Published *online* on the journal's web page : http://innovatics.unsil.ac.id

Innovation in Research of Informatics (INNOVATICS)

| ISSN (Online) 2656-8993 |

SECURE EYE: An Intelligent Real-Time Surveillance System for Cyber Threat Detection and Monitoring

Rafi Adinata Rachmat[1], Dedy Hariyadi[2]

[1,2] *Universitas Jenderal Achmad Yani Yogyakarta, Yogyakarta, Indonesia*
[1]*rafiadinatar@gmail.com*, [2]*dedy@unjaya.ac.id*

ARTICLE INFORMATION

Article History:
Received: February 24th, 2026
Last Revision: March 20th, 2026
Published *Online*: March 30rd, 2026

KEYWORDS

YOLOv8,
Real-time surveillance,
Face recognition,
Spoofing detection,
CCTV

CORRESPONDENCE

Phone: 081358142269
E-mail: rafiadinatar@gmail.com

ABSTRACT

Conventional campus CCTV systems operate passively and remain vulnerable to presentation attacks, including printed images and digital replay. Although recent studies have advanced human detection, face recognition, and liveness verification, these components are typically evaluated independently, with limited validation of their integration under resource-constrained, CPU-only deployment environment. This study proposes SECURE EYE, an integrated real-time surveillance framework combining YOLOv8n-based human detection, SFace embedding with Support Vector Machine (SVM) classification, and rule-based liveness verification. The system was evaluated on a dataset of 1,050 facial images from 10 identities (80:20 split) and further assessed using 1080p RTSP video streams on a virtual machine (8 vCPU, 8 GB RAM, no GPU). The results show that the system achieves 84.29% face recognition accuracy, with F1-scores ranging from 0.75 to 0.93. The liveness detection module yields a True Positive Rate (TPR) of 80.00% and a True Negative Rate (TNR) of 87.50%, with a False Acceptance Rate (FAR) of 12.50% and a False Rejection Rate (FRR) of 20.00%. Real-time performance reaches 4.77 FPS with 0.21-second latency and moderate CPU utilization (16.89%). These findings demonstrate the feasibility of integrated surveillance with anti-spoofing capabilities in CPU-only deployment environment. However, the FRR of 20% indicates a significant usability limitation, highlighting the need for further optimization.

1. INTRODUCTION

CCTV systems are widely deployed in campus environments to support security and monitoring activities. However, in practice, many of these systems still function primarily as passive recording tools, often resulting in delayed threat detection and limited responsiveness. The integration of face recognition into surveillance systems further introduces reliability challenges, particularly due to vulnerability to presentation attacks such as printed images and digital replay [1]. These limitations hinder the ability of conventional systems to support cyber intelligence capabilities for proactive threat monitoring and behavioral analysis.

Recent advances in deep learning have significantly improved key components of intelligent surveillance systems. Object detection has benefited from the efficiency of YOLO-based models [2], [3], while embedding-based approaches such as SFace enable more discriminative and robust face recognition [4], [5]. In parallel, various liveness detection methods have been developed to mitigate spoofing attacks, ranging from traditional techniques to more advanced deep learning-based approaches [6], [7]. Several recent surveys have also highlighted rapid progress in face anti-spoofing, anomaly detection in surveillance, and robust face recognition under challenging [5].

Despite these advancements, most existing studies still evaluate detection, recognition, and liveness verification as separate modules or integrate only a subset of these components. For instance, several recent systems combine detection with recognition but omit liveness verification [8], [9], while others evaluate liveness detection in isolation without integrating it into a complete pipeline [10], [11]. Consequently, there remains limited understanding of how these components perform when deployed as a unified end-to-end system, particularly under

CPU-only deployment environment. This limitation becomes increasingly critical in real-world institutional environments, where computational resources are constrained and GPU acceleration is not always available. Achieving a balance between accuracy, robustness, and real-time performance under such conditions therefore remains a significant challenge.

This research proposes the SECURE EYE framework, which integrates human detection, face recognition, and liveness verification into a single pipeline optimized for CPU-only deployment environment. The system employs YOLOv8n for efficient human detection due to its favorable trade-off between speed and accuracy [12], [13]. For face recognition, SFace embeddings are combined with a Support Vector Machine (SVM) classifier to enable lightweight yet effective identity classification [4], [14]. In addition, a rule-based liveness verification mechanism based on Eye Aspect Ratio (EAR) and Laplacian variance is utilized to mitigate common presentation attacks while maintaining low computational overhead [10], [11]. All processing is performed locally to reduce latency and minimize reliance on external infrastructure, aligning with edge-computing principles.

This research develops and evaluates an end-to-end intelligent CCTV surveillance system under resource-constrained conditions. The evaluation focuses on face recognition accuracy, liveness detection robustness (TPR, TNR, FAR, and FRR), and computational performance (FPS, latency, and resource utilization). In addition, failure cases and trade-offs between security and usability are analyzed to provide a more comprehensive understanding of system behavior. This study bridges the gap in CPU-only intelligent surveillance by presenting an integrated and operationally validated framework that demonstrates how detection, recognition, and anti-spoofing can be effectively combined for real-time deployment in resource-constrained environments.

2. RELATED WORK

The development of intelligent surveillance systems has been significantly influenced by advances in deep learning-based computer vision. In modern CCTV analytics, three core components are commonly involved: human detection, biometric-c identity recognition, and liveness verification to mitigate presentation attacks [1], [5]. Although each of these components has demonstrated substantial progress, their integration into a unified and operationally validated framework remains limited.

Real-time human detection serves as the foundational stage in automated surveillance pipelines. The YOLO family of single-stage object detectors is widely adopted due to its efficiency and real-time processing capability [2], [3]. In CCTV-based environments, YOLOv8 provides competitive detection accuracy while maintaining computational feasibility under hardware constraints [12], [13]. Recent studies have further enhanced detection robustness through the incorporation of attention mechanisms and architectural refinements [5]. However, most existing implementations assess detection performance in isolation, without examining how detection inaccuracies propagate to downstream processes such as recognition and liveness verification. This limitation suggests that high detection accuracy alone does not necessarily translate into reliable end-to-end system performance.

