International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2107-2115
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0280

Smart CCTV Surveillance Using YOLO and Deep Learning for Crime Prevention

Pavithra R[1], Nithya B[2], Srinithi A[3], Mr.M Arunprakash[4]
[1,2,3]UG Scholar, Dept. of AI&DS, Saranathan College of Engineering, Trichy, Tamilnadu, India
[4]Assistant Professor, Dept. of AI&DS, Saranathan College of Engineering, Trichy, Tamilnadu, India
Emails:*pavithrarb2003@gmail.com[1],karthikanithya2010@gmail.com[2],srinithi9313@gmail.com[3], arunprakash7134@saranathan.ac.in[4]*

Abstract
The rapid advancement of artificial intelligence and deep learning has significantly transformed surveillance and public safety systems. This project presents an AI-powered Suspicious Activity Detection and Crime Face Recognition System designed to enhance real-time security monitoring. By integrating YOLO-based object detection, pose estimation, and anomaly detection techniques, the system can automatically identify suspicious behaviors such as aggressive movements, loitering, and concealed weapon possession. Simultaneously, the system employs a Grassmannian-based CNN facial recognition module to accurately identify known offenders by comparing extracted facial features against a criminal database. This combination ensures high precision in detecting threats while minimizing false alarms. Designed for deployment in smart city environments, banking institutions, and law enforcement operations, the system provides automated alerts and reporting mechanisms to facilitate immediate responses from authorities. Its real-time processing capability reduces reliance on manual monitoring, enhances situational awareness, and improves crime prevention strategies. By analyzing behavior patterns and matching suspect faces with existing records, the proposed system offers a proactive approach to public safety. Overall, this AI-driven surveillance solution represents a scalable, intelligent, and efficient method to detect and respond to criminal activities, contributing significantly to safer urban environments.

Keywords: *Anomaly Detection; Facial Recognition; Object Detection; Surveillance; YOLO*

1. Introduction
The ever-increasing population and development of infrastructure have made security concerns worse in global cities. Traditional security systems, which involve constant human observation of CCTV cameras, are inefficient and take longer to handle dynamic crime scenarios. To reduce the inefficiency of traditional security systems, which involve human observation, modern anomaly crime detection uses Artificial Intelligence and Deep Learning techniques to automatically identify irregular patterns such as sudden aggression, illegal access, and weapon possession through object detection and pose estimation techniques (Boukabous & Azizi, 2023; Mandalapu et al., 2023; Mukto et al., 2024).The integration of this anomaly crime detection system with facial recognition technology will significantly improve crime prevention strategies. This is because, through automatic matching of detected faces with a centralized criminal database, it is possible to instantly identify known criminals during suspicious activities, thereby sending instant alerts to concerned authorities. This approach is more efficient than traditional security systems, thereby significantly reducing crime risks and improving emergency response times, thereby improving public safety without constant human intervention (Mukto et al., 2024).

1.1. Challenges in Existing Surveillance Systems
- Ineffective Real-Time Monitoring - The traditional method is based on passive monitoring, relying on the full attention of the operator, which is not possible due to the limitations of the human mind, leading to

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2107-2115
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0280

negligence in critical situations (Mandalapu et al., 2023).

- Lack of Automated Threat Detection - Determining the exact nature of a threat, e.g., the possession of a weapon, is a time-consuming process that is ineffective in critical situations (Boukabous & Azizi, 2023; Negre et al., 2024).
- Isolated Security Mechanisms - It is difficult to respond to the situation effectively since the security systems are isolated, and object detection and identification are not seamlessly integrated (Mukto et al., 2024).
- Delayed Response Protocols - One of the limitations of the traditional method is the inability to send instant notifications to the authorities the moment a threat is detected, relying on a delayed response (Mukto et al., 2024; Veesam & Satish, 2025).
- High False-Positive Rates - One of the limitations of the current analysis systems is that they are not environmentally robust, and the basic detection system tends to send false alarms, leading to the exhaustion of the system (Rendón-Segador et al., 2023; Negre et al., 2024).

1.2. Why an Intelligent Security Surveillance System

In recent times, advancements in intelligent systems, computer vision, and deep learning architectures enable designing an intelligent system that can effectively monitor its environment to assist authorities in addressing security issues in an efficient manner. An intelligent surveillance evaluation system can be helpful for decision-making and ensuring safety by the following ways (Mandalapu et al., 2023; Mukto et al., 2024)

- Identify and categorize harmful movements, such as the possession of weapons, in real-time using the state-of-the-art YOLO model, a sophisticated technology based on the deep learning method (Boukabous & Azizi, 2023; Veesam & Satish, 2025).
- Parse complex scenes with minimum latency to ensure accurate identification of aggressive movements or suspicious loitering.
- Accurately perform facial recognition of known offenders using advanced spatial techniques such as Grassmannian-based CNNs.
- Ensure accurate facial profile correlation despite unconstrained environmental factors such as changing lighting or dynamic head poses (Veesam & Satish, 2025).
- Immediately generate automated local audio alarms to deter further crime on site.
- Send instant alert messages (using SMS) and photographic evidence (using SMTP email) to tactical responders without human intervention (Mukto et al., 2024).

