relying on the False Reject Rate (FRR) alongside the False Accept Rate (FAR) metrics, in accordance with ISO/IEC [24]:

$$\boldsymbol{FRR} = \frac{\boldsymbol{False\ rejections}}{\boldsymbol{Total\ Genuine\ attempts}} \times \mathbf{100\%} \tag{1}$$

$$\boldsymbol{FAR} = \frac{\boldsymbol{False\ acceptances}}{\boldsymbol{Total\ impostor\ attempts}} \times \mathbf{100\%} \tag{2}$$

Testing involved 3 registered users and 3 unregistered users for security evaluation, plus 3 additional unregistered subjects for independent face model testing, each performing multiple authentication attempts under controlled conditions.

3. Results and Discussion

All experiments were conducted under the following standardized conditions: indoor environment with stable artificial lighting, Logitech C270 HD 720p webcam positioned statically at a horizontal level,

and an operational distance of 30 to 50 cm between the subject and the camera lens.

3.1 Hand Gesture Recognition Results

Gesture recognition was tested across four hand orientation angles (0°, 45°, 90°, 180°) for both hands.

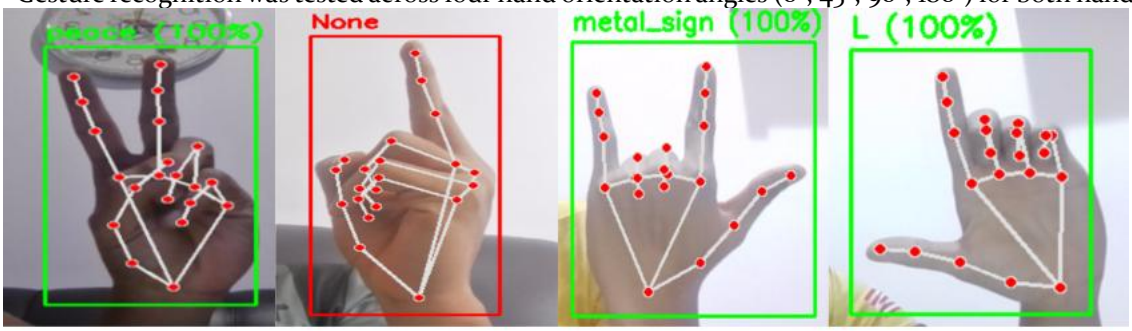

Figure. 6 Sample gesture captures at different orientations: (a) Peace at 0°, (b) Peace at 90°, (c) Metal Sign at 180°, (d) L at 0°

Figure 6 presents representative gesture captures at different hand rotation angles. Image (a) shows the peace gesture at 0° (palm facing camera), successfully recognized. Image (b) shows the same gesture at 90° (hand sideways), where the system failed because only the edge of the hand is visible. Image (c) demonstrates the metal sign gesture at 180° (back of hand facing camera), still successfully recognized due to its distinctive finger pattern. Image (d) shows the L gesture at 0°, the only orientation at which this gesture could be reliably detected. Table 1 summarizes the complete results.

Table 1. Hand Gesture Recognition Results Across Orientations

Gesture	Hand	0° (Palm)	45° (Tilted)	90° (Side)	180° (Back)
Peace	Right	✓	✓	✗	✓
Peace	Left	✓	✓	✗	✓
Metal Sign	Right	✓	✓	✗	✓
Metal Sign	Left	✓	✓	✗	✓
L	Right	✓	✗	✗	✗
L	Left	✓	✗	✗	✗

The orientation tolerance results reveal a clear distinction between gesture types based on their geometric properties. Peace and metal sign gestures achieved a 75% orientation recognition rate (recognized at 0°, 45°, and 180°), while L was limited to 25% (recognized only at 0°). This disparity is attributable to the fundamental difference in finger orientation patterns: peace and metal sign rely on vertically extended fingers that protrude above the hand silhouette regardless of palm orientation, maintaining recognizable landmark configurations even from the dorsal view. In contrast, the L gesture depends on the angular relationship between a laterally extended thumb and vertically extended index finger. At 45°, the thumb undergoes perspective foreshortening as it points partially toward or away from the camera, distorting the characteristic 90° angular pattern. At 180°, the thumb is occluded behind the palm mass, eliminating the defining spatial feature entirely.

This finding is consistent with Zhang et al[7]. who noted that MediaPipe landmark accuracy degrades when fingers are occluded or foreshortened. The 90° failure across all gestures confirms that edge-on hand orientations provide insufficient visual information for reliable landmark extraction, a known limitation of single-camera 2D-projected 3D landmark estimation. Compared to Ilhamy et al. [9], who reported gesture recognition primarily in frontal orientations, this study extends the evaluation to include multi-angle robustness analysis, providing a more comprehensive assessment of real-world deployment constraints.