For identity recognition, embedding-based approaches have become dominant due to their ability to produce discriminative feature representations. The SFace framework improves feature separability using a sigmoid-constrained hypersphere loss function, resulting in more stable identity encoding [4]. In closed-set recognition scenarios, embedding vectors are frequently combined with Support Vector Machine (SVM) classifiers to achieve efficient and robust classification with relatively low computational overhead [14]. While several surveillance systems integrate detection and recognition [8], [9], these approaches often exclude explicit liveness verification as part of the authentication pipeline. This indicates that recognition performance is commonly optimized independently, without considering its interaction with anti-spoofing mechanisms in practical deployments.

Biometric authentication systems remain vulnerable to presentation attacks, including printed images and replayed digital media [1]. Recent surveys highlight the rapid advancement of face anti-spoofing techniques, as well as the increasing sophistication of attack strategies [6], [7]. Deep learning-based approaches, including transformer-based models, have demonstrated improved robustness by capturing global facial representations [15]. However, such methods typically require substantial computational resources, limiting their applicability in resource-constrained environments. Lightweight alternatives, such as blink-based liveness detection using Eye Aspect Ratio (EAR) and texture-based analysis, have shown feasibility for real-time and CPU-based implementations [10], [11]. However, these methods are often evaluated independently rather than as part of a fully integrated pipeline combining detection and recognition. Importantly, even these lightweight approaches are rarely integrated with detection and recognition modules within a single end-to-end surveillance pipeline. Consequently, liveness verification is frequently treated as an auxiliary component, rather than being systematically validated within a unified surveillance framework.

Despite strong performance across individual components, comprehensive end-to-end integration of detection, recognition, and anti-spoofing within a single real-time CCTV system remains limited. Existing studies tend to evaluate modules separately or combine only partial components, without rigorously validating their interaction under CPU-only deployment environment [8], [9]. This gap highlights the need for an integrated, computationally efficient, and operationally validated surveillance framework capable of delivering reliable performance in real-world, resource-constrained environments such as campus systems.

3. METHODOLOGY

This research proposes an integrated real-time CCTV surveillance framework that combines human detection, identity recognition, and liveness verification within a unified operational architecture. The system is designed for CPU-only deployment environment and executed entirely within a Linux-based Virtual Machine (VM), simulating realistic campus infrastructure constraints.

The methodological contribution lies in the architectural integration of detection, embedding-based recognition, and anti-spoofing verification into a single end-to-end processing system validated under operational conditions.

3.1 System Deployment Architecture

The proposed system follows a layered deployment architecture consisting of: Data Acquisition Layer, Processing and Intelligence Layer, and Output and Monitoring Layer. The overall architecture is illustrated in Figure 1.

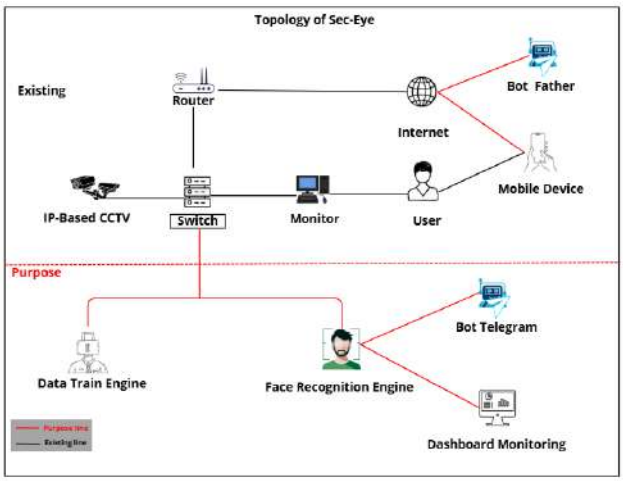

FIGURE 1. DEPLOYMENT ARCHITECTURE

In the data acquisition layer, IP-based CCTV cameras transmit video streams via RTSP through the local network infrastructure. These streams are received by a Linux-based Virtual Machine (8 vCPU, 8 GB RAM, no GPU) serving as the centralized processing engine. All computations are performed locally to reduce latency and minimize reliance on external infrastructure, following edge-computing principles that enhance system reliability. This layered separation minimizes redundant processing and allows each layer to be optimized independently for CPU-only deployment environment[16].

The output layer provides real-time visualization overlays, Telegram-based notification alerts, and structured logging of identity metadata. This modularity ensures a unified operational workflow. This architecture also supports cyber intelligence capabilities for proactive threat monitoring and behavioral analysis by enabling continuous identity tracking, anomaly detection, and real-time alert generation.

3.2 Research Design and Evaluation

This research follows an experimental design in which system performance is evaluated under both controlled conditions and real-time operational scenarios. While Figure 1 illustrates the deployment architecture, the internal analytical workflow is described by the sequential processing pipeline shown in Figure 2.

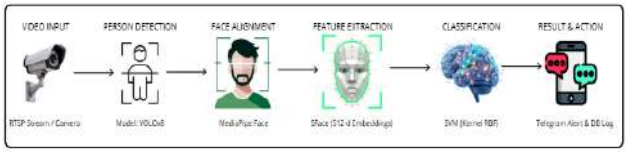

FIGURE 2. END-TO-END PROCESSING PIPELINE OF THE PROPOSED FRAMEWORK

Each incoming video frame undergoes the following stages:

1. Human Detection.

YOLOv8n is employed to detect objects belonging to the "person" class. YOLOv8n (nano) is selected for its fastest inference among YOLOv8 variants, making it suitable for CPU-only deployment environment [12], [13]. Only detected human regions are forwarded to subsequent stages, reducing unnecessary computation [2], [3].

2. Face Localization and Feature Extraction.

Faces within detected human bounding boxes are extracted and normalized using a pre-trained face detector [17]. A pre-trained SFace model generates fixed-dimensional embedding vectors representing identity features [4]. Embedding-based methods demonstrate robustness in surveillance scenarios [5].

