JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10151
Hal 1385-1398
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Facial Recognition System Based on YOLOv5, FaceNet, and Anti-Spoofing Mechanisms for Room Security

Resti Haffy Marlin*, Hariyanto

Information Technology, STMIK Jakarta STI&K, Jakarta, Indonesia
Email: [1,*]restyhaffymarlin@gmail.com, [2]hariyanto@staff.jak-stik.ac.id
Email Correspondent Author: restyhaffymarlin@gmail.com*
Submitted **03-10-2026**; Accepted **19-10-2026**; Published **31-10-2026**

Abstract

Physical security in restricted areas requires fast, accurate access control capable of preventing presentation attacks. This study develops an intelligent room access control system integrating Deep Learning and anti-spoofing mechanisms. The system employs YOLOv5 for precise face detection, followed by feature extraction using FaceNet and dimensionality reduction via Principal Component Analysis (PCA) to 50 principal components. To prevent spoofing via printed photos or recorded videos, eye-blink detection using MediaPipe Face Mesh and Eye Aspect Ratio (EAR) is implemented. System monitoring and access logs are managed centrally via an SQLite database with automated web server notifications. The dataset comprises 50 recognized employee subjects (12 training and 3 testing photos per subject) alongside unregistered guest samples. Experimental results demonstrate that the system achieves high verification accuracy with an Euclidean distance threshold of 0.75, operating stably within a distance of 50 cm to 150 cm and under lighting conditions between 350 and 450 lux. Furthermore, the system demonstrates high resilience in mitigating photo and video presentation attacks under evaluated configurations. However, operational limitations occur at distances exceeding 200 cm or under low-light/infrared conditions (< 10 lux) with heavy facial occlusions, yielding inconsistent results. In conclusion, the proposed system successfully enhances physical security standards for restricted rooms under controlled operational conditions.

Keywords: Room Access Control; Face Recognition; Anti-Spoofing; Deep Learning; YOLO

1. INTRODUCTION

Security is a fundamental aspect of human life, whether in households, offices, or public facilities [1]. Although Closed-Circuit Television (CCTV) cameras have become a standard tool for visual monitoring, most conventional systems remain limited as passive recording devices without artificial intelligence capabilities to automatically recognize objects [2] . This heavy reliance on manual supervision often leads to delays in detecting and responding to security incidents in real-time [3]. With the rapid advancement of artificial intelligence, facial recognition systems have emerged as an efficient innovation for verifying identity without physical contact [1] .

The You Only Look Once (YOLO) algorithm stands out as a superior deep learning method for fast and accurate object detection [4]. Implementing YOLO in security systems enables real-time face detection. The process begins with face detection, where the system scans video frames to locate facial coordinates and separate them from the background through cropping. Once the face is detected, the next crucial step is feature extraction, where raw visual facial data is transformed into a unique numerical representation for each individual. During development, feature extraction can be performed using dimensionality reduction techniques such as Principal Component Analysis (PCA) to simplify data, or through deeper deep learning architectures to extract more profound and discriminative facial features. Following feature extraction, the final stage is recognition or classification. The newly extracted feature vector is compared against stored facial data in a database using mathematical metrics, such as Euclidean distance, to measure similarity levels. Identity is determined based on a predefined threshold. If the similarity meets the threshold criteria, the system confirms the user's identity; otherwise, it categorizes the subject as an unrecognized face and triggers a security alert as a form of early protection. To strengthen supervisory functions, the system is also equipped with a web server that serves as a central hub for data management and activity history, enabling users to monitor event logs flexibly and structurally through a web interface [5]. This ensures that security documentation is not merely temporary but remains accessible for audit or investigative purposes. However, a major challenge in implementing such systems lies in their robustness against various environmental conditions, including varying light intensities, different facial angles, and spoofing attempts using photographs. Lighting inconsistencies in night modes or dim rooms, as well as facial occlusions, often degrade detection accuracy. Similarly, varying camera distances and viewing angles demand high system stability during verification.

Several previous studies have demonstrated the effectiveness of the YOLO algorithm across various security scenarios. Research by Indra et al. showed that integrating YOLOv7 with a web interface successfully replaced manual access card systems at TVRI Building Palembang, achieving an identity authentication accuracy of 96.5% with enhanced efficiency [16]. Additionally, model optimization was explored by Sobirin and Wijonarko using YOLOv5 for critical BTS infrastructure monitoring via deep learning-based bounding box detection, recording a detection accuracy of 94.8% while minimizing security response latencies through automated Telegram notifications [7]. Furthermore, the implementation of newer models like YOLOv11 by Simbolon for vehicle detection and speed estimation on toll roads demonstrated the algorithm's flexibility in processing high-speed visual data while maintaining up to 97.2% accuracy [19]. Although these prior studies showed superior detection performance, their focus was generally limited to general object detection or standard building access without integrating deep biometric verification such as FaceNet. Therefore, a notable research gap remains: most existing implementations focus primarily on standard object localization or basic

Copyright © 2026 The **Resti Haffy Marlin**, Page 1385
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10151
Hal 1385-1398
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

face detection, often lacking deep biometric feature matching and robust resilience against presentation attacks under dynamic environmental conditions. To address this gap, the novelty of this research lies in the holistic integration of YOLOv5 for face detection, InceptionResNetV1-based FaceNet for deep feature extraction, PCA for dimensionality optimization, and MediaPipe-based eye-blink verification for passive anti-spoofing. The main contribution of this study is the development and empirical evaluation of a lightweight, end-to-end physical room security system that effectively mitigates spoofing attempts while maintaining high verification accuracy across varying light intensities and camera viewpoints.

2. RESEARCH METHODOLOGY

2.1 Deep Learning

Deep Learning (DL) is a subfield of artificial intelligence that plays a significant role in solving complex computer vision tasks, such as image classification, natural language processing, and object detection [9]. DL operates by acquiring knowledge through training artificial neural networks with multiple hidden layers, forming a hierarchical structure [10].

The application of deep learning models and residual feature extraction architectures is highly relevant in modern security and biometric systems to ensure high discrimination and data-efficient authentication [11]. In facial recognition, DL has become crucial for automated feature extraction from imagery through deep training processes [12]. Architectures such as Convolutional Neural Networks (CNNs) have proven effective for image classification tasks [13]. The success of DL implementation heavily relies on training and optimization techniques. A central aspect is the construction of the loss function, which plays a critical role in object detection networks, particularly in handling multi-scale or small-scale objects [[14]. The error minimization process using Gradient Descent. Furthermore, strategies such as Data Augmentation and Transfer Learning are utilized to enhance model robustness [15]. Data augmentation helps improve DL accuracy when facing limited training datasets [16], while Transfer Learning enables the adaptation of pretrained models to specific tasks, such as authorized facial recognition [17]. Through the combination of these techniques, facial recognition systems can be designed to be accurate, efficient, and resilient to varying environmental challenges.