3.2 Face Recognition on Registered Users

Face recognition was tested across three registered subjects (Ines, Rafyendra, and Septian) at three head orientations: 0° (straight), 45° rightward tilt, and 45° leftward tilt. Table 2 presents the

Euclidean distance values.

Table 2. Euclidean Distance Values for Registered Users

Subject	0° (Straight)	45° Right	45° Left	Threshold
Ines	0.228	0.257	0.269	0.34
Rafyendra	0.266	0.291	0.310	0.34
Septian	0.241	0.352	0.320	0.34
Average	0.245	0.300	0.300	0.34

The results in Table 2 reveal two analytically significant patterns. First, the monotonic increase in Euclidean distance from 0° (average 0.245) to 45° (average 0.300) quantifies the degradation effect of head orientation on embedding similarity. This 22.4% increase in average distance demonstrates that facial feature asymmetry—where the visible proportion of eyes, nose, and jawline changes with rotation—produces measurably different embedding vectors, consistent with findings by Mensah et al. [16] and Schroff et al.[22] on pose-induced embedding variance.

Second, the inter-subject variability is notable: Septian's 45° right distance (0.352) exceeded the threshold while Ines and Rafyendra remained within bounds at the same angle. This suggests that individual facial geometry significantly influences orientation tolerance, a factor that Kyrkou et al. [23] identified as critical for threshold calibration in dynamic environments. The overall average distance of 0.282 across all subjects and orientations provides a 17.1% margin below the 0.34 threshold, indicating that the system accommodates natural pose variation for compliant users.

3.3 Face Recognition on Unregistered Users

To evaluate the face model's standalone rejection capability, three unregistered subjects (Suci, Surya, and Daffa) were tested independently from the FAR evaluation. Each subject was tested at the same three orientations (0°, 45° right, 45° left). The system computed Euclidean distances against all stored references and displayed per-user distance breakdowns. Table 3 presents the closest Euclidean distance values.

Table 3. Euclidean Distance Values for Unregistered Users

Subject	0° (Straight)	45° Right	45° Left	Threshold
Suci	0.457	0.417	0.482	0.34
Surya	0.477	0.479	0.459	0.34
Daffa	0.529	0.479	0.459	0.34
Average	0.488	0.458	0.467	0.34

All unregistered subjects produced Euclidean distances well above the 0.34 threshold across every orientation, with an overall average of 0.471. The separation ratio between the registered mean (0.282) and unregistered mean (0.471) yields a discriminability factor of 1.67×, indicating clear class separation. For comparison, Schroff et al. [22] reported typical inter-class to intra-class ratios of 1.5–2.0× on the LFW benchmark, placing this system's discriminability within the expected range for ResNet-based face verification.

A critical finding emerged from the per-user distance analysis. Subject Suci produced a minimum distance of 0.416 against registered user Ines at 45° orientation, and subject Surya produced 0.418 against Ines at 0° orientation. Both values fall below the initial model-trained threshold of 0.42, meaning these unregistered subjects would have been falsely accepted under the default configuration. This vulnerability—where partial facial similarity between unrelated individuals approaches the decision boundary—has been documented in face verification literature as a persistent challenge, particularly in small enrollment databases where limited reference diversity narrows the inter-class margin [16][24].

This empirical evidence directly motivated the threshold recalibration from 0.42 to 0.34. The 0.08 reduction creates a registered-to-threshold margin of 0.058 (0.34 – 0.282) and an unregistered-to-threshold margin of 0.131 (0.471 – 0.34), effectively establishing a separation gap that eliminates the identified false positive risk while maintaining acceptance for compliant registered users.

3.4 System Security Testing (FAR and FRR)

Security testing evaluated the complete system (gesture + face verification) under controlled conditions.

Figure. 7 System interface: (a) Access granted, (b) Access denied

Figure 7 presents the system interface during access decisions. Panel (a) shows registered users (Ines, Septian, Rafyendra) successfully passing both verification layers, displaying "AKSES DIIZINKAN" with match percentages (73%, 76%, 69%). Panel (b) shows an unregistered user receiving "AKSES DITOLAK" despite entering the correct gesture sequence, confirming that face verification independently blocks unauthorized access.

Table 4. FRR Test Results

Registered User	Gesture Sequence	Attempts	Accepted	Rejected
Ines	Peace → L	10	10	0
Septian	Metal Sign → L	10	10	0
Rafyendra	Peace → Metal Sign	10	10	a0
	Total	30	30	0

FRR = 0/30 × 100% = 0%

Table 5. FAR Test Results

Unregistered User	Gesture Sequence	Attempts	Accepted	Rejected
Suci	Peace → L	10	0	10
Surya	Metal Sign → L	10	0	10
Daffa	Peace → Metal Sign	10	0	10
	Total	30	0	30

FAR = 0/30 × 100% = 0%.