3. Identity Classification.

Embedding vectors are classified using a Support Vector Machine (SVM) with a radial basis function kernel, suitable for closed-set recognition problems. SVM with RBF kernel is chosen for its low training overhead and fast inference on small embedding spaces [14]. The decision function of the SVM classifier is defined as:

$$f(x) = w^T x + b \tag{1}$$

4. Liveness Verification.

To mitigate presentation attacks, rule-based liveness detection is applied using Eye Aspect Ratio (EAR) for blink detection and Laplacian variance for texture-based spoof detection. A blink is detected when the EAR value falls below 0.2 for at least three consecutive frames. A face is classified as live if the Laplacian variance exceeds 50; otherwise, it is classified as a spoof. These thresholds were empirically determined through preliminary experiments to balance false acceptance and false rejection rates under varying lighting conditions.

Rule-based liveness detection is selected over deep learning-based alternatives to avoid GPU dependency and to maintain real-time performance under CPU-only deployment environment [10], [11]. The EAR is computed as defined in Formula (2).

$$EAR = \frac{||p_2 - p_6|| + ||p_3 - p_5||}{2||p_1 - p_4||} \tag{2}$$

A subject is authenticated only when classification confidence exceeds the defined threshold and liveness criteria are satisfied. Integrating anti-spoofing enhances biometric reliability [18]. This unified pipeline ensures that detection, recognition, and spoof prevention are validated as a single operational system rather than independent modules.

3.3 Dataset and Annotation Process

The dataset comprises 1,050 facial images collected from 10 predefined individuals within a campus environment using CCTV-comparable camera specifications, yielding an average of approximately 105 images per individual. Identity labeling was performed manually, and image resolution varies depending on the camera capture conditions. Figure 3.3 presents representative samples from the dataset, illustrating the variability in pose, illumination, and facial appearance.

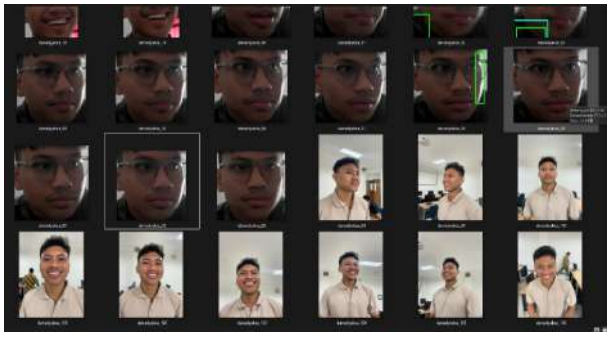

FIGURE 3. PRESENTS REPRESENTATIVE SAMPLES DATASET

The dataset was split into 80% for training (840 images) and 20% for testing (210 images). Pre-trained YOLOv8 and SFace models were used without fine-tuning, with only the SVM classifier trained using embedding vectors. This dataset size reflects a closed-set surveillance scenario commonly found in campus environments, where the number of registered individuals is limited and controlled.

In addition to static dataset evaluation, real-time IP-based CCTV streams were used to assess system stability and spoof detection performance. All experiments were conducted within a Linux-based Virtual Machine configured with Ubuntu Linux, 8 vCPU, 8 GB RAM, and no GPU acceleration, reflecting practical deployment conditions in educational institutions with limited hardware capacity.

3.4 Evaluation Metrics

System performance was evaluated across three dimensions:

a. Face Identification Metrics

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (6)$$

A confusion matrix was used for per-class performance analysis.

b. Liveness Detection Metrics

$$TPR = \frac{TP}{TP + FN} \quad (7)$$

$$FAR = \frac{FP}{FP + TN} \quad (8)$$

$$FRR = \frac{FN}{TP + FN} \quad (9)$$

Evaluation included genuine face presentations and spoof attempts using printed photographs and smartphone display replay under varying distances, angles, and lighting conditions.

c. Computational Performance Metrics

To evaluate feasibility under a CPU-only deployment environment, system performance was measured across four key indicators: frames per second (FPS), processing latency per frame, CPU utilization, and memory consumption. All evaluations were conducted under identical virtual machine configurations to ensure reproducibility and consistency across experiments.

4. RESULTS AND DISCUSSION

This section presents the empirical evaluation of the proposed integrated CCTV surveillance framework. The analysis focuses on three main aspects: (1) multi-class face recognition performance, (2) liveness detection robustness against presentation attacks, and (3) real-time computational feasibility under CPU-only deployment environment. All experiments were conducted under the configuration described in Section 3.

4.1 Face Recognition Performance

The face recognition module was evaluated as a closed-set multi-class classification problem involving ten registered identities. Each identity class contained 21 testing samples (support = 21), ensuring balanced class distribution across the 10 registered identities.

4.1.1 Classification Report

The system achieved an overall accuracy of 84.29%. Detailed per-class performance is presented in Table 1.

TABLE 1. MULTI-CLASS FACE RECOGNITION CLASSIFICATION

Class	Precision	Recall	F1-Score	Support
ariefikhwan	1.00	0.71	0.83	21
dedyhariyadi	0.75	0.86	0.80	21
oktavia	0.78	1.00	0.88	21
ramasahtyawan	0.88	1.00	0.93	21
trianahertiani	0.75	0.86	0.80	21
adityawahyuningrat	1.00	0.71	0.83	21
ahmadfazal	1.00	0.71	0.83	21
bayumurti	0.86	0.86	0.86	21
danadyaksa	0.67	0.86	0.75	21
rafiadinata	1.00	0.86	0.92	21
Overall Accuracy				**0.8429**

Precision values ranged from 0.67 to 1.00, while recall ranged from 0.71 to 1.00. Most F1-scores exceeded 0.80, indicating generally balanced classification performance across identities. This consistency between per-class F1-scores and overall accuracy suggests stable inter-class separability without dominance by a single identity class, reflecting the ability of the embedding space to preserve discriminative identity features.

