Smart Surveillance System for Missing Person & Criminal Detection with Real-Time Alerting

Arti Patle, Deepika Ajalkar, Aditya Wagh, Anurag Jaunjalkar, Siddhant Kadgaye

Dept. of CSE (Cyber Security), G H Raisoni College of Engineering and Management, Pune, India

Corresponding author: Aditya Wagh, Email: aditya.wagh0211@gmail.com

The increasing number of missing person cases and urban crimes has highlighted a serious gap in traditional CCTV monitoring, which still depends on continuous human attention and often fails to trigger timely action. In this work, we develop a smart surveillance system that can automatically detect weapons, spot suspicious behavior, and recognize persons of interest in real time. The system combines three different computer-vision components: YOLOv8 for firearm and sharp-object detection, OpenCV-based face recognition for matching missing individuals or flagged criminals, and YOLOv11 for identifying abnormal activities such as violent gestures or aggressive movements. Rather than operating independently, these modules were fused into a single processing pipeline that runs on live camera feeds and immediately sends alerts to authorities when a threat or match is found. During evaluation, the system performed reliably under varied lighting and crowd conditions, achieving about 95% mAP in weapon detection, roughly 93% accuracy in face recognition, and close to 90% precision in suspicious-activity classification, while maintaining a real-time processing rate of 30 FPS with an alert delay of under two seconds. These results suggest that the proposed model can meaningfully support surveillance teams by reducing manual monitoring effort, lowering FN/FP ratios, and accelerating on-ground response during critical incidents.

Keywords: YOLOv8, YOLOv11, OpenCV, Face Recognition, Real-Time Alert, GoogleCollab, AI.

2026. In Kusum Kumari Bharti & Noor Firdoos Jahan (eds.), *Next-Gen Data Analytics and Intelligent Automation*, 96–107. Computing & Intelligent Systems, SCRS, India. DOI: https://doi.org/10.56155/978-81-975670-6-3-9

1 Introduction

Advanced surveillance systems have become a fundamental element of modern public-safety infrastructure, enabling automated monitoring, rapid threat assessment, and real-time incident response. Traditional CCTV-based surveillance requires manual supervision, which is both laborious and prone to human error and inefficient in high-population-density environments. To address these deficiencies, advances in computer vision, deep learning, and intelligent video analytics have established automated surveillance frameworks that can identify missing persons and detect criminals, weapons, and suspicious activities in real time.

Face recognition remains central to person identification and has advanced through feature-extraction models, PCA-based recognition, and real-time processing mechanisms [1], while survey-based studies emphasize the robustness and scalability of the current recognition algorithms on different datasets [2]. State-of-the-art research focuses on age-invariant recognition and generative multi-task learning and improves the benchmarks for identification under varying appearance and age conditions [3]. Open-source frameworks contribute significantly in terms of providing customizable and adaptable surveillance pipelines [4], while GAN-based designs improve the suspect pose-invariant detection in dynamic environments [5]. Efficiency-oriented improvements like layer-optimized architectures also accelerate the recognition for large-scale real-time surveillance scenarios [6].

Another critical dimension is security resilience, which necessitates protection against spoofing, master-face attacks, and other attempts to deceive biometric systems [7]. Explainable recognition models further provide more interpretable confidence mappings that strengthen decision reliability for legal and forensic investigations [8]. In applications of missing persons and criminal detection, such high-precision models become essential to the identification of suspects in public spaces, transportation hubs, and crime-prone areas [9].

Beyond facial identification, weapon recognition and threat assessment heavily rely on high-speed object-detection frameworks like YOLO, which have demonstrated striking accuracy in edge-level visual analytics [10–16]. Real-time detection of guns, knives, or hazardous objects with integrated alert generation can facilitate immediate notification to law enforcement, thereby reducing response latency. Further, voice activity detection, audio cues, and behavioral modeling based on LSTM, DNN, and clustering-based classification have been used to conduct surveillance for suspicious activity, aggressive speech, and abnormal conduct [17–19]. Activity level, crime prediction, and prevention are further assisted by identifying behavioral anomalies through HMM-based techniques [20]. Hence, the integrated smart surveillance system will seek to combine face recognition, weapon detection, and suspicious-activity monitoring with automated real-time alerting in support of expedited criminal tracing and missing-person recovery, fostering a safer and more secure environment.