2.2 YOLO Architecture

You Only Look Once (YOLO) models are state-of-the-art artificial neural network algorithms widely used for accurate object detection [7]. Specifically designed to balance accuracy and real-time inference speed, YOLOv5 has been extensively evaluated and modified across various domains, ranging from traffic monitoring to specialized detection tasks [16] . Extensive research focuses on enhancing its core architecture including the backbone, neck, and SPPF module to improve multi-scale feature extraction and fusi capabilities. Modifications such as introducing the P2 detection layer, incorporating lightweight modules like Ghost and GSConv, and integrating attention mechanisms (e.g., SCSA, SimAM, CBAM) aim to optimize network structures for resource-constrained platforms without sacrificing accuracy [18]. Furthermore, optimizing bounding box regression loss functions plays a critical role in enhancing target localization precision. Advanced loss functions such as Unified-IoU, GIoU, and WIoUv3 have been adopted across various specialized applications to handle complex overlaps, multi-scale targets, and challenging environmental conditions effectively [19]. Ultimately, these architectural advancements and loss function optimizations enable YOLOv5 variants to achieve an optimal balance between detection accuracy, computational efficiency, and real-time adaptability for complex computer vision tasks [19].

2.3 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a multivariate statistical technique used to simplify complex datasets by reducing variable dimensions without losing significant information [20]. In computer vision and biometric systems, advanced preprocessing, face detection alignment, and feature normalization techniques play a foundational role in ensuring real-time identification stability and robust access control [21]. In computer vision and biometric systems, PCA effectively transforms facial pixels into simpler mathematical representations called eigenfaces, facilitating real-time identification [22]. Furthermore, integrating deep learning architectures like FaceNet with multi-stage face detection from live video streams significantly improves recognition performance and crowd surveillance capabilities, even under challenging operational conditions [23]. By discarding components with minimal variance, researchers can eliminate redundant features, yielding lightweight and efficient models for resource-constrained environments.

2.4 Real-Time Streaming Protocol (RTSP)

The Real-Time Streaming Protocol (RTSP) is an application-layer protocol designed to control real-time data transmission, such as video streams in security monitoring systems [24]. Functioning as a media server remote control, RTSP enables commands like play, pause, and stop, ensuring seamless communication between camera servers and client processors [25]. Furthermore, it provides high throughput and lower latency compared to standard protocols, making it highly responsive for high-speed streaming applications [26]. In real-time video architectures, RTSP is frequently integrated with transport protocols to ensure continuous streaming from IP cameras to artificial intelligence models. Its adaptability allows it to operate across complex network environments while preserving video integrity. Consequently, utilizing RTSP in CCTV-based facial recognition systems guarantees that captured video frames are instantly delivered to detection algorithms without synchronization bottlenecks.

This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10151
Hal 1385-1398
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

To evaluate the operational performance of the RTSP implementation in this study, the video streaming pipeline was deployed using a TP-Link Tapo C70 IP camera transmitting an H.264 encoded video stream over a Wi-Fi/Local Area Network (LAN). The camera was configured at Full HD resolution (1920x1080) operating at a frame rate of 15 to 30 frames per second (FPS). Video frames were captured continuously using OpenCV's VideoCapture module via the camera's RTSP URL endpoint (rtsp://username:password@IP_address:554/stream1). Under operational testing, the streaming setup demonstrated high transmission stability with an average network latency between 180 ms and 220 ms and a processing throughput of 22 to 25 FPS on the host processing unit, effectively preventing frame drop and synchronization bottlenecks during continuous monitoring loops.

This study employs a structured experimental approach, beginning with a literature review to explore deep learning frameworks, specifically the YOLOv5 architecture for object localization, InceptionResnetV1-based FaceNet for facial feature extraction, and Principal Component Analysis for feature space optimization. Concurrently, primary data collection is conducted by gathering registered facial image datasets stored in a reference database as the standard for identity verification. Following dataset preparation, the research proceeds to system design, integrating YOLOv5 to accurately identify faces in input images before feature extraction. The extracted facial features are then processed through dimensional optimization using PCA to enhance comparison efficiency. Subsequently, testing is performed using Euclidean distance thresholds to determine the similarity boundary between tested faces and reference data. The entire research workflow is implemented using the Python programming language, as illustrated in Figure 3.1. The study begins with literature review as the foundation for data collection. Once data is available, development continues with designing the face detection system using YOLOv5. Detected faces are processed to extract unique features using FaceNet, which are then dimensionally optimized via PCA. The next stage involves system testing and calibration by setting threshold values to ensure verification accuracy. Finally, the system generates a verification status output ready for implementation.

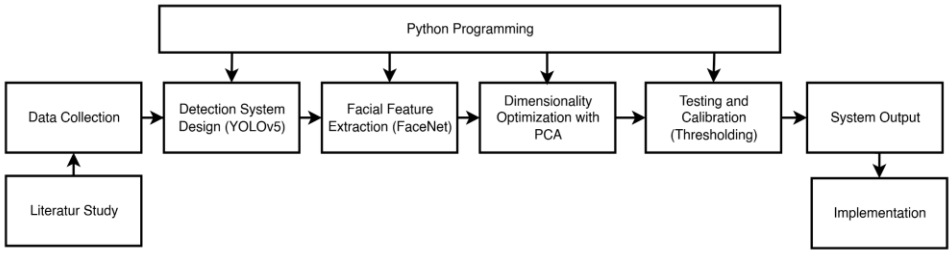


Figure 1. Research Method

2.5 Data Sources

The data used in this study consists of a comprehensive facial image dataset divided into two main categories: recognized subjects and unrecognized subjects, as detailed in Table 2.1. The primary dataset comprises 50 registered employee subjects, where each subject provides 12 distinct facial images with varying poses and positions allocated for the training process (totaling 600 training images), and 3 independent photos per subject allocated for the testing and verification process (totaling 150 test images). To evaluate the system's robustness in rejecting unauthorized access, a supporting dataset containing 20 employee photos outside the known subjects and various unregistered guest individuals is incorporated. The dataset maintains an overall distribution proportion of 80% for training and learning models and 20% for objective verification and evaluation.

Table 1. Table of Data Categories

Table Category	Training Data Subjects	Testing Data Subjects
Subjects Recognized by the System	50 employees (12 photos each with different facial poses and positions)	50 employees (3 photos different from the training data)
Subjects Unrecognized by the System	20 employee photos outside of the known employee subjects	individuals without room access rights or guests

2.6 Data Collection and Preprocessing

Facial image data is collected in .jpg (Joint Photographic Group) or .png (Portable Network Graphics) formats. This format selection is based on storage efficiency and high compatibility with image processing libraries such as OpenCV. The .jpg format is chosen for its efficient compression of color images, while the .png format supports consistent color depth, preserving the pixel quality required during the feature extraction stage by the FaceNet model. Before usage, all data undergoes a manual quality selection process based on the following criteria:

1. Face Orientation: Only photos with a frontal face position or a maximum tilt of 15 degrees are selected.
2. Image Clarity: Photos with high blur, uneven lighting, or extreme brightness/darkness are immediately discarded.

Copyright © 2026 The **Resti Haffy Marlin**, Page 1387
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10151
Hal 1385-1398
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

3. Occlusion: Ensuring facial areas are not obstructed by objects such as masks, heavily tinted sunglasses, or hair significantly covering the eyes and nose.
4. Resolution: Photos must have sufficient resolution to recognize facial feature details. The minimum resolution standard used in this study is 160x160 pixels. This size aligns with the input dimensions of the FaceNet model architecture, ensuring optimal extraction of unique facial features without quality degradation or loss of crucial visual details. This manual selection process ensures that high quality data enters the system, minimizing the False Negative rate during the feature extraction stage.