The combined 0% FRR and 0% FAR demonstrates the effectiveness of the dual-layer sequential fusion approach under controlled conditions. The FAR result is particularly significant because unregistered subjects performed the correct gesture sequences, confirming that gesture password knowledge alone is insufficient for access—the face verification layer independently and reliably prevents unauthorized entry [4][6]. This aligns with the theoretical advantage of multimodal fusion described by Ross and Jain [6], where combining independent biometric modalities multiplicatively reduces error rates compared to unimodal systems.

However, it is important to contextualize these results within the experimental constraints. The 0% FRR is conditional upon adherence to operational procedures (frontal face positioning, stable lighting, palm facing camera). As demonstrated in Table 2, Septian was rejected at 45° right tilt, indicating that FRR would increase under non-ideal positioning. The sample size of 60 total trials, while sufficient for prototype validation, falls below the ISO/IEC 19795-1:2021 [24] recommendation for statistically robust biometric performance claims. These results should therefore be interpreted as prototype-level validation rather than production-grade certification metrics.

3.5 Hardware Actuation

The solenoid door lock was tested across 20 trials (10 accepted, 10 denied)

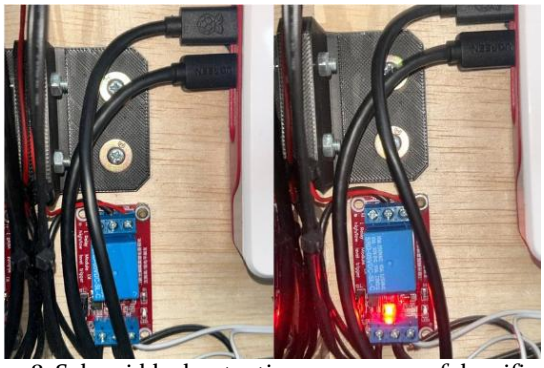

Figure. 8 Solenoid lock actuation upon successful verification

Figure 8 shows the relay module (SRD-05VDC) receiving the activation signal from the Raspberry Pi 5. When access is granted, the relay activates as indicated by the illuminated LED, subsequently energizing the solenoid to unlock the door for 2 seconds. In all 10 accepted trials, the solenoid successfully retracted. In all 10 denied trials, the solenoid remained locked. The total inference time from initial gesture capture to solenoid actuation was approximately 1.6 seconds (1.3 seconds for dual gesture recognition and 0.3 seconds for face verification), demonstrating real-time responsiveness suitable for operational deployment. No mechanical failures were observed, confirming reliable software-hardware integration[25].

3.6 Limitation

Several limitations should be acknowledged. The L gesture's strict orientation requirement (0° only) reduces operational flexibility compared to peace and metal sign. The 0.34 threshold, while enhancing security, increases sensitivity to head tilts beyond 45°, as evidenced by Septian's rejection at that angle. The evaluation involves 6 subjects and 60 authentication trials—sufficient for prototype validation but insufficient for generalizable statistical claims per ISO/IEC 19795-1:2021[24] standards. The system remains untested under variable lighting, outdoor environments, or extended continuous operation. Critically, no anti-spoofing or liveness detection mechanism is implemented; presentation attacks using photographs or video playback were not evaluated [3][4]. Future implementations should address these gaps through expanded participant diversity, environmental robustness testing, and integration of liveness detection techniques.

4. Conclusion

This study demonstrates that sequential fusion of hand gesture recognition and face recognition on a single edge device provides a viable dual-layer biometric access control solution for high-security facilities. The architecture's core design principle—conditional activation of face verification only upon successful gesture authentication—proved effective in both reducing computational overhead and establishing independent security barriers that cannot be circumvented by compromising a single modality. Empirical testing revealed that default model-trained thresholds may be insufficient for security-critical deployments, leading to the adoption of an operationally validated threshold that eliminated identified false acceptance vulnerabilities. The system achieved complete separation between registered and unregistered user populations under controlled conditions, with reliable hardware actuation confirming end-to-end integration feasibility. Future work should prioritize expanding participant diversity for statistically robust performance validation, implementing anti-spoofing countermeasures such as liveness detection and randomized gesture challenges, and evaluating system robustness under variable environmental conditions to establish deployment readiness beyond the prototype stage.

References

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Trans. Circuits*

Syst. Video Technol., vol. 14, no. 1, pp. 4–20, Jan. 2004, doi: 10.1109/TCSVT.2003.818349.