High-performing classes such as ramasahtyawan (F1 = 0.93) and rafiadinata (F1 = 0.92) demonstrate strong feature separability in the SFace embedding space [8], indicating that the extracted embeddings maintain robust discriminative capability even without task-specific fine-tuning. In contrast, the lower precision observed in the danadyaksa class (0.67) indicates a higher proportion of false positive predictions, suggesting partial overlap between feature representations of this class and other identities. This behavior may be influenced by similar facial geometry between identities, illumination variation and limited intra-class diversity.

Despite the modest dataset size and CPU-only deployment environment, the results confirm that embedding-based recognition combined with SVM classification remains effective for small-scale closed-set authentication scenarios in resource constrained environments.

4.1.2 Confusion Matrix Analysis

Figure 4 presents the confusion matrix for multi-class face recognition, illustrating the classification performance across the ten registered identities.

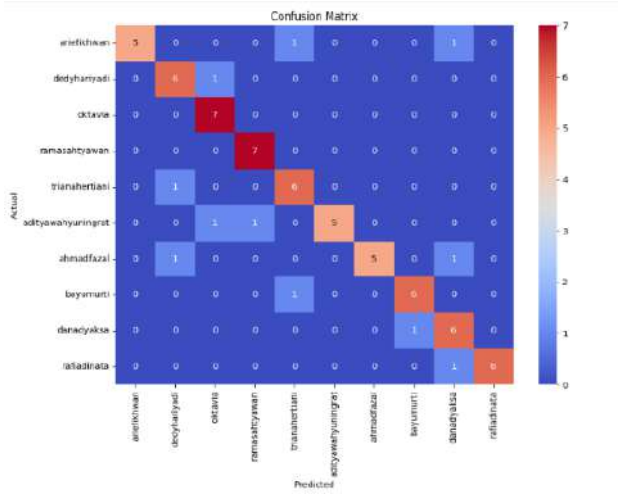

FIGURE 4. CONFUSION MATRIX FOR MULTI-CLASS FACE RECOGNITION

The confusion matrix shows that most predictions are concentrated along the principal diagonal, indicating correct class assignments. Off-diagonal elements represent misclassifications, primarily occurring between visually similar individuals. These misclassifications suggest that the extracted embedding features are not fully separable for certain identity pairs, leading to ambiguity in the SVM decision boundary. Such errors may be influenced by similar facial structures or accessories, illumination and pose variation and limited training diversity per class. Overall, the matrix confirms that the recognition module performs reliably in a controlled campus environment, although performance could further improve with larger and more diverse datasets.

4.2 Liveness Detection Performance

The liveness detection module was evaluated as a binary classification problem distinguishing genuine (live) faces from spoof presentations. Spoof samples included printed photographs and smartphone screen replay attacks.

4.2.1 Confusion Matrix

Figure 5 presents the confusion matrix for the liveness detection module, summarizing the classification results for genuine and spoof samples.

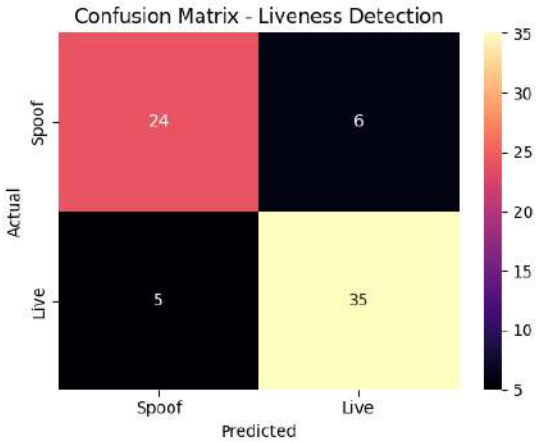

FIGURE 5. CONFUSION MATRIX FOR LIVENESS DETECTION

The matrix indicates that most live and spoof samples were correctly classified, confirming that the combined Eye Aspect Ratio (EAR) and Laplacian variance mechanism effectively differentiates genuine facial presence from static or replayed media.

4.2.2 Liveness Evaluation Metrics

The system achieved an overall accuracy of 84.29%, with a True Positive Rate (TPR) of 80.00% and a True Negative Rate (TNR) of 87.50%. The False Acceptance Rate (FAR) and False Rejection Rate (FRR) were recorded at 12.50% and 20.00%, respectively. These results indicate that the system demonstrates a reasonable balance between correctly identifying registered individuals and rejecting unauthorized access attempts, making it suitable for controlled closed-set surveillance scenarios such as campus environments.

The evaluation metrics were computed based on the following results:

$$\left[TPR = \frac{TP}{TP + FN} = \frac{24}{24 + 6} = 0.80\right] \quad (10)$$

$$\left[TNR = \frac{TN}{TN + FP} = \frac{35}{35 + 5} = 0.875\right] \quad (11)$$

$$\left[FAR = \frac{FP}{FP + TN} = \frac{5}{5 + 35} = 0.125\right] \quad (12)$$

$$\left[FRR = \frac{FN}{FN + TP} = \frac{6}{6 + 24} = 0.20\right] \quad (13)$$

The higher TNR compared to TPR indicates that the system is more effective at rejecting spoof attempts than consistently accepting all genuine users. This imbalance suggests that the system prioritizes security by minimizing false acceptance, albeit at the cost of rejecting some legitimate users. From a security standpoint, maintaining a relatively low FAR is desirable, as it reduces the risk of unauthorized access.

Although transformer-based anti-spoofing methods [12] may achieve higher robustness, they typically require substantially greater computational resources. In contrast, the proposed rule-based approach provides a practical balance between security and computational efficiency, making it well-suited for real-time deployment in CPU-only deployment environment.

4.3 Real-Time Computational Performance

The system was evaluated under live RTSP camera input processed entirely within a Linux-based Virtual Machine (8 vCPU, 8 GB RAM, no GPU).

4.3.1 FPS and Latency

After initialization, the system stabilized at an average FPS of 4.77 and an average latency per frame of 0.21 seconds. An initial latency spike (~0.95 seconds) was observed during model loading and memory allocation. After approximately 200 frames, the system reached steady-state performance with minimal variance. Figure 6 illustrates the convergence of frame rate over time, plotting FPS against the frame index, showing a stable throughput after the warm-up phase.