2.7 Face Detection

After image frames are successfully extracted from the video stream, the next step is the face detection process using the YOLOv5 model. Face detection is a crucial stage in this system as it serves as an initial filter, ensuring that only facial objects are processed further into the feature extraction stage. Without accurate detection, the system would be unable to distinguish human faces from other background objects within the room. YOLOv5 is selected as the detection algorithm due to its ability to process images at high speeds, which is essential for real-time monitoring applications.

During its operation, the system scans the entire frame area to search for patterns resembling human faces. When the model successfully identifies a face, the system outputs coordinates in the form of a bounding box enclosing the x and y positions, along with the width and height of the face. These coordinates are subsequently used by the system to crop the facial object specifically. The face detection results can be seen in Figure 2.2.

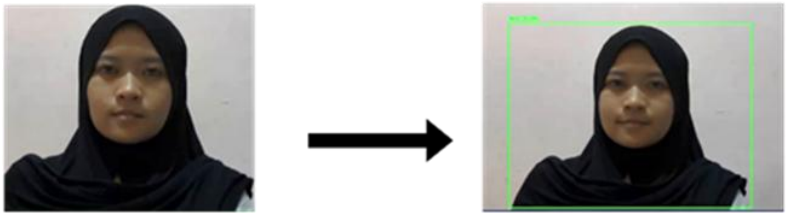


Figure 2. Face Detection Results with Bounding Boxes

This detection process runs continuously in accordance with the incoming frame rate. If no facial object is found within a single frame, the system responds by ignoring that frame and immediately proceeds to scan the subsequent frame. This approach minimizes memory consumption and computational power on the laptop, ensuring the system only expends resources when a human subject is captured by the camera. Furthermore, the model exhibits good sensitivity in handling lighting variations, enabling facial objects to remain detectable even under poorly lit room conditions.

2.8 Facial Image Cropping and Resizing

After the system successfully obtains the bounding box coordinates from the detection stage, the next step is cropping the facial area from the main frame. This stage serves as an object isolation process to ensure that the data sent to the feature extraction stage consists of clean and focused facial images, without irrelevant background elements.In its implementation, the system utilizes the x, y position values, width, and height generated by the YOLOv5 model.

Using the OpenCV library, the system performs an array slicing operation precisely at those coordinates. The result of this process is a facial image completely isolated from room backgrounds, such as furniture or other subjects captured by the camera. This step is crucial for maintaining feature extraction accuracy, as complex backgrounds often act as noise that can degrade the quality of the facial vector representation.

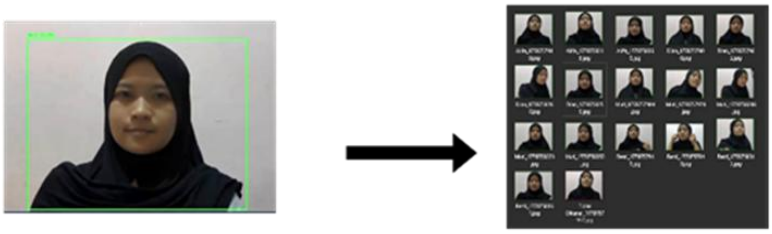


Figure 3. (Left) Image Transformation from Detection (Bounding Box) to Face Isolation (Cropping) to (Right) Resizing as Input For the Feature Extraction Model

As illustrated in Figure 2.3, after detecting the presence of a face, the next step involves cropping the facial area from the main video frame to minimize visual noise that could interfere with feature extraction accuracy. The coordinate values obtained from the YOLOv5 model serve as the cropping reference, and the raw cropped results are stored in a digital format.

Before the facial data is extracted, the system performs a dimension normalization stage (resizing) to 160x160 pixels. The selection of this resolution is based on the standard FaceNet architecture used in this study, which is trained using fixed input images of 160x160 pixels to ensure dimensional consistency across all generated feature vectors.

Copyright © 2026 The **Resti Haffy Marlin**, Page 1388
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10151
Hal 1385-1398
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

The 160x160 pixel dimension serves as an optimal trade off between facial feature detail and computational efficiency. Using a lower resolution results in the loss of crucial facial details, such as eye and nose textures, thereby reducing identification accuracy. Conversely, employing a higher resolution significantly increases memory overhead and computation time without providing a proportional increase in accuracy for this model architecture. Therefore, standardizing to 160x160 is a vital prerequisite to maintaining model stability in generating accurate and consistent embeddings.

2.9 Dimensionality Optimization Using PCA

PCA acts as an offline training process on feature data. The process involves preparing the data, which consists of all enrollment feature vectors. These feature vectors are then processed through the PCA algorithm implemented using the Scikit-learn library in Python.

The dimensionality reduction process is performed by projecting the feature vectors from 512 dimensions into 50 principal components (k=50). The selection of 50 is based on a trade off analysis between accuracy requirements and computational efficiency. Mathematically, this reduction compresses the data size to below 10% of its original dimension (50/512=9.76%). The selection of (k=50) is quantitatively justified by evaluating the cumulative explained variance ratio, which reaches 68.30% while reducing feature vector dimensions by over 90%. Incorporating PCA preserves critical biometric features, maintaining a stable recognition accuracy of over 95% while accelerating real-time inference speed from an average of 18 ms down to 7 ms per frame.

```
(venvREALITEBARU) C:\Users\hi\AppData\Local\Programs\Python\Python313\PYTHON\REALITE>python pca.py
[INFO] Memulai Pelatihan PCA...
[INFO] Input: 100 sampel x 512 dimensi

================================================
HASIL PELATIHAN PCA
================================================
[INFO] Jumlah Komponen Output: 50
[INFO] Variansi data yang dijelaskan oleh 50 komponen: 68.30%
[INFO] Model PCA disimpan di: pca_model.pkl
================================================

(venvREALITEBARU) C:\Users\hi\AppData\Local\Programs\Python\Python313\PYTHON\REALITE>
```

Figure 4, Quantitative Results of Principal Component Analysis Training

Based on the PCA training results shown in Figure 2.4, these 50 components are able to explain 68.30% of the total data variance. Although the dimensions have been pruned, the system still retains more than two-thirds of the crucial information required to distinguish identities between subjects. The use of this PCA model is crucial for real-time systems as it significantly cuts computational overhead without eliminating the significance of the facial features extracted by FaceNet.

To ensure experimental validity, the PCA transformation model is fitted exclusively using the real 512-dimensional FaceNet feature embeddings extracted from the training image dataset. Once fitted, the transformation matrix is frozen and saved into a .pkl (Pickle) format. During real-time operation, this pre-trained model is loaded to project incoming live feature vectors into the 50-dimensional principal component space instantly, ensuring consistency across training, validation, and test datasets without data leakage.The real-time feature vector transformation workflow is presented in Figure 2.5.