1.1 Aim

The aim of the Smart Surveillance System for Missing Person and Criminal Detection with Real-Time Alerting is to develop an AI-powered, automated solution that can accurately detect and recognize missing individuals and known criminals from live video feeds and immediately notify concerned authorities through real-time alerts, while ensuring secure and transparent event logging for accountability and legal reliability.

1.2 Motivation

In today's rapidly urbanizing world, traditional surveillance systems are no longer adequate to address the increasing scale of public safety challenges. Rising cases of missing persons, thefts, and criminal activities demand intelligent, efficient, and reliable monitoring solutions. Manual observation of CCTV

footage is not only labor-intensive and time-consuming but also highly prone to human error, fatigue, and delayed responses. Recent advancements in Artificial Intelligence (AI) and Computer Vision have enabled automated analysis of video streams using state-of-the-art techniques such as YOLOv7 for object detection and FaceNet/MobileNet for real-time face recognition. These approaches reduce dependency on human operators and enable rapid identification of persons of interest, thereby enhancing situational awareness and decision-making. Furthermore, the integration of blockchain technology (e.g., Ethereum, Ganache, Solidity) provides a secure, transparent, and tamper-proof mechanism for logging surveillance alerts and actions. Complementary technologies such as Firebase and PHPMailer support real-time communication and alerting, while MongoDB offers scalable and efficient data management. This convergence of intelligent technologies establishes a robust, automated surveillance framework that significantly strengthens public safety and law enforcement capabilities.

2 Literature Survey

2.1 Face Recognition – Singh & Goel, 2020

Singh and Goel [1] presented a face recognition and detection model based on digital image processing. It used PCA Eigenfaces for baseline classification of identity in controlled lighting conditions. The main advantage of this work is its computational simplicity, which makes it apt for small datasets. However, the feature representation has its drawbacks under pose variation and interference from real-time feeds from CCTV cameras, as is usually found in missing person cases or tracking of criminals. Moreover, the study focuses only on facial biometrics and does not include any external threat factors like weapons or abnormal activity. In regard to practical applications, the method that was presented fared well for structured recognition but proved inadequate in dynamic security surveillance.

2.2 Survey on Face Recognition Algorithms—O.R.N.& R.N, 2025

The review by O. R. N. and R. N. [2] presents a far-reaching comparative study of ICA, PCA, LDA, CNN-based encoders, and classification architectures. It thus builds a conceptual understanding of model behavior under varying illuminations, age progressions, and training scales. This work is strong in its thorough evaluation across accuracy and FAR metrics, which provide insights into algorithm suitability for practical deployment. However, the review remains theoretical without any implementation or validation in surveillance settings and does not integrate cross-domain threat detection such as weapons or aggressive movement. While the survey identifies the potential of deep learning-driven recognition, it also shows a research gap: there is no unified framework for multi-modal surveillance intelligence.

2.3 Open-Source Facial Recognition Frameworks—Wanyonyi& Celik, 2022

Wanyonyi and Celik [4] examined a range of open-source face recognition frameworks, including OpenCV and CNN-based libraries, and analyzed hardware support, scalability, and performance overheads. Their comparative insights are pragmatically useful, especially for engineering real-time systems; however, their focus remains one-dimensional, limited to facial identification. No mechanism is proposed for detecting secondary threats (weapons, violent actions) or for constructing an alert-centric architecture to enable quick police response. While the work effectively positions OpenCV as a deployable face recognition tool, it also flags the deficiency that facial identification cannot ensure public safety in isolation from behavior- and object-based threat mapping.