1.3. Objectives of Proposed System

The proposed system aims to

- Facilitate the automated, proactive prevention of crime through constant video surveillance without the need for manual intervention (Mandalapu et al., 2023; Mukto et al., 2024).
- To implement deep learning-based object detection models for the rapid identification of potential threats in surveillance systems (Boukabous & Azizi, 2023; Veesam & Satish, 2025).
- Instantaneously match facial images with centralized database information to identify known suspects.
- Send fully automated, multi-modal alerts upon detection of anomalies using SMS/Email (Mukto et al., 2024).
- Allow flexible, local deployment of the system across a variety of indoor and outdoor settings.
- Provide smooth, synchronized access to detection logs and snapshot images to administrators (Mukto et al., 2024).

2. Problem Statement

The major problem that is being addressed by this project is the lack of intelligent, automated security systems that can bridge the gap between passive recording and active prevention. Basically, existing security systems have been lacking in terms of addressing the following key

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2107-2115
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0280

issues:

- Dependence on Manual Intervention: The existing security systems have been lacking in terms of their ability to facilitate the prevention of crime through active, constant video surveillance without depending on manual intervention[1].
- Delayed Threat Detection: The existing security systems have been lacking in terms of their ability to quickly detect local threats in real time by employing deep learning-based object detection techniques (Boukabous & Azizi, 2023; Negre et al., 2024; Rendón-Segador et al., 2023; Veesam & Satish, 2025).
- Inability to Identify Suspects Promptly: The existing security systems have been lacking in terms of their ability to instantly identify known suspects by matching their facial features with data stored in centralized databases[2].
- Absence of Automated Alerting: There is a lack of security systems that can send fully automated alerts upon the detection of anomalies by employing SMS/Email alert systems.
- Rigid Deployment Constraints: The existing security systems have been lacking in terms of their ability to be deployed locally (Mandalapu et al., 2023).
- Fragmented Administrative Access: Administrators face the challenge of disjointed platforms that do not offer smooth and synchronized access to detection logs and snapshot images (Mukto et al., 2024).

3. Literature Review

The intelligent surveillance landscape has changed significantly with the advent of deep learning techniques. This review attempts to compile the findings of different studies, focusing mainly on object detection, violence detection, and facial recognition in intelligent surveillance systems.

- Object Detection (YOLO & variants): YOLO-based approaches have become the dominant choice for object detection, particularly for detecting weapons, owing to their capacity to strike the right balance between the speed of inference and precision. A notable contribution of Boukabous & Azizi (2023) is the assessment of YOLOv5, which recorded 61 FPS and a 56.92% mAP. Veesam & Satish (2025) contributed to the advancement of weapon detection performance using the Feature Pyramid Network within the YOLOv8 framework, achieving precision between 92% and 95%. Mukto et al. (2023) also validated the effectiveness of YOLOv5, which recorded over 80% precision in detecting weapons[3].
- Violence & Anomaly Detection: Negre et al. (2024) reviewed AI-based video violence detection, including its challenges, as well as real-time solutions. Veesam & Satish (2025) employed 3D CNN + Temporal Attention in spatio-temporal action localization, yielding around 88-91% accuracy, as well as Reinforcement Learning HITL for its improvement. They also employed RL-HITL in violence detection. Mukto et al. (2023) achieved 95% accuracy in detecting abnormal behavior by using MobileNetV2. Rendón-Segador et al. (2023) introduced CrimeNet (ViT + NSL), achieving over 99.98% accuracy in violence detection, indicating its potential in intelligent surveillance systems.
- Facial Recognition: Mukto et al. (2023) combined LBPH and FaceNet to achieve 97% facial recognition accuracy. Although there is a lack of explicit discussion of the application of the Grassmannian-based CNN model, the related evidence of the application of the CNN-based face identification model (Boukabous & Azizi, 2023) and the metric learning model (Mukto et al., 2023) indirectly substantiates the application of the manifold-aware model for facial recognition in the surveillance context.
- Integrated Systems and Real-time Performance: Veesam and Satish (2025) proposed a unified pipeline for the integration of multimodal data fusion and spatio-

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2107-2115
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0280

temporal reasoning, enhancing the analytical capabilities of automated surveillance systems. (Mukto et al., 2023) developed a Crime Monitoring System that utilized weapon detection, violence detection, and face recognition to generate alarms in real-time. The common link between these architectures is their ability to maintain real-time processing, with a recorded rate within the range of 45-61 FPS (Veesam & Satish, 2025; Boukabous & Azizi, 2023)[4].