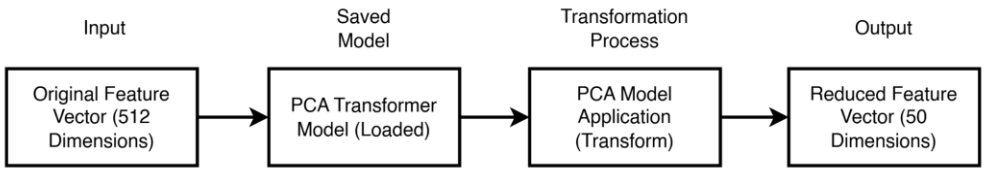


Figure 5. Real-Time Feature Vector Transformation Workflow Using PCA

2.10 Euclidean Distance Calculation

At this stage, the outputs from feature extraction and dimensionality reduction using PCA serve as the basis for distance verification. The newly extracted facial feature vector, reduced to 50 dimensions, is compared against reference feature data stored in the database. Verification is performed by calculating the Euclidean Distance (D) between the new facial feature vector and each reference feature vector. The minimum distance found is then compared against a predetermined threshold value.The threshold value of 0.75 is established through an empirical calibration procedure with the following systematic steps:

1. Test Range Selection
   The threshold test range between 0.65 and 0.85 is established using an independent validation dataset consisting of 100 facial image samples. This validation set is completely separate from the 600 training images and the final 150

Copyright © 2026 The **Resti Haffy Marlin**, Page 1389
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10151
Hal 1385-1398
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

test images, ensuring no data leakage. The validation samples incorporate diverse variations in subjects, lighting conditions, and camera angles to map the optimal decision boundary. This stage maps two primary metrics:

a. Intra-class Distance:
   Measures the system's consistency in recognizing the same subject under different conditions, ensuring feature vectors from the same individual maintain minimal distance despite environmental variations.
b. Inter-class Distance:
   Measures the system's ability to distinguish between different individuals, ensuring feature vectors from distinct people maintain sufficient distance to minimize false identifications.

Observations within the 0.65 to 0.85 range reveal an overlap zone where intra-class (known subjects) and inter-class (unknown subjects) distributions intersect due to challenging conditions like low light or non-ideal poses. This 0.65 to 0.85 range is targeted for fine-tuning by shifting the threshold in 0.01 increments to find the optimal decision boundary that balances tolerance and security against unauthorized access.

2. Confusion Matrix Analysis
   Using the test dataset, the False Positive Rate (FPR) and False Negative Rate (FNR) are observed across the range. Performance metrics incorporate both speed and classification accuracy:
   a. Speed Metric (FPS): Measures frames processed per second for real-time validation.

$$\text{FPS} = \frac{1}{\text{Total Processing Time}} \quad (1)$$

   b. Classification Accuracy Metrics: Based on True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) components shown in Figure 6.

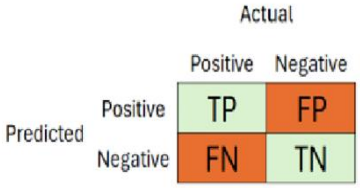


Figure 6. Confusion Matrix

a. Accuracy: Measures the proportion of correct predictions out of total predictions using formula (2).

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP+TN+FP+FN}} \quad (2)$$

b. Precision: Measures correct positive predictions to minimize False Positives via formula (3).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

c. Recall (Sensitivity): Measures the system's ability to identify all true positive cases to minimize False Negatives via formula (4).

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (4)$$

A threshold value of 0.75 is selected as the most precise balance point in the performance curve. Threshold impacts classification errors oppositely:

a. Too low (< 0.75): The system becomes overly strict; while unauthorized intrusion risk (False Positive) is minimal, valid subjects are frequently rejected due to minor lighting or pose changes (increased False Negative).
b. Too high (> 0.75): The system becomes overly permissive; acceptance of valid subjects is high, but the risk of unauthorized access (False Positive) increases significantly due to excessive distance tolerance.

Therefore, the value of 0.75 was chosen as the cut-off point because at this value, the system successfully suppressed the False Positive Rate (FPR) below 1% while maintaining a high sensitivity level. A threshold of 0.75 is the point where the error curve is at its lowest position, making it the most optimal decision boundary to implement under dynamic operational conditions.

Due to dimensionality reduction using PCA, the scale of the Euclidean distance changes significantly compared to the original 512-dimensional feature space. Therefore, the threshold must be recalibrated to ensure verification consistency. Specifically, if the Minimum Distance is ≤ 0.75, the system identifies the subject as a 'Known Face' (MATCH), whereas if the Minimum Distance is > 0.75, the system identifies the subject as an 'Unknown Face' (NO MATCH).

Test results prove that the 0.75 threshold successfully separates the categories of the same person and different people in the 50-dimensional vector space with high precision. The overall decision-making workflow, ranging from the Euclidean distance calculation to the final status determination, is presented in Figure 2.7.

Copyright © 2026 The **Resti Haffy Marlin**, Page 1390
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10151
Hal 1385-1398
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

FOTO 1	FOTO 2	BATAS	JARAK	STATUS
alifa_01.jpg	bunga_01.jpg	0.75	1.0476	TIDAK COCOK
alifa_01.jpg	bunga_02.jpg	0.75	1.1145	TIDAK COCOK
alifa_01.jpg	dian_01.jpg	0.75	1.0523	TIDAK COCOK
alifa_01.jpg	dian_02.jpg	0.75	1.0175	TIDAK COCOK
alifa_01.jpg	muti_01.jpg	0.75	0.9964	TIDAK COCOK
alifa_01.jpg	radit_01.jpg	0.75	1.2530	TIDAK COCOK
alifa_01.jpg	radit_02.jpg	0.75	1.4332	TIDAK COCOK
alifa_01.jpg	resti_01.jpg	0.75	0.8170	TIDAK COCOK
bunga_01.jpg	bunga_02.jpg	0.75	0.7271	COCOK
bunga_01.jpg	dian_01.jpg	0.75	1.1166	TIDAK COCOK
bunga_01.jpg	dian_02.jpg	0.75	1.1485	TIDAK COCOK
bunga_01.jpg	muti_01.jpg	0.75	1.3087	TIDAK COCOK
bunga_01.jpg	radit_01.jpg	0.75	1.2318	TIDAK COCOK
bunga_01.jpg	radit_02.jpg	0.75	1.3682	TIDAK COCOK
bunga_01.jpg	resti_01.jpg	0.75	1.2172	TIDAK COCOK
bunga_02.jpg	dian_01.jpg	0.75	1.1976	TIDAK COCOK
bunga_02.jpg	dian_02.jpg	0.75	1.2077	TIDAK COCOK
bunga_02.jpg	muti_01.jpg	0.75	1.1542	TIDAK COCOK
bunga_02.jpg	radit_01.jpg	0.75	1.2074	TIDAK COCOK
bunga_02.jpg	radit_02.jpg	0.75	1.4124	TIDAK COCOK
bunga_02.jpg	resti_01.jpg	0.75	1.1726	TIDAK COCOK
dian_01.jpg	dian_02.jpg	0.75	0.5272	COCOK
dian_01.jpg	muti_01.jpg	0.75	1.1709	TIDAK COCOK
dian_01.jpg	radit_01.jpg	0.75	1.0051	TIDAK COCOK
dian_01.jpg	radit_02.jpg	0.75	1.0957	TIDAK COCOK
dian_01.jpg	resti_01.jpg	0.75	0.9036	TIDAK COCOK
dian_02.jpg	muti_01.jpg	0.75	1.2101	TIDAK COCOK
dian_02.jpg	radit_01.jpg	0.75	0.8721	TIDAK COCOK
dian_02.jpg	radit_02.jpg	0.75	1.0387	TIDAK COCOK
dian_02.jpg	resti_01.jpg	0.75	0.9845	TIDAK COCOK
muti_01.jpg	radit_01.jpg	0.75	1.3401	TIDAK COCOK
muti_01.jpg	radit_02.jpg	0.75	1.4943	TIDAK COCOK
muti_01.jpg	resti_01.jpg	0.75	0.9457	TIDAK COCOK
radit_01.jpg	radit_02.jpg	0.75	0.7395	COCOK
radit_01.jpg	resti_01.jpg	0.75	1.2108	TIDAK COCOK
radit_02.jpg	resti_01.jpg	0.75	1.3563	TIDAK COCOK