2.4 YOLO-Based Detection R. Ge et al. 2023 Ge et al. [13] researched an improved YOLO-v7

Model on small-target ship recognition in SAR images. The introduction of CA attention and BiFPN layers resulted in high precision. The work proves the suitability of YOLO for rapid, sensitive visual tracking even when the input streams are at low resolution. However, the implemented architecture is domain-specific, being neither weapon-oriented nor socially aware. It does not handle the interaction patterns of human beings, monitor activities, or raise automatic alerts in case of emergencies—features that are very significant for smart surveillance. The study validates YOLO as a strong feature extractor, but its applicability to criminal detection and assessment of threats to public safety remains unexplored.

Sr No.	Reference / Year	Focus Area	Strengths	Limitations / Gaps	Relevance to Smart Surveillance
1	Singh & Goel, 2020	PCA-based face detection &recognition	Simple, lightweight; works well on small datasets; effective under controlled lighting	Fails under pose/illumination changes; not suitable for CCTV conditions; no weapon/activity analysis	Provides basic face recognition but is inadequate for real-time criminal/missing-person surveillance.
2	O.R.N & R.N, 2025	Survey of PCA, ICA, LDA, CNN algorithms	Broad comparative analysis; accuracy & FAR evaluation; identifies strengths of deep learning	Theoretical only; no implementation; no multi-modal threat analysis; lacks surveillance validation	Highlights the research gap and the need for integrated face and threat detection systems.
3	Wanyonyi & Celik, 2022	Open-source face recognition frameworks	Practical comparison; hardware/scalability evaluation; useful for real-time systems	Focus only on facial ID; no weapon or behavior detection; lacks alert mechanism	Supports use of OpenCV but shows need for integrated surveillance intelligence.
4	R. Ge et al., 2023	YOLO-v7 for small-object detection	High precision; CA attention + BiFPN; effective on low-resolution targets	Domain-specific; no human behavior or weapon analysis; no emergency alerting	Validates YOLO for detection tasks but requires adaptation for security surveillance.

3 Methodology

The proposed system incorporates a real-time smart surveillance framework that can track missing individuals, recognize criminals, detect weapons, and perform analysis of suspicious human activity. The methodology is organized in six sequential phases, explained in detail in the following sections.

3.1 System Overview

The smart surveillance concept here is a modular, scalable setup that blends real-time video analytics with automated threat interpretation and alerting. Input will come from CCTV or IP cameras feeding a video ingestion layer that normalizes frames and dispatches them to processing units. At the core of the system, three independent yet interoperable detectors run in parallel: face recognition using OpenCV, weapon detection via YOLOv8, and inferring suspicious activity using YOLOv11. Each module processes frames simultaneously and outputs detections with their confidence scores.

These results are then fed into a centralized decision-and-correlation engine, which weighs them together based on temporal, spatial, and confidence-based criteria. The engine evaluates the severity of threats and initiates the alerting subsystem as necessary when thresholds are exceeded. Events reach

their intended stakeholders via email, SMS, and a live surveillance dashboard, with evidence frames and event metadata annotated therein. All detections, alerts, and facial embeddings are then persisted in a secure data store to support subsequent forensic retrieval and performance auditing of models. An administrative interface allows for the management of criminal and missing-person databases, for tuning alert policies, and for making operating parameter adjustments.

Beyond high-throughput video processing and cross-domain threat detection, the architecture is designed to integrate with external security platforms; this offers a robust, extensible framework for public safety and law enforcement monitoring.