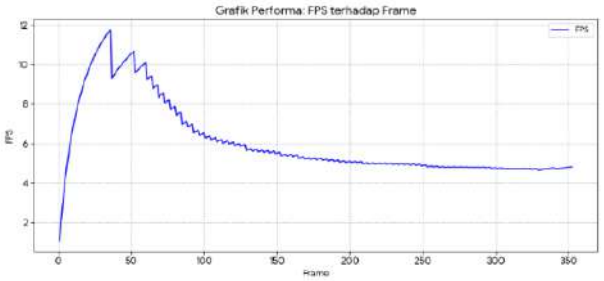

FIGURE 6. FPS VS FRAME INDEX

Figure 7 presents the latency per frame, demonstrating a significant decrease after initialization and stabilization at approximately 0.21 seconds per frame.

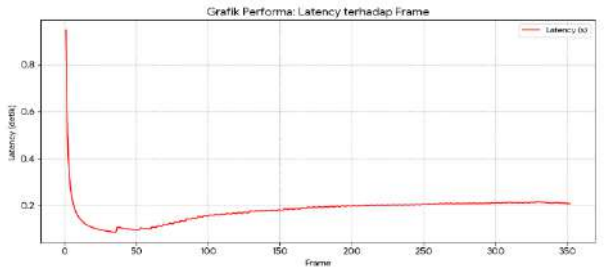

FIGURE 7. LATENCY PER FRAME

Although GPU-accelerated systems typically achieve higher throughput [2], the results confirm that integrated human detection, face recognition, and liveness verification can operate reliably within a CPU-only deployment environment.

4.3.2 Resource Utilization

Resource usage was monitored during continuous operation.

TABLE 2. REAL-TIME PERFORMANCE SUMMARY

Parameter	Average	Min.	Max.	Status
CPU Usage	16.89%	00.00%	69.40%	Controlled (8 spikes > 50%)
RAM Usage	78.20%	77.60%	78.60%	Highly stable (Deviation < 0.3%)

CPU usage remained moderate with occasional short spikes above 50%. RAM usage was highly stable without progressive growth, indicating no memory leakage. These results validate that the proposed framework can be deployed on institutional virtual servers without requiring GPU acceleration.

4.4 Operational System Performance and Validation

Beyond classification metrics, operational validation was conducted to assess real-world deployment feasibility. The results demonstrate that the proposed framework operates not only as a classification model but as a deployable surveillance system suitable for resource-constrained campus environments. These results further demonstrate that the system can support cyber intelligence capabilities through continuous monitoring, identity tracking, and real-time alert generation.

4.4.1 Continuous Operational Stability

The system was executed continuously for approximately 30 minutes using live RTSP input. Throughout this period, no runtime crash or unexpected termination was observed, and frame processing remained within the previously reported stable FPS range. Memory consumption showed no progressive growth during monitoring, and all processing stages including human detection, embedding extraction, identity classification, and liveness verification operated sequentially without any synchronization failure.

These results confirm the structural stability of the system under sustained CPU-only deployment conditions. The absence of degradation across all monitored indicators demonstrates that the proposed framework can maintain consistent performance over extended operation, reinforcing its feasibility for practical deployment in resource-constrained environments such as educational institutions.

4.4.2 Identity Logging and Database Validation

Each successful detection event generated structured metadata including, Timestamp, Identity label, Liveness status, and Camera source. These records were automatically stored in a PostgreSQL database. Validation confirmed consistency between detection events and stored entries, demonstrating reliable logging and audit functionality.

4.4.3 Real-Time Notification Delivery

The system's ability to deliver real-time alerts is illustrated in Figure 8, which shows the operational detection results for various scenarios.

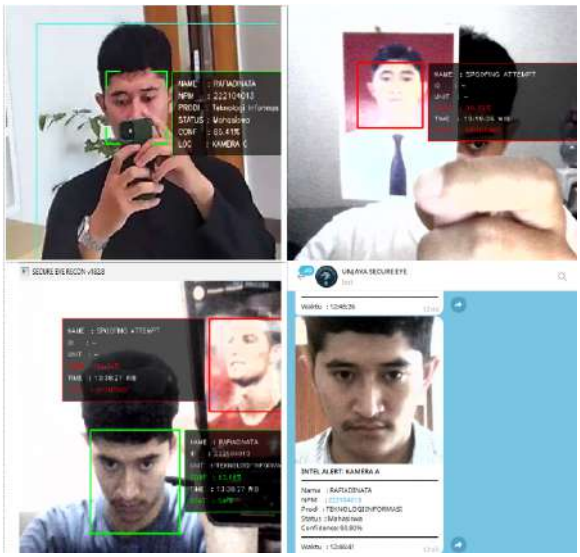

FIGURE 8. OPERATIONAL DETECTION RESULTS: (A) SUCCESSFUL LIVE IDENTITY RECOGNITION, (B) PRINT-BASED SPOOF DETECTION, (C) SCREEN REPLAY SPOOF DETECTION, AND (D) REAL-TIME TELEGRAM NOTIFICATION DELIVERY

The system was integrated with the Telegram Bot API to enable automated remote alerts triggered upon successful identity recognition and spoof detection. During testing, notifications were delivered consistently without transmission failure, no duplicate alerts were observed, and the average notification delay ranged between 1 and 2 seconds.

This responsiveness confirms near real-time remote monitoring capability, demonstrating that the system extends its functionality beyond local visualization. The reliable and low-latency alert mechanism reinforces the practical utility of the proposed framework in real-world

surveillance scenarios where immediate notification is critical for timely response.

4.4.4 End-to-End Functional Integration

The evaluation confirms the cohesive integration of all system components, where YOLOv8-based human detection filters candidate regions from incoming frames, SFace embedding extraction generates identity representations, SVM classification determines the predicted identity, and liveness verification validates authenticity. Metadata logging and notification modules execute automatically upon inference completion, ensuring end-to-end operational continuity.