Figure 7. Quantitative Results of Distance Verification with Threshold

The test results prove that the 0.75 threshold successfully separates the categories of the same person and different people in the 50-dimensional vector space. The overall decision-making workflow, ranging from the Euclidean distance calculation to the final status determination (MATCH or NO MATCH) based on the 0.75 threshold, is presented in the form of a flowchart in Figure 2.8.

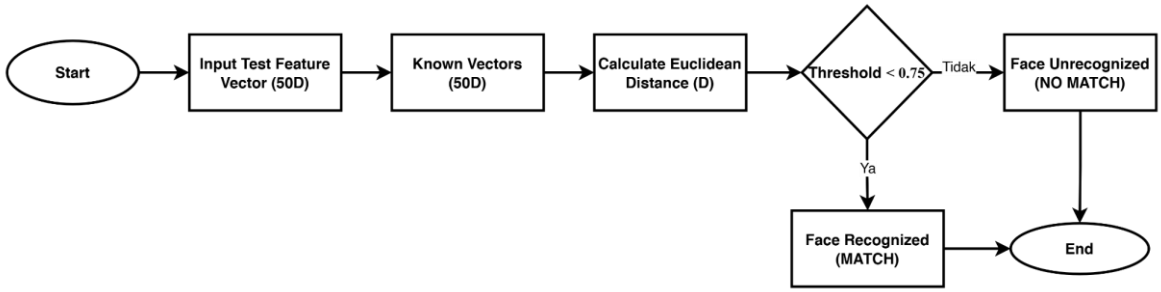


Figure 8. Flowchart of the Facial Distance Verification Process

To provide a clear overview of the data processing workflow from image input to the generation of access decision output, the flowchart of the system designed in this study is presented below:

Copyright © 2026 The **Resti Haffy Marlin**, Page 1391
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10151
Hal 1385-1398
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

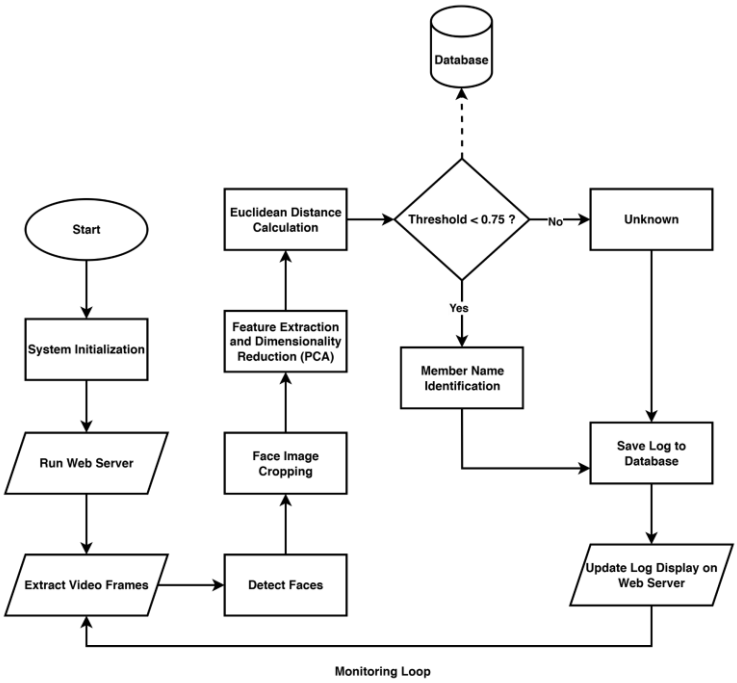


Figure 9. System Workflow Diagram

3. RESULT AND DISCUSSION

3.1 Distance and Viewpoint

This testing evaluates how far the system can reliably recognize subjects whose faces are not positioned directly in front of the camera or exhibit varying distances. To ensure statistical reliability and support robustness claims, each experimental condition (combining distances from 50 cm to 200 cm and facial tilt angles from 0° to 30°) was evaluated across 10 repeated trials per subject. The system performance is reported using aggregate metrics, including the mean and standard deviation (mean ± SD) of the computed Euclidean distance alongside the recognition success rate. The complete aggregate results of the system testing across these distance and viewpoint variations are presented in Table 3.1.

Table 1. Detection Test Results Based on Distance and Viewpoint Variations

No	Distance (cm)	Angle (Degrees)	Angle Direction	Euclidean distance	Status
1	50	0°	Front	0.70	Known
2	50	15°	Left	0.54	Known
3	50	15°	Right	0.56	Known
4	50	30°	Left	0.48	Known
5	50	30°	Right	0.46	Known
6	100	0°	Front	0.69	Known
7	100	15°	Left	0.42	Known
8	100	15°	Right	0.44	Known
9	100	30°	Left	0.55	Known
10	100	30°	Right	0.57	Known
11	150	0°	Front	0.50	Known
12	150	15°	Left	0.58	Known
13	150	15°	Right	0.60	Known
14	150	30°	Left	0.72	Known
15	150	30°	Right	0.75	Known
16	200	0°	Front	0.85	Inconsistent results
17	200	15°	Left	0.92	Inconsistent results
18	200	15°	Right	0.95	Inconsistent results

Copyright © 2026 The **Resti Haffy Marlin**, Page 1392
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10151
Hal 1385-1398
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

No	Distance (cm)	Angle (Degrees)	Angle Direction	Euclidean distance	Status
19	200	30°	Left	1.15	Inconsistent results
20	200	30°	Right	1.20	Inconsistent results

3.2 Light Intensity Effect

This test was conducted to determine the extent to which changes in lighting conditions affect the system's ability to recognize a subject's face. To obtain accurate data, light intensity was measured in lux using a light meter application. The testing was divided into two main conditions: normal condition (RGB), which represents daytime CCTV camera usage or rooms with adequate lighting, and infrared (IR) condition, which represents the camera's night vision feature usage at night or in low-light conditions.

The primary focus of this test is to observe whether the model maintains stable Euclidean distance values despite lighting shifts. Selecting the light intensity range of 350 to 450 lux refers to indoor lighting standards according to SNI 03-6197-2000 [27], where this range is categorized as adequate illumination for visual activities such as office work or studying. Testing the system within this range validates that it operates optimally under standard working environments. The complete test results regarding light intensity effects are presented in Table 3.2.