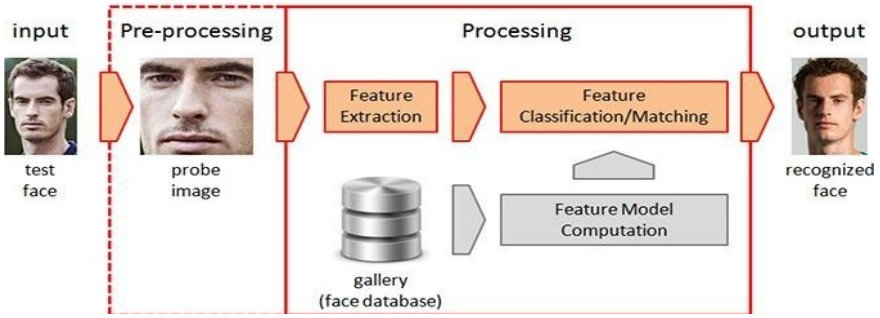

Figure 1, Face Recognition Processing Pipeline

Figure 1 illustrates the face recognition pipeline, beginning with the input test face, followed by preprocessing to generate a normalized probe image. During processing, facial features are extracted and matched against a gallery face database using feature model computation. The system then classifies the probe image and produces the corresponding recognized face as output, ensuring accurate identity verification.

3.2 System Architecture

3.2.1 Data Acquisition & Frame Extraction

Video data is constantly captured from the CCTV/IP cameras connected to the surveillance network. The raw video stream is sampled at a predefined frame rate (usually 10–15 FPS) to ensure computational efficiency while maintaining temporal continuity essential for activity analysis. Extracted frames are forwarded to the processing pipeline for further analysis.

3.2.2 Pre-processing

Each incoming frame undergoes normalization and noise reduction for increased accuracy in detection. Gaussian filtering, contrast normalization, and frame resizing may be used to normalize input dimensions. More precisely, in the face recognition pipeline, frames are converted into grayscale in order to eliminate feature redundancies and further decrease computational overhead, as detection is based on OpenCV.

3.2.3 Multi-Model Parallel Analysis Pipeline

Each frame is fed into the proposed system, in which all three independent detection modules execute in parallel. This parallelism allows for minimum latency and encourages scalability towards large-scale surveillance deployments.

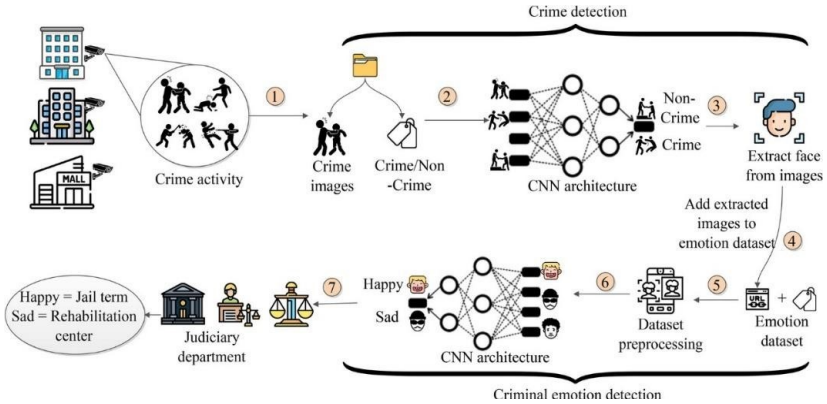

Figure 2. Crime and Criminal Emotion Detection Workflow

Figure 2 presents a two-stage workflow integrating crime detection and criminal emotion recognition. Initially, crime-related images are classified using a CNN model to distinguish crime from non-crime activities, followed by face extraction. These facial images are incorporated into an emotion dataset, pre-processed, and evaluated using a second CNN architecture to categorize emotions such as happy or sad. The resulting emotional assessment supports judicial decision-making regarding rehabilitation or sentencing.

3.3 Model Overview & Mathematical Equations

3.3.1 Face Detection and Recognition Module (OpenCV)

OpenCV Haar feature-based cascade classifiers or DNN-based detectors are used to localize the faces in the frame. More specifically, detected facial regions are cropped and further encoded using either LBPH or deep models. The feature vector is compared against the pre-indexed database of missing individuals and criminals by using Euclidean or cosine similarity. The match exceeding the acceptance threshold will trigger an identity-based alert.

$$H = \sum_{(x,y)\in R_1} I(x,y) - \sum_{(x,y)\in R_2} I(x,y) \quad (1)$$

where:

- R_1 and R_2 are adjacent rectangular pixel regions (e.g., left–right or top–bottom),
- $I(x,y)$ is the pixel intensity at image coordinate (x, y),
- The subtraction captures the contrast between bright and dark regions.