Unlike studies that evaluate modules independently [8], [9], this work validates the complete processing chain under real operational conditions. The results demonstrate that the proposed framework operates not only as a classification model but as a fully deployable surveillance system, confirming its suitability for resource-constrained campus environments where both performance reliability and hardware efficiency are essential requirements.

4.5 Limitations and Future Work

Several limitations should be acknowledged. First, the system was evaluated using ten identities in a controlled indoor campus environment. Broader generalization requires more diverse datasets with varying illumination, occlusions, and demographic diversity, Robust face recognition under challenging surveillance conditions remains an open research problem, particularly under illumination variation, occlusion, and cross-domain deployment scenarios [19]. Second, the liveness detection mechanism relies on rule-based features and was evaluated only against print and screen replay attacks. More advanced spoof techniques such as 3D mask attacks were not considered. Third, scalability under multi-camera simultaneous processing has not been empirically validated, although the architecture supports RTSP-based integration. Future research will focus on expanding dataset diversity, incorporating lightweight deep learning-based anti-spoofing models, and evaluating multi-stream scalability in distributed operational environments.

4.6 Comparative Analysis

To further evaluate the effectiveness of the proposed framework, a comparative analysis is conducted against existing approaches in CCTV-based surveillance systems. Prior studies have demonstrated strong performance in individual components, particularly object detection using YOLO-based models [12], [13], [20], [21]. Face recognition has also been extensively studied using deep learning-based approaches and survey works [5], [18], [19], [22]. Meanwhile, face anti-spoofing and liveness detection have rapidly evolved, as highlighted in recent surveys [23], [24].

However, most existing works primarily focus on optimizing individual modules rather than evaluating a fully integrated end-to-end system operating under CPU-constrained deployment scenarios. This limitation is also discussed in surveillance-oriented studies [25]-[27], where real-time integration and computational efficiency remain key challenges.

To address this gap, this study emphasizes not only performance but also deployment feasibility. Table 4.3 presents a comparative evaluation of classifiers applied to SFace embeddings using the same dataset and experimental setup. All comparative experiments were conducted using the same dataset and CPU-only deployment environment to ensure fair and consistent evaluation.

TABLE 3. COMPARATIVE EVALUATION OF CLASSIFIERS

Classifier	Accuracy	Inference Time (ms)	Suitability
SVM (RBF)	84.29%	1.2	High
k-NN (k=5)	71.34%	3.4	Moderate
MLP	83.81%	1.5	Moderate

The results show that SVM achieves the highest accuracy (84.29%) with the fastest inference time. Although MLP provides comparable performance, it requires higher computational cost during training and inference. Meanwhile, k-NN shows lower efficiency due to distance calculations against all training samples. These findings indicate that SVM offers the best balance between accuracy and computational efficiency for CPU-only deployment environment. In addition to classifier comparison, liveness detection methods were also evaluated against alternative approaches. Table 4.4 summarizes the comparison between the proposed rule-based method and other lightweight techniques.

TABLE 4. COMPARATIVE OF LIVENESS DETECTION METHODS

Method	TPR	TNR	FAR	FRR	Latency (ms)
EAR + Laplacian	80%	87.50%	12.50%	20.00%	10
MobileNetV2	83%	85.00%	15.00%	16.70%	38
LBP Histogram	75%	80.00%	20.00%	25.00%	7

These results demonstrate that lightweight methods can provide practical security performance while maintaining low computational overhead in real-time surveillance scenarios. The proposed method achieves the highest True Negative Rate (87.50%) and lowest False Acceptance Rate (12.50%), indicating strong resistance to spoofing attacks.

Although MobileNetV2 achieves a lower False Rejection Rate, it introduces significantly higher latency and computational cost. The LBP-based method is faster but less accurate overall. Therefore, the rule-based approach provides the most suitable trade-off for real-time CPU-based surveillance systems.

Overall, the proposed framework prioritizes deployment feasibility, computational efficiency, and real-time performance over maximizing accuracy in isolated benchmark scenarios. This design choice aligns with edge computing principles, where resource constraints require lightweight yet reliable solutions [16].

4.7 Failure Case Analysis

Failure case analysis is conducted to further understand the limitations of the proposed system under real-world deployment conditions. Although the system demonstrates stable performance in controlled environments, several failure scenarios were identified during testing.

Case 1: Misclassification on visually similar faces. The system occasionally produces incorrect predictions when different individuals exhibit highly similar facial characteristics. This issue arises from the limited discriminative separation within the embedding space,

where feature representations between identities may overlap. This limitation is more pronounced due to the relatively small dataset used in this study.

Case 2: Performance degradation under poor lighting conditions. Recognition performance decreases significantly in environments with low or uneven illumination. Low-light conditions introduce noise and reduce contrast in facial regions, which degrades the quality and stability of SFace embeddings. As a result, the SVM decision boundary becomes less reliable, increasing the likelihood of misclassification. This limitation is particularly relevant in CCTV-based systems, where lighting conditions are often uncontrolled and dynamically changing.

Case 3: Liveness detection failure (False Rejection Rate -20%). The system records a False Rejection Rate (FRR) of 20%, indicating that genuine users may be incorrectly rejected. This occurs when natural variations in user behavior, such as delayed blinking or minimal facial movement, do not satisfy the predefined rule-based thresholds. This reflects a trade-off between system security and usability.

Case 4: Vulnerability to advanced spoofing scenarios. While the rule-based liveness detection mechanism is effective against basic attacks such as printed images and screen replay, it may struggle against more advanced spoofing techniques, including high-quality video replay with realistic motion patterns. In such scenarios, temporal facial cues such as eye blinking and subtle head movements can closely mimic genuine user behavior, allowing the attack to satisfy predefined EAR and texture thresholds. This limitation indicates that rule-based approaches alone may be insufficient against motion-based spoofing attacks, highlighting the need for more robust temporal or learning-based liveness verification methods.

In addition to these specific cases, broader system limitations were also observed. The system experiences increased computational load when processing crowded scenes with multiple faces, which may reduce real-time performance on CPU-only deployment environment. Furthermore, the dataset used in this study is relatively limited and was collected independently, which may restrict the generalization capability of the model in more diverse real-world environments.