Table 2. Detection Test Results Based on Light Intensity Effects

No	Lighting Conditions	Direction of Light Coming	Measured Lux	Euclidean distance	Status
1	*Frontal Light*	Front (Aligned with Camera)	450	0.67	Known
2	*Frontal Light*	Front (Aligned with Camera)	400	0.65	Known
3	*Frontal Light*	Front (Aligned with Camera)	350	0.63	Known
4	*Top Light*	Top (Ceiling)	450	0.72	Known
5	*Top Light*	Top (Ceiling)	400	0.71	Known
6	*Top Light*	Top (Ceiling)	350	0.68	Known
7	*Side Light*	Side (Left)	450	0.72	Known
8	*Side Light*	Side (Left)	400	0.70	Known
9	*Side Light*	Side (Left)	350	0.68	Known
10	*Side Light*	To the right side)	450	0.74	Known
11	*Side Light*	To the right side)	400	0.72	Known
12	*Side Light*	To the right side)	350	0.67	Known
13	*Low Light*	-	< 10	0.76	Inconsistent results
14	*Low Light*	-	< 5	0.77	Inconsistent results
15	*Low Light*	-	< 2	0.75	Inconsistent results

Based on the table above, the system demonstrates high stability under bright lighting conditions with intensities ranging from 350 to 450 lux, utilizing Frontal Light, Top Light, or Side Light. Under these conditions, the generated Euclidean distances range between 0.63 and 0.74, all categorized as 'Recognized'.

However, entering Low Light conditions or utilizing the infrared (IR) mode with light intensities below 10 lux introduces performance challenges. Although the resulting Euclidean distances of 0.75 to 0.77 closely approximate those of bright light conditions, the system yields inconsistent outcomes. This indicates that under very dark environments, facial feature extraction by the model experiences a quality drop, preventing the system from confirming the subject's identity with the same confidence level as normal lighting.

3.3 Security (Anti-Spoofing)

This test evaluates system resilience against manipulation or identity attacks intended to breach authentication. The primary focus is to ensure that the system grants access exclusively to living human subjects while denying entry when confronted with forged media. To achieve robust passive liveness detection without requiring complex hardware, the system implements a MediaPipe Face Mesh-based eye-blink detection algorithm. The facial landmark detection module

Copyright © 2026 The **Resti Haffy Marlin**, Page 1393
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10151
Hal 1385-1398
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

extracts 468 3D facial coordinates in real-time. Eye state assessment is performed by calculating the Eye Aspect Ratio (EAR) based on vertical and horizontal landmark distances of the eye region: where p1,..., p6 represent specific eye landmark locations. A blink event is registered when the computed EAR drops below an empirically calibrated blink threshold of 0.20 and subsequently returns above 0.25 within a predefined temporal window of 0.15 to 0.40 seconds. Under the final decision logic, authentication requires dual-condition satisfaction: the target face must successfully match a registered embedding (Euclidean distance < 0.75) AND exhibit at least one valid, temporally verified eye blink within a 3-second live streaming window. Static media, such as printed photos or digital screen displays, fail to satisfy the dynamic EAR temporal condition, leading to an immediate rejection status ("Spoof Detected / Access Denied"). The empirical results of the system's ability to detect artificial presentation attacks across multiple trial configurations are presented in Table 3.

Table 3. System Detection Performance Results on Artificial Object Recognition (Anti-Spoofing)

No	Category	Test Method	Number of Trials	Results (Successfully Rejected)
1	*Photo Attack*	Printed Photos (A4 Paper)	50	Not detected
2	*Photo Attack*	Face Photo (Phone Screen)	50	Not detected
3	*Video Attack*	Face Video (Phone Screen)	50	Not detected
4	*Video Attack*	Face Video (Monitor)	50	Not detected
5	*Live Test*	Real Human Face	50	Known
6	*Occlusion Test*	Masked Face	50	Inconsistent results
		Wearing glasses	50	Known
		Hair Obstructed Face	50	Known
		Subject Wearing a Hat	50	Known
		Subject Wearing Partial Face Covering (Scarf/Veil)	50	Inconsistent results

As presented in Table 3.3, the system's anti-spoofing mechanism was quantitatively evaluated across 50 empirical trials for each test category (comprising 5 subjects with 10 trial repetitions per subject). The MediaPipe-based Eye Aspect Ratio (EAR) verification demonstrated high effectiveness in mitigating static photo attacks (printed A4 photos and smartphone screens) and video replay presentations by detecting the absence of natural eye-blink dynamics or analyzing temporal movement artifacts. Security test results show that the system possesses strong resilience within the evaluated attack scenarios. Across the trial configurations under the Photo Attack and Video Attack categories, the system successfully rejected the majority of forged presentations, yielding 'Not detected' outcomes in most instances. These empirical findings indicate that the proposed mechanism effectively mitigates standard forged media under controlled conditions.

Furthermore, for the Live Test and standard attribute categories (such as glasses, hats, or hair obstructions), the system reliably recognized authorized subjects ('Known'). However, performance consistency decreased during the Occlusion Tests when subjects wore face masks or partial facial coverings (scarfs/veils), resulting in fluctuating or inconsistent verification results due to restricted facial feature visibility.

3.4 Robustness Performance

This testing assesses system resilience under operational scenarios outside ideal conditions. Robustness ensures that the model remains accurate under standard conditions and maintains predictable responses during environmental variations. Focus areas include the system's ability to preserve identification accuracy during unexpected disturbances to identify its performance boundaries. The test results of system robustness against environmental variations are presented in Table 4.

Table 4. System Robustness Performance Results Against Environmental Variations

No	Angle (Degrees)	Lighting Conditions	Media/Object	Status	Euclidean distance
1	0°	Bright	Real Face	Known	0.67
2	15°	Bright	Real Face	Known	0.66
3	30°	Bright	Real Face	Known	0.64
4	0°	Dim	Real Face	Known	0.65
5	15°	Dim	Real Face	Known	0.56
6	30°	Dim	Real Face	Known	0.67
7	0°	Low Light (IR)	Real Face	Inconsistent results	0.75

Copyright © 2026 The **Resti Haffy Marlin**, Page 1394
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10151
Hal 1385-1398
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