[2] A. Saldamli, G.; Dalli, A.; Taha, K.; Al-Fuqaha, "Security and privacy issues in physical and cyber biometric systems," pp. 282–289, 2020, doi: 10.1109/ICIoT48696.2020.9089455.

[3] J. Galbally, S. Marcel, and J. Fierrez, "Biometric Antispoofing Methods: A Survey in Face Recognition," *IEEE Access*, vol. 2, pp. 1530–1552, 2014, doi: 10.1109/ACCESS.2014.2381273.

[4] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A Survey of face manipulation and fake detection," *Inf. Fusion*, vol. 64, pp. 131–148, Dec. 2020, doi: 10.1016/j.inffus.2020.06.014.

[5] A. R. and A. Jain, "Multimodal biometrics: An overview," in *12th EUSIPCO*, 2004, pp. 1221–1224. doi: 0908.1417.

[6] M. S. H. and G. Muhammad, "An audio-visual emotion recognition system using deep learning fusion for a cognitive wireless framework," *IEEE Wirel. Commun.*, vol. 26, no. 3, pp. 62–68, 2021.

[7] F. Zhang *et al.*, "MediaPipe Hands: On-device Real-time Hand Tracking," Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.10214

[8] C. Lugaresi *et al.*, "MediaPipe: A Framework for Building Perception Pipelines," Jun. 2019, [Online]. Available: http://arxiv.org/abs/1906.08172

[9] A. A. Ilham, I. Nurtanio, Ridwang, and Syafaruddin, "Applying LSTM and GRU Methods to Recognize and Interpret Hand Gestures, Poses, and Face-Based Sign Language in Real Time," *J. Adv. Comput. Intell. Intell. Informatics*, vol. 28, no. 2, pp. 265–272, Mar. 2024, doi: 10.20965/jaciii.2024.p0265.

[10] B. Sundar and T. Bagyammal, "American Sign Language Recognition for Alphabets Using MediaPipe and LSTM," *Procedia Comput. Sci.*, vol. 215, pp. 642–651, 2022, doi: 10.1016/j.procs.2022.12.066.

[11] H.-J. Kim and S.-W. Baek, "Application of Wearable Gloves for Assisted Learning of Sign Language Using Artificial Neural Networks," *Processes*, vol. 11, no. 4, p. 1065, Apr. 2023, doi: 10.3390/pr11041065.

[12] D. E. King, "High quality face recognition with deep metric learning." Accessed: Mar. 15, 2026. [Online]. Available: http://blog.dlib.net/2017/02/high-quality-face-recognition-with-deep.html

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[14] H. Li, K. Ota, and M. Dong, "Learning IoT in Edge: Deep Learning for the Internet of Things with Edge Computing," *IEEE Netw.*, vol. 32, no. 1, pp. 96–101, Jan. 2018, doi: 10.1109/MNET.2018.1700202.

[15] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge Computing: Vision and Challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016, doi: 10.1109/JIOT.2016.2579198.

[16] J. A. Mensah, J. K. Appati, E. K. . Boateng, E. Ocran, and L. Asiedu, "FaceNet recognition algorithm subject to multiple constraints: Assessment of the performance," *Sci. African*, vol. 23, p. e02007, Mar. 2024, doi: 10.1016/j.sciaf.2023.e02007.

[17] A. Jha, Ishita, P. G. Shenwai, A. Batra, S. Kotian, and P. Modi, "GesSure: A Robust Face-Authentic Enabled Dynamic Gesture Recognition GUI Application," *Int. J. Cybern. Informatics*, vol. 11, no. 4, pp. 19–30, Aug. 2022, doi: 10.5121/ijci.2022.110402.

[18] and A. S. A. Indriani, M. Harris, "Applying hand gesture recognition for user guide application using MediaPipe," in *2nd ISSAT*, 2021, pp. 101–108.

[19] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," May 2016, [Online]. Available: http://arxiv.org/abs/1605.08695

[20] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.

[21] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.

[22] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2015, pp. 815–823. doi: 10.1109/CVPR.2015.7298682.

[23] J. Diez-Tomillo, J. M. Alcaraz-Calero, and Q. Wang, "Dynamic-Distance-Based Thresholding for UAV-Based Face Verification Algorithms," *Sensors*, vol. 23, no. 24, p. 9909, Dec. 2023, doi: 10.3390/s23249909.

[24] ISO/IEC, "ISO/IEC 19795-1:2021 Information technology — Biometric performance testing and reporting," 2021. [Online]. Available: https://www.iso.org/standard/73515.html

[25] M. S. U. and W. Slany, "Visual programming for human detection using FaceNet in Pocket Code," *Int. J. Interact. Mob. Technol.*, vol. 18, no. 13, pp. 382–396, 2024, doi: 10.3991/ijim.v18i13.49277.