Overall, these findings indicate that while the proposed system is effective for controlled and moderately complex scenarios, further improvements are required to enhance robustness, scalability, and generalization.

4.8 Trade-off Analysis

The proposed system involves several trade-offs between accuracy, computational efficiency, and security robustness. While the system achieves real-time performance under CPU-only constraints, it introduces a relatively high False Rejection Rate (20%), reflecting a clear trade-off between usability and security.

First, the use of YOLOv8n enables efficient real-time human detection, but it may exhibit lower detection accuracy compared to larger YOLO variants, particularly in complex scenes, as reported in recent studies [2], [3].

Second, the combination of SFace embeddings with SVM classification reduces computational overhead and enables fast inference, but it limits scalability compared to fully end-to-end deep learning models that can adapt more effectively to large-scale datasets [5].

Third, the rule-based liveness detection approach offers minimal computational cost and supports real-time execution on CPU-only infrastructure. However, it is inherently less robust than deep learning-based anti-spoofing methods when dealing with sophisticated attacks, such as high-quality video replay or motion-based spoofing [6], [7].

Finally, the system achieves stable real-time performance at 4.77 FPS under CPU-only deployment. Although this throughput is lower than GPU-accelerated systems, it demonstrates that integrated surveillance functionality combining detection, recognition, and liveness verification can be achieved without specialized hardware, aligning with edge-computing principles for resource-constrained environments [16].

Overall, these trade-offs indicate that the proposed framework prioritizes deployment feasibility, computational efficiency, and operational reliability over maximizing isolated performance metrics. This design choice is appropriate for real-world institutional environments, where hardware limitations necessitate lightweight yet effective solutions.

5 CONCLUSIONS

This research presents SECURE EYE, an integrated real-time CCTV surveillance framework combining YOLOv8n-based human detection, SFace embedding with SVM classification, and rule-based liveness verification, optimized for CPU-only deployment environment. The system was evaluated on 1,050 facial images from 10 identities and validated using live RTSP streams on a virtual machine (8 vCPU, 8 GB RAM, no GPU). The proposed framework achieved 84.29% face recognition accuracy (F1-score: 0.75–0.93) and liveness performance of TPR 80.00% and TNR 87.50% (FAR 12.50%, FRR 20.00%). Real-time operation was maintained at 4.77 FPS with 0.21 s latency and 16.89% CPU utilization, confirming stable performance under CPU-only deployment environment. These results demonstrate that the system can operate as a practical surveillance solution while supporting cyber intelligence capabilities.

However, several limitations remain. The FRR of 20% indicates a usability constraint due to rigid threshold-based liveness rules. The system relies on a limited dataset (10 identities) and was evaluated under controlled conditions, which may restrict generalization. In addition, the rule-based liveness approach is less robust against advanced spoofing attacks, and multi-camera scalability has not yet been validated. Future work should focus on integrating lightweight deep learning-based anti-spoofing, expanding dataset diversity, and applying adaptive or temporal liveness strategies to reduce FRR while maintaining low FAR. Evaluation under multi-stream deployment and the integration of behavioral and anomaly-based analytics are also recommended to further enhance cyber intelligence capabilities.

Overall, this study demonstrates that an end-to-end surveillance system can be effectively deployed in resource-constrained environments, providing a practical foundation for CPU-efficient and cyber intelligence

enabled CCTV systems. This work contributes not only to intelligent surveillance systems but also to the advancement of cyber intelligence frameworks in resource-constrained environments.

REFERENCES

[1] M. Andrejevic and N. Selwyn, "Facial recognition technology in schools: critical questions and concerns," *Learn. Media Technol.*, vol. 45, no. 2, pp. 115–128, 2020, doi: 10.1080/17439884.2020.1686014.

[2] R. Khanam and M. Hussain, "What is YOLOv5: A deep look into the internal features of the popular object detector," pp. 3–10, 2024, [Online]. Available: http://arxiv.org/abs/2407.20892

[3] R. Sapkota *et al.*, "YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series," *Artif. Intell. Rev.*, vol. 58, no. 9, 2025, doi: 10.1007/s10462-025-11253-3.

[4] Y. Zhong, W. Deng, J. Hu, D. Zhao, X. Li, and D. Wen, "SFace: Sigmoid-Constrained Hypersphere Loss for Robust Face Recognition," *IEEE Trans. Image Process.*, vol. 30, pp. 2587–2598, 2021, doi: 10.1109/TIP.2020.3048632.

[5] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, and D. Zhang, "Biometrics recognition using deep learning: a survey," *Artif. Intell. Rev.*, vol. 56, no. 8, pp. 8647–8695, 2023, doi: 10.1007/s10462-022-10237-x.

[6] A. Keresh and P. Shamoi, "Liveness Detection in Computer Vision: Transformer-Based Self-Supervised Learning for Face Anti-Spoofing," *IEEE Access*, vol. 12, pp. 185673–185685, 2024, doi: 10.1109/ACCESS.2024.3513795.

[7] H. Qi, R. Han, Y. Shi, and X. Qi, "A Novel High-Performance Face Anti-Spoofing Detection Method," *IEEE Access*, vol. 12, no. April, pp. 67379–67391, 2024, doi: 10.1109/ACCESS.2024.3400285.

[8] R. Ullah *et al.*, "A Real-Time Framework for Human Face Detection and Recognition in CCTV Images," *Math. Probl. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/3276704.

[9] H. J. Mun and M. H. Lee, "Design for Visitor Authentication Based on Face Recognition Technology Using CCTV," *IEEE Access*, vol. 10, no. October, pp. 124604–124618, 2022, doi: 10.1109/ACCESS.2022.3223374.

[10] S. Chakraborty and D. Das, "An Overview of Face Liveness Detection," *Int. J. Inf. Theory*, vol. 3, no. 2, pp. 11–25, 2014, doi: 10.5121/ijit.2014.3202.