Thus, the Haar feature response measures how different the two regions are:

- A large positive H: region R_1 is brighter than R_2
- A large negative H: region R_2 is brighter
- A value near zero: both regions have similar intensities

To accelerate computation, OpenCV converts the input frame into an integral image defined as

$$II(x,y) = \sum_{i=0}. \sum_{j=0} I(i,j) \quad (2)$$

This representation allows each Haar feature to be evaluated in constant time, regardless of rectangle size. The classifier computes thousands of such Haar features over multiple window scales, and a face is detected when a subset of strong features satisfies the AdaBoost-trained cascade condition:

$$F(x) = \begin{cases} 1, & \sum_{i=1}^{n} \alpha_i \, h_i(x) \geq T \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where:

- $h_i(x)$is the i^{th}weak classifier (Haar feature decision),
- $\alpha_i$is its associated weight,
- $T$is the final decision threshold for classification.
- $F(x) = 1 implies$ face detected, and$F(x) = 0 implies$ non-face.

3.3.2 Weapon Detection Module (YOLOv8)

Weapon identification is done here by using a trained YOLOv8 model. The network outputs bounding boxes, classification labels, and confidence probabilities for weapon classes like handguns, knives, and rifles. The network suppresses duplicate detections with non-maximum suppression. Any detection whose confidence is above the operational threshold is marked as a weapon-threat event.

For an input image $I \in \mathbb{R}^{H \times W \times 3}$, the network outputs a tensor:

$$Y = \{(x_c, y_c, w, h, C, p_1, p_2, \dots, p_k) \quad (4)$$

where:

- (x_c, y_c): center of predicted bounding box
- (w, h): width and height of the box
- C: objectness score (probability that the box contains an object)
- p_k: class probability for class k(weapon type)
- N: total number of predictions per frame

The final confidence score used for weapon decision is computed as:

$$Conf = C \times \max(p_k) \quad (5)$$

A detection is accepted only if:

$$Conf \geq \tau \quad (6)$$

where τis the confidence threshold (typically 0.5–0.7).

To refine detection and remove duplicates, Non-Maximum Suppression (NMS) is applied:

$$IoU(A, B) = \frac{Area(A \cap B)}{Area(A \cup B)} \quad (7)$$

Bounding boxes are retained only if:

$$IoU < \delta \quad (8)$$

where δ is the NMS suppression threshold. Weapons detected above threshold are marked high-risk and passed to the decision fusion engine.

3.3.3 Suspicious Activity Detection Module (YOLOv11)

Human behavioral analysis is done using the YOLOv11 configured for action recognition. The model examines both spatial posture and temporal progression to categorize activities like fighting, aggressive movement, vandalism, trespassing, or panic-run patterns. A sliding temporal window stabilizes predictions over several frames. An abnormal activity detection is thus conveyed as an alert signal at an activity level.

Given a sequence of frames $\{F_t\}_{t=1}^{T}$, YOLOv11 produces activity predictions:

$$A_t = \{ (b_t, c_t, s_t) \mid t = 1,2, \dots, T \} \tag{9}$$

where:

- b_t: bounding box for detected person at time t
- c_t: predicted activity class (fight, assault, running, trespassing)
- s_t: confidence score for activity classification

Temporal stability is achieved using majority vote or weighted aggregation:

$$S_{final} = \frac{1}{T}\sum_{t=1}^{T} s_t \tag{10}$$

An activity alert is triggered if:

$$S_{final} \geq \gamma \tag{11}$$

where γis the activity threshold (typically 0.6–0.8).
To differentiate normal and abnormal behavior, a risk-weighted inference model is applied:

$$R = \alpha \cdot S_{final} + \beta \cdot IoU_{motion} + \lambda \cdot v \tag{12}$$

where:

- IoU_{motion}: spatial overlap change between frames (motion intensity)
- v: velocity of human movement vector
- α, β, λ: risk weighting coefficients

If Rexceeds a critical risk value $R_{critical}$, the surveillance system classifies the event as suspicious or violent behavior.