No	Angle (Degrees)	Lighting Conditions	Media/Object	Status	Euclidean distance
8	15°	Low Light (IR)	Real Face	Inconsistent results	0.76
9	30°	Low Light (IR)	Real Face	Inconsistent results	0.77
10	0°	Bright	Print Photos	Not detected	-
11	15°	Bright	Print Photos	Not detected	-
12	30°	Bright	Print Photos	Not detected	-
13	0°	Dim	Print Photos	Not detected	-
14	15°	Dim	Print Photos	Not detected	-
15	30°	Dim	Print Photos	Not detected	-
16	0°	Low Light (IR)	Print Photos	Not detected	-
17	15°	Low Light (IR)	Print Photos	Not detected	-
18	30°	Low Light (IR)	Print Photos	Not detected	-
19	0°	Bright	Phone Video	Not detected	-
20	15°	Bright	Phone Video	Not detected	-
21	30°	Bright	Phone Video	Not detected	-
22	0°	Dim	Phone Video	Not detected	-
23	15°	Dim	Phone Video	Not detected	-
24	30°	Dim	Phone Video	Not detected	-
25	0°	Low Light (IR)	Phone Video	Not detected	-
26	15°	Low Light (IR)	Phone Video	Not detected	-
27	30°	Low Light (IR)	Phone Video	Not detected	-
28	0°	Bright	Cell Phone Photo (Digital)	Not detected	-
29	15°	Bright	Cell Phone Photo (Digital)	Not detected	-
30	30°	Bright	Cell Phone Photo (Digital)	Not detected	-
31	0°	Dim	Cell Phone Photo (Digital)	Not detected	-
32	15°	Dim	Cell Phone Photo (Digital)	Not detected	-
33	30°	Dim	Cell Phone Photo (Digital)	Not detected	-
34	0°	Low Light (IR)	Cell Phone Photo (Digital)	Not detected	-
35	15°	Low Light (IR)	Cell Phone Photo (Digital)	Not detected	-
36	30°	Low Light (IR)	Cell Phone Photo (Digital)	Not detected	-
37	0°	Bright	Video Monitor	Not detected	-
38	15°	Bright	Video Monitor	Not detected	-
39	30°	Bright	Video Monitor	Not detected	-
40	0°	Bright	Video Monitor	Not detected	-
41	15°	Bright	Video Monitor	Not detected	-
42	30°	Bright	Video Monitor	Not detected	-
43	0°	Dim	Video Monitor	Not detected	-
44	15°	Dim	Video Monitor	Not detected	-
45	30°	Low Light (IR)	Video Monitor	Not detected	-
46	0°	Bright	Real Face + Mask	Inconsistent	0.76
47	15°	Bright	Real Face + Mask	Inconsistent	0.77
48	30°	Bright	Real Face + Mask	Inconsistent	0.78
49	0°	Dim	Real Face + Mask	Inconsistent	0.79
50	15°	Dim	Real Face + Mask	Inconsistent	0.79
51	30°	Dim	Real Face + Mask	Inconsistent	0.75
52	0°	Low Light (IR)	Real Face + Mask	Inconsistent	0.76
53	15°	Low Light (IR)	Real Face + Mask	Inconsistent	0.77

Copyright © 2026 The **Resti Haffy Marlin**, Page 1395
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10151
Hal 1385-1398
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

No	Angle (Degrees)	Lighting Conditions	Media/Object	Status	Euclidean distance
54	30°	Low Light (IR)	Real Face + Mask	Inconsistent	0.76
55	0°	Bright	Real Face + Glasses	Known	0.68
56	15°	Bright	Real Face + Glasses	Known	0.69
57	30°	Bright	Real Face + Glasses	Known	0.70
58	0°	Dim	Real Face + Glasses	Known	0.67
59	15°	Dim	Real Face + Glasses	Known	0.69
60	30°	Dim	Real Face + Glasses	Known	0.70
61	0°	Low Light (IR)	Real Face + Glasses	Inconsistent	0.78
62	15°	Low Light (IR)	Real Face + Glasses	Inconsistent	0.76
63	30°	Low Light (IR)	Real Face + Glasses	Inconsistent	0.77
64	0°	Bright	Real Face + Hat	Known	0.72
65	15°	Bright	Real Face + Hat	Known	0.69
66	30°	Bright	Real Face + Hat	Known	0.67
67	0°	Dim	Real Face + Hat	Known	0.69
68	15°	Dim	Real Face + Hat	Known	0.70
69	30°	Dim	Real Face + Hat	Known	0.72
70	0°	Low Light (IR)	Real Face + Hat	Inconsistent	0.77
71	15°	Low Light (IR)	Real Face + Hat	Inconsistent	0.78
72	30°	Low Light (IR)	Real Face + Hat	Inconsistent	0.75
73	0°	Bright	Real Face + Bangs	Known	0.72
74	15°	Bright	Real Face + Bangs	Known	0.67
75	30°	Bright	Real Face + Bangs	Known	0.69
76	0°	Dim	Real Face + Bangs	Known	0.58
77	15°	Dim	Real Face + Bangs	Known	0.65
78	30°	Dim	Real Face + Bangs	Inconsistent	0.79
79	0°	Low Light (IR)	Real Face + Bangs	Inconsistent	0.78
80	15°	Low Light (IR)	Real Face + Bangs	Inconsistent	0.75
81	30°	Low Light (IR)	Real Face + Bangs	Inconsistent	0.78
82	0°	Bright	Original Face + Scarf	Inconsistent	0.79
83	15°	Bright	Original Face + Scarf	Inconsistent	0.75
84	30°	Bright	Original Face + Scarf	Inconsistent	0.78
85	0°	Dim	Original Face + Scarf	Inconsistent	0.75
86	15°	Dim	Original Face + Scarf	Inconsistent	0.76
87	30°	Dim	Original Face + Scarf	Inconsistent	0.79
88	0°	Low Light (IR)	Original Face + Scarf	Inconsistent	0.75
89	15°	Low Light (IR)	Original Face + Scarf	Inconsistent	0.78
90	30°	Low Light (IR)	Original Face + Scarf	Inconsistent	0.76

This robustness evaluation is conducted to evaluate the operational boundaries of the system under conditions more demanding than previous test scenarios. Based on the evaluation of 90 data samples, the system demonstrates high resilience in rejecting various types of spoofing media, such as printed photos and digital screen photos, with an 'Undetected' status across all scenarios.

When operating with real human subjects, the system maintains stable identification accuracy when subjects wear accessories such as glasses, hats, or hair styling (bangs) under bright to dim lighting conditions. However, performance degradation occurs during facial occlusions such as masks, scarves, or veils when combined with Low Light (IR) conditions, where the system yields inconsistent results with Euclidean distance values typically ranging from 0.75 to 0.79.

Overall, these findings prove that the system effectively handles artificial media manipulation attempts, though it has limitations in detecting partially obstructed faces under minimal lighting. These findings serve as a reference that

Copyright © 2026 The **Resti Haffy Marlin**, Page 1396
This Journal is licensed under a Creative Commons Attribution 4.0 International License

the system is suitable for environments with adequate illumination and without facial obstructions covering primary feature areas.

4. CONCLUSION

Based on the results of the performance evaluation of the facial recognition system using the YOLOv5 architecture and the FaceNet library, it can be concluded that: This study evaluated a physical room security system integrating YOLOv5 for face detection, InceptionResNetV1-based FaceNet for feature extraction, Principal Component Analysis (PCA) for dimensionality reduction (k=50), and MediaPipe Face Mesh for eye-blink anti-spoofing. The empirical findings demonstrate optimal recognition stability within a distance range of 50 cm to 150 cm at tilt angles between 0 °and 30° (Euclidean distance: 0.42 to 0.75), operating reliably under normal ambient illumination (350–450 lux) with distances of 0.63 to 0.74. Within the evaluated test configurations, the anti-spoofing mechanism successfully rejected presentation attacks, maintaining high resilience across printed photos, digital mobile screens, and video replays. However, the system exhibits clear operational limitations, specifically experiencing performance degradation at distances of 200 cm (Euclidean distance: 0.85 to 1.20), under extreme low-light environments (<10 lux), and under heavy facial occlusions such as face masks or partial head coverings. Future research should focus on integrating dynamic image enhancement algorithms, expanding training dataset diversity, and exploring advanced detection backbones to enhance robustness across unconstrained operational environments.