[11] V. Rahayu, E. Dolphina, and R. Pramunendar, "Implementasi Sistem Absensi Berbasis Face Recognition Dengan Mekanisme Blink Detection Menggunakan Svm Secara Real-Time," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 11, no. 1, pp. 755–769, 2026, doi: 10.36341/rabit.v11i1.7127.

[12] M. R. Sholahuddin *et al.*, "Optimizing YOLOv8 for Real-Time CCTV Surveillance: A Trade-off Between Speed and Accuracy," *J. Online Inform.*, vol. 8, no. 2, pp. 261–270, 2023, doi: 10.15575/join.v8i2.1196.

[13] Q. Wang, G. Feng, and Z. Li, "A Lightweight Person Detector for Surveillance Footage Based on YOLOv8n," *Sensors*, vol. 25, no. 2, 2025, doi: 10.3390/s25020436.

[14] Z. Ramadhani, L. Safira, A. D. Hartanto, and H. Hartatik, "Implementasi Algoritma SVM Dalam Pengembangan Sistem Presensi Berbasis Face Recognition," *Intechno J. (Information Technol. Journal)*, vol. 4, no. 2, pp. 42–47, 2022, doi: 10.24076/intechnojournal.2022v4i2.1561.

[15] D. Wang *et al.*, "Wild Face Anti-Spoofing Challenge 2023: Benchmark and Results," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2023-June, pp. 6380–6391, 2023, doi: 10.1109/CVPRW59228.2023.00679.

[16] 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, 2016, doi: 10.1109/JIOT.2016.2579198.

[17] V. Sanikommu, S. Mummaneni, N. Jacob, K. C. Emmanuel Sanjay Raj, B. Kumar, and R. D. P. Variam, "Deep Learning-Based Control System for Context-Aware Surveillance Using Skeleton Sequences from IP and Drone Camera video," *Int. Arab J. Inf. Technol.*, vol. 22, no. 5, pp. 919–929, 2025, doi: 10.34028/iajit/22/5/6.

[18] H. L. Gururaj, B. C. Soundarya, S. Priya, J. Shreyas, and F. Flammini, "A Comprehensive Review of Face Recognition Techniques, Trends, and Challenges," *IEEE Access*, vol. 12, no. June, pp. 107903–107926, 2024, doi: 10.1109/ACCESS.2024.3424933.

[19] A. Zhalgas, B. Amirgaliyev, and A. Sovet, "Robust Face Recognition Under Challenging Conditions: A Comprehensive Review of Deep Learning Methods and Challenges," *Appl. Sci.*, vol. 15, no. 17, pp. 1–27, 2025, doi: 10.3390/app15179390.

[20] M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," *Machines*, vol. 11, no. 7, 2023, doi: 10.3390/machines11070677.

[21] D. Nimma *et al.*, "Object detection in real-time video surveillance using attention based transformer-YOLOv8 model," *Alexandria Eng. J.*, vol. 118, no. November 2024, pp. 482–495, 2025, doi: 10.1016/j.aej.2025.01.032.

[22] L. Li, X. Mu, S. Li, and H. Peng, "A Review of Face Recognition Technology," *IEEE Access*, vol. 8, pp. 139110–139120, 2020, doi: 10.1109/ACCESS.2020.3011028.

[23] Z. Yu, Y. Qin, X. Li, C. Zhao, Z. Lei, and G. Zhao, "Deep Learning for Face Anti-Spoofing: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 5609–5631, 2023, doi: 10.1109/TPAMI.2022.3215850.

[24] H. Xing, S. Y. Tan, F. Qamar, and Y. Jiao, "Face Anti-Spoofing Based on Deep Learning: A Comprehensive Survey," *Appl. Sci.*, vol. 15, no. 12, pp. 1–42, 2025, doi: 10.3390/app15126891.

[25] N. Choudhry, J. Abawajy, S. Huda, and I. Rao, "A Comprehensive Survey of Machine Learning Methods for Surveillance Videos Anomaly Detection," *IEEE Access*, vol. 11, no. October, pp. 114680–114713, 2023, doi: 10.1109/ACCESS.2023.3321800.

[26] C. Rajesh, B. R. T. Bapu, S. Asha, and R. K. Veluri, "Deep learning-based anomaly detection in video surveillance," *Int. J. Electron. Secur. Digit. Forensics*, vol. 17, no. 4, pp. 522–534, 2025, doi: 10.1504/IJESDF.2025.147181.

[27] F. Hidayat, U. Elviani, and F. Agil Alunjati, "Perspective Chapter: Advancement in CCTV Video Analytics - Leveraging Face Recognition for Enhanced Security Solutions," *Image Sensors - Digit. Imaging Syst. Appl.*, pp. 1–16, 2025, doi: 10.5772/intechopen.1007978.

AUTHORS

Rafi Adinata Rachmat

Rafi Adinata Rachmat is currently an undergraduate student in Information Technology at Universitas Jenderal Achmad Yani Yogyakarta. With a strong focus on Cyber Security, he possesses expertise in Ethical Hacking, Digital Forensics, and OSINT Analysis. Beyond technical security, he is proficient in networking and programming. His multidisciplinary background includes professional skills in photography and videography, enhancing his capabilities in visual data analysis and digital content processing. He has been actively involved in cybersecurity research and real-time AI surveillance development, reflecting his commitment to advancing digital safety and intelligence.

Dedy Hariyadi

Dedy Hariyadi received a bachelor's degree in informatics engineering (S.T.) from Universitas Pembangunan Nasional" Veteran" Yogyakarta, in 2004, a master's in digital forensics (M.Kom.) from Universitas Islam Indonesia in 2016, and engineer profession (Ir.) from Universitas Muhammadiyah Yogyakarta, in 2024. Currently continuing his doctoral studies in Cyber Security (Dr.) at the Department of Computer Science and Electronics, Universitas Gadjah Mada. He is currently a lecturer and researcher in Cyber Security and Digital Forensics at Universitas Jenderal Achmad Yani Yogyakarta. He can be contacted by email: dedy@unjaya.ac.id.