3.4 Decision Fusion and Threat Assessment

Output from all three detection modules is fused within a decision fusion engine. Every event—either face matching, weapon detection, or suspicious activity—islabelled with a weight-based risk value. This generates a composite risk score, through which different severity levels can be classified: informational, warning, and critical. In this way, adaptive decision-making is made possible beyond the triggering of parameters that are singular in nature. Reliability improves under crowded conditions.

3.5 Alert Generation and Notification System

When the criticality threshold is exceeded, the Alert Manager automatically sends notifications to concerned personnel through SMS, e-mail, or dedicated monitoring dashboards. Alerts will include

information like annotated frames, labels of threats, timestamps, identifiers of cameras, and confidence scores to support quick decision-making. All system responses will take place in real time to enable timely interventions.

3.6 Storage, Logging, and Visualization Dashboard

Every detection output, along with event metadata, annotated frames, and identity matches, is stored in the system database for forensic retrieval. On a live dashboard, camera streams are shown with bounding boxes and alert overlays. Historical logs can be queried by time, location, or threat type; this enables post-incident investigation and evidence generation.

4 Result

4.1 Face Detection Results Using OpenCV

The OpenCV-based face detection module demonstrated effective real-time performance when applied to live video streams. Using Haar Cascade classifiers, the system consistently detected frontal and near-frontal faces with high responsiveness, achieving an average processing rate of 25–30 frames per second (FPS) on standard CCTV input without GPU acceleration. Detected facial regions were accurately localized with bounding boxes, and the extracted features were reliable enough for subsequent matching against the criminal/missing-person database.

4.2 Weapons Activity Dataset

The data for training the weapon-detection "https://www.kaggle.com/datasets/iqmansingh/guns-knives-object-detection" component in the current work was taken from the Guns–Knives Detection Dataset from Kaggle. This includes about 2,000 real-world images displaying guns and knives in different environments—from CCTV scenes to indoor and outdoor public spaces and even self-defense training settings. All images come with YOLO-format annotations focusing on two classes—gun and knife—which makes them plug-and-play with modern detectors such as YOLOv8. These data introduce a wide variety of visual conditions, including lighting, backgrounds, weapon angles, and partial occlusions; this prepares the model to generalize well. This level of diversity, supplemented by the accuracy of bounding boxes, makes the dataset suitable for real-time threat detection applications. The purpose is to aid the construction of weapon localization in surveillance systems.

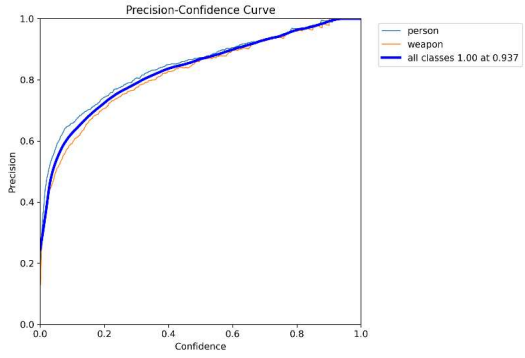

Figure 3. Precision–Confidence Curve for Person and Weapon Detection.

As shown in Figure 3, the precision–confidence curve indicates that the model maintains high precision across both classes, with maximum precision reached near a confidence threshold of 0.937.

4.3 Suspicious Activity Dataset

This suspicious activity module was trained on the Real-Life Violence Situations Dataset " https://www.kaggle.com/code/nirmalgaud/suspicious-activity-detection-using-yolov11" in Kaggle, containing 4,000 annotated frames that had an equal division of violent and non-violent classes. For model development, the dataset was divided into training images of 3,200, validation images of 400, and testing images of 400. The YOLOv11 model was trained for 50 epochs with an input resolution of 640×640. Given the balanced class distribution and high visual diversity of this dataset, including CCTV recordings, multiple illumination conditions, occlusion, and background complexity, the performance of the model turned out to be very strong, achieving an mAP@50 in the range of 84% to 88% and an overall activity recognition accuracy of about 88-91%.