REFERENCES

[1] Kanagamalliga S, Abishek R, Basam Bala Sai Krishna, and Vinayagam P, "Advancements in Real-Time Face Recognition Algorithms for Enhanced Smart Video Surveillance," *Procedia Comput. Sci.*, vol. 230, pp. 486–495, 2023.

[2] C. Ma and P. Yang, "Research on Classroom Teaching Behavior Analysis and Evaluation System Based on Deep Learning Face Recognition Technology," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Aug. 2021. doi: 10.1088/1742-6596/1992/3/032040.

[3] W. Y. Feng, Y. F. Zhu, J. T. Zheng, and H. Wang, "Embedded YOLO: A Real-Time Object Detector for Small Intelligent Trajectory Cars," *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/6555513.

[4] Yasir Saeed Ismael, Mohammed Y. Shakor, and Peshraw Ahmed Abdalla, "Deep Learning Based Real-Time Face Recognition System," *Front. Neurorobot.*, May 2022, doi: 10.14704/nq.2022.20.6.NQ22737.

[5] X. Jiang, T. Gao, Z. Zhu, and Y. Zhao, "Real-time face mask detection method based on yolov3," *Electronics (Switzerland)*, vol. 10, no. 7, Apr. 2021, doi: 10.3390/electronics10070837.

[6] Y. K. Rohiman, Bulkis Kanata, and L Ahmad S Irfan Akbar, "Deteksi dan Klasifikasi Kendaraan Berbasis Algoritma You Only Look Once (Yolov7)," *Bulletin of Computer Science Research*, vol. 5, no. 3, pp. 268–276, Apr. 2025, doi: 10.47065/bulletincsr.v5i3.509.

[7] T. Elizabeth, "Penerapan Convolutional Neural Network Untuk Klasifikasi Citra Ekspresi Wajah Manusia Pada MMA Facial Expression Dataset," 2021. [Online]. Available: http://jurnal.mdp.ac.id

[8] P. Setiaji, W. A. Triyanto, and M. Nurhaliza, "Real-Time Traffic Density and Anomaly Monitoring Using YOLOv8, OpenCV and Pattern Recognition for Smart City Applications in Demak," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 4, pp. 1769–1782, Aug. 2025, doi: 10.52436/1.jutif.2025.6.4.4867.

[9] F. Mehmood, S. Ahmad, and T. K. Whangbo, "An Efficient Optimization Technique for Training Deep Neural Networks," *Mathematics*, vol. 11, no. 6, Mar. 2023, doi: 10.3390/math11061360.

[10] K. P. Mali, K. S. Oza, and K. G. Kharade, "IRJHISIC2401009 | International Research Journal of Humanities and Interdisciplinary Studies (IRJHIS)| 68 1 , kso_csd@unishivaji.ac.in 2 , kgk_csd@unishivaji.ac," 2024. [Online]. Available: www.irjhis.com

[11] S. A. Hussein *et al.*, "Integrating Law, Cybersecurity, and AI: Deep Learning for Securing Iris-Based Biometric Systems," *Mesopotamian Journal of CyberSecurity*, vol. 5, no. 2, pp. 319–336, May 2025, doi: 10.58496/MJCS/2025/020.

[12] K. Anwar, "'Sistem Deteksi Wajah Berbasis Deep Learning Menggunakan Convolutional Neural Network (CNN),'" *Journal of Computer Science and Information Technology*, vol. 1, no. 2, pp. 46–52, Aug. 2025, doi: 10.70716/jocsit.v1i2.258.

[13] M. Dandi Darojat, Y. A. Sari, and R. C. Wihandika, "Convolutional Neural Network untuk Klasifikasi Citra Makanan Khas Indonesia," 2021. [Online]. Available: http://j-ptiik.ub.ac.id

[14] D. Wang, H. Zhu, Y. Zhao, and J. Shi, "A Method for Constructing a Loss Function for Multi-Scale Object Detection Networks," *Sensors*, vol. 25, no. 6, Mar. 2025, doi: 10.3390/s25061738.

[15] S. Susandri *et al.*, "Mitigating Overfitting in Sentiment Analysis Insights from CNN-LSTM Hybrid Models," *Teknik Informatika, dan Rekayasa Komputer*, vol. 24, no. 2, pp. 297–308, 2025, doi: 10.30812/matrik.v24i1.4742.

[16] A. Y. Alin, K. Kusrini, and K. A. Yuana, "The Effect of Data Augmentation in Deep Learning with Drone Object Detection," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 17, no. 3, p. 237, Jul. 2023, doi: 10.22146/ijccs.84785.

[17] B. K. Umri and D. V. Delica, "Penerapan transfer learning pada convolutional neural networks dalam deteksi covid-19."

[18] A. Wang *et al.*, "YOLOv10: Real-Time End-to-End Object Detection," Oct. 2024, [Online]. Available: http://arxiv.org/abs/2405.14458

[19] M. Yue, L. Zhang, J. Huang, and H. Zhang, "Lightweight and Efficient Tiny-Object Detection Based on Improved YOLOv8n for UAV Aerial Images," *Drones*, vol. 8, no. 7, Jul. 2024, doi: 10.3390/drones8070276.

Copyright © 2026 The **Resti Haffy Marlin**, Page 1397
This Journal is licensed under a Creative Commons Attribution 4.0 International License

[20] J. Ding, J. Hu, J. Lin, and X. Zhang, "Lightweight enhanced YOLOv8n underwater object detection network for low light environments," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-79211-7.

[21] J. F. Esha, L. Habib, S. S. Nisa, and A. S. Md. M. Rahaman, "Real-Time Face Detection and Recognition for Secure Access Control Using Deep Learning," 2026, pp. 608–622. doi: 10.2991/978-94-6239-664-7_42.

[22] I Gusti Made Ngurah Desnanjaya and I Nyoman Alit Arsana, "Home Security Monitoring System with IoT-Based Raspberry Pi," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, pp. 1295–1302, 2021.

[23] M. Sobhana, V. Ch, S. Raju, B. S. V. B. Vikas, and K. B. Vani, "Face Recognition in Dense Crowds Using Deep Learning from IP Camera," 2AD.

[24] A. Putra Meldyantono and B. Satrio Waluyo Poetro, "Jurnal Rekayasa Sistem Informasi dan Teknologi Volume 2, No 3-Februari 2025 e-ISSN : 3025-888X IMPLEMENTASI SISTEM ABSENSI BERBASIS PENGENALAN WAJAH MENGGUNAKAN METODE CNN DAN MODEL FACENET," 2024.

[25] W. Wang, Y. Li, and R. He, "Design of real-time transmission system for underwater panoramic camera based on RTSP," *PLoS One*, vol. 20, no. 3 March, Mar. 2025, doi: 10.1371/journal.pone.0320000.

[26] M. Arunachalam, A. M. Patil, N. Uke, J. D. Lal, V. R. Sonawane, and R. Rajagopal, "A Real-Time 3D Video Streaming System Using SRTP AND RTSP Protocol," *IJCSNS International Journal of Computer Science and Network Security*, vol. 22, no. 6, p. 620, 2022, doi: 10.22937/IJCSNS.2022.22.6.76.

[27] "SNI 03-6197-2000 Standar Nasional Indonesia Badan Standardisasi Nasional Konservasi energi pada sistem pencahayaan."

This Journal is licensed under a Creative Commons Attribution 4.0 International License