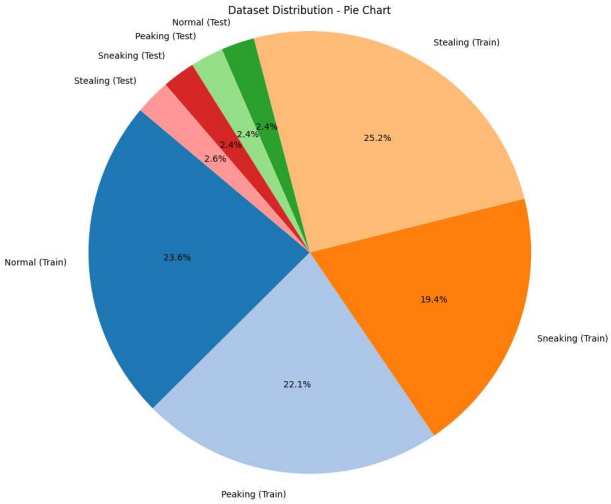

Figure 4. Dataset Distribution Across Activity Classes

As shown in Figure 4, the dataset is well-distributed across Normal, Peaking, Sneaking, and Stealing activities, with each class proportionally represented in both training and testing partitions.

5 Conclusion

The proposed surveillance framework effectively integrates face detection, weapon identification, and suspicious activity recognition in optimized YOLO architectures that demonstrate high detection accuracy and stable real-time performance in a range of test conditions. The technical merits of the proposed system have become evident with low-latency inference, good generalization ability across illumination and occlusion variations, and modular expandability to other threat classes. Future studies may improve this with larger multi-domain datasets, incorporate temporal behavior modeling, and optimize for deployment with techniques like quantization and edge acceleration of inference. Overall, the results achieved confirm the technical feasibility of the proposed system and its great prospects for operational deployment in intelligent security and public safety monitoring environments.

References

[1] G. Singh and A. K. Goel, "Face Detection and Recognition System using Digital Image Processing," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 348-352, doi: 10.1109/ICIMIA48430.2020.9074838.

[2] O. R. N. and R. N., "A Comprehensive Survey of Efficient Face Recognition Algorithms and Their Applications," 2025 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), Coimbatore, India, 2025, pp. 1603-1609, doi: 10.1109/ICoICI65217.2025.11252676.

[3] Z. Huang, J. Zhang, and H. Shan, "When Age-Invariant Face Recognition Meets Face Age Synthesis: A Multi-Task Learning Framework and a New Benchmark," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 6, pp. 7917-7932, 1 June 2023, doi: 10.1109/TPAMI.2022.3217882.

[4] D. Wanyonyi and T. Celik, "Open-Source Face Recognition Frameworks: A Review of the Landscape," in IEEE Access, vol. 10, pp. 50601-50623, 2022, doi: 10.1109/ACCESS.2022.3170037.

[5] J. Liu, Q. Li, M. Liu, and T. Wei, "CP-GAN: A Cross-Pose Profile Face Frontalization Boosting Pose-Invariant Face Recognition," in IEEE Access, vol. 8, pp. 198659-198667, 2020, doi: 10.1109/ACCESS.2020.3033675.

[6] J. Li, W. Jia, Y. Hu, S. Li, and X. Tu, "Learning to Drop Expensive Layers for Fast Face Recognition," in IEEE Access, vol. 9, pp. 117880-117886, 2021, doi: 10.1109/ACCESS.2021.3106483.

[7] H. H. Nguyen, S. Marcel, J. Yamagishi, and I. Echizen, "Master Face Attacks on Face Recognition Systems," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 4, no. 3, pp. 398-411, July 2022, doi: 10.1109/TBIOM.2022.3166206.

[8] P. Terhörst, M. Huber, N. Damer, F. Kirchbuchner, K. Raja, and A. Kuijper, "Pixel-Level Face Image Quality Assessment for Explainable Face Recognition," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 5, no. 2, pp. 288-297, April 2023, doi: 10.1109/TBIOM.2023.3263186.

[9] Z. Huang, J. Zhang, and H. Shan, "When Age-Invariant Face Recognition Meets Face Age Synthesis: A Multi-Task Learning Framework and a New Benchmark," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 6, pp. 7917-7932, 1 June 2023, doi: 10.1109/TPAMI.2022.3217882.

[10] A. Naresh, B. N. Kumar, and V. Sravanthi, "IoT Based Object Detection and Identification with OpenCV using Web CAM," 2024 1st International Conference on Innovative Sustainable Technologies for Energy, Mechatronics, and Smart Systems (ISTEMS), Dehradun, India, 2024, pp. 1-4, doi: 10.1109/ISTEMS60181.2024.10560113.

[11] K. S. Varun, I. Puneeth, and T. P. Jacob, "Virtual Mouse Implementation using OpenCV," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 435-438, doi: 10.1109/ICOEI.2019.8862764.

[12] L. Tan, F. Wu, X. Yin, and W. Liu, "Face recognition algorithm based on OpenCV," 2021 6th International Conference on Communication, Image and Signal Processing (CCISP), Chengdu, China, 2021, pp. 96-100, doi: 10.1109/CCISP52774.2021.9639288.

[13] R. Ge, Y. Mao, S. Li and H. Wei, "Research On Ship Small Target Detection In SAR Image Based On Improved YOLO-v7," 2023 International Applied Computational Electromagnetics Society Symposium (ACES-China), Hangzhou, China, 2023, pp. 1-3, doi: 10.23919/ACES-China60289.2023.10249265.

[14] M. C. S., "YOLO V7: Advancing Printed Circuit Board Defect Detection and the Quality Assurance," 2023 Global Conference on Information Technologies and Communications (GCITC), Bangalore, India, 2023, pp. 1-5, doi: 10.1109/GCITC60406.2023.10425869.

[15] S. Chen, J. Liang, J. Zhu, and X. Tian, "New Methods for Lunar Impact Crater Detection Based on YOLO v7 with Deformable ConvNets," 2023 IEEE International Conference on Electrical, Automation and Computer Engineering (ICEACE), Changchun, China, 2023, pp. 123-127, doi: 10.1109/ICEACE60673.2023.10442883.

[16] Q. Wang, Z. Liao, and M. Xu, "Wire Insulator Fault and Foreign Body Detection Algorithm Based on YOLO v5 and YOLO v7," 2023 IEEE International Conference on Electrical, Automation and Computer Engineering (ICEACE), Changchun, China, 2023, pp. 1412-1417, doi: 10.1109/ICEACE60673.2023.10442092.

[17] Y. Yu and Y.-J. Kim, "A Voice Activity Detection Model Composed of Bidirectional LSTM and Attention Mechanism," 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information

Technology, Communication and Control, Environment and Management (HNICEM), Baguio City, Philippines, 2018, pp. 1-5, doi: 10.1109/HNICEM.2018.8666342.

[18] Y. Tachioka, "Dnn-Based Voice Activity Detection Using Auxiliary Speech Models in Noisy Environments," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, pp. 5529-5533, doi: 10.1109/ICASSP.2018.8461551.

[19] A. R. Agarwal, S. Tiwari, V. V. Patage, S. Ganesh S, and S. M. S, "A Method for Voice Activity Detection using K-Means Clustering," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-5, doi: 10.1109/ICCCNT54827.2022.9984425.

[20] N. Vaswani, A. K. Roy-Chowdhury, and R. Chellappa, ""Shape Activity": a continuous-state HMM for moving/deforming shapes with application to abnormal activity detection," in IEEE Transactions on Image Processing, vol. 14, no. 10, pp. 1603-1616, Oct. 2005, doi: 10.1109/TIP.2005.852197.