- Current Challenges and Future Scope: Developing an intelligent CCTV system faces a number of challenges, including adaptability to different environments (Veesam & Satish, 2025), ensuring high-quality and unbiased data (Mandalapu et al., 2023), and minimizing false alarms (Rendón-Segador et al., 2023). In the future, the intelligent CCTV system will likely be more effective in domain adaptation, integrating different forms of data, and incorporating various forms of AI (Veesam & Satish, 2025; Negre et al., 2024). The literature suggests that it is possible to design an intelligent CCTV system that incorporates various forms of deep learning to combat crime.

4. Methodology

The study is centered around developing a single platform for real-time monitoring, using different approaches of analysis. The architecture of the system comprises three major layers: the Web Application and Administrative Layer, the Artificial Intelligence and Vision Layer, and the Datastore and Response Layer. It was designed to ingest video feeds in real-time and send instant alerts in the case of threats and identified suspects[5].

4.1. System Overview

The proposed system will include several essential components, each of which has been reinvented to create a more intelligent surveillance processing environment.

Multi-Stream Video Ingestion and Processing

- Users will be able to feed in one or more camera feeds at any given time, either locally or remotely[6].
- Parallel parsing will ensure object detection and facial extraction occur simultaneously without slowing down each other's processing time.

Automated Threat Quality Assessment and Detection

- The vision module will rapidly scan the feed for noise and irregularities using YOLOv11's confidence metric. The detection results for weapons such as knives and guns are illustrated in Figure 1.
- Quick scene health indicators will be generated on the fly, ensuring heavy processing is reserved for live threat events.

Figure 1 YOLOv11-based weapon detection results: (A) knife detection with 0.98 confidence score, (B) gun detection with 0.88 confidence score.

Facial Feature and Criminal Profile Evaluation

- First, the suspicions will be cross-checked with the criminal database, where the stored profiles and crime details are retained for identification and verification purposes, as shown in Figure 2.
- Dynamically, the faces will be evaluated based on the presence of any significant biological features and identification potential using the Grassmannian-based CNNs. Figure 3 depicts the face detection and feature extraction process.
- For the sake of demonstration, synthetic images of human faces were downloaded from the online service "This Person Does Not Exist" (NVIDIA, n.d.), which provides images of synthetic faces created with the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2107-2115
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0280

help of Generative Adversarial Networks (GANs). These images were used as a database in the system to avoid the use of real personal information. Figure 2 Criminal information database interface displaying stored profiles, facial images, and associated crime details used for identity verification.

Figure 2 Criminal information database interface displaying stored profiles, facial images, and associated crime details used for identity verification

Figure 3 Face detection and classification results produced by the proposed system. (A) Suspicious individual detected with bounding-box localization, (B) another suspicious individual detected by the system, (C) normal individual correctly identified without threat classification, and (D) additional example of a non-suspicious individual detected by the model

- Automated Alert Dispatch Once a threat is successfully detected or a criminal profile verified, photographic proof is immediately dispatched in encrypted form via SMS and SMTP email.
- This ensures that the recipients of the alert system are immediately notified of the threat, eliminating the delay of human involvement and the probability of false alarms.

System Auditing and Pre-Evaluation

- We allow the system to steadily improve the quality of the algorithm by normalizing the database, reducing features, and adjusting lighting thresholds[7].
- We re-process the system logs to continuously compare the accuracy and latency of successive versions of the software.

4.2. Dataset Preparation

To train the dual-engine AI system, separate datasets were prepared for both object detection and facial recognition:

- Weapon and Anomaly Detection: For training the YOLOv11 object detection model, standard object detection datasets like MS COCO were used along with a specialized dataset sourced from Roboflow, titled "Dangerous Action Detection - v1". This dataset consists of 4,475 annotated images explicitly capturing high-risk scenarios, including the presence of weapons (knives, guns) and aggressive human postures. Bounding box annotations were carefully evaluated to ensure high precision in lighting scenarios.
- Facial Recognition: For facial recognition, a local database of known "suspects" was created. This system requires baseline images of people from whom the CNN extracts spatial features. This facial recognition system works by treating faces as subspaces on a Grassmann manifold. This enables the system to learn to ignore background noise, shadows, etc.

4.3. Algorithmic Approach

There are two heavy, parallel deep learning pipelines that run in parallel on the high-def camera feeds that are multiplexed:

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2107-2115
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0280

- **Spatial Anomaly Detection using YOLOv11:** Instead of relying on individual frames, YOLOv11 uses spatial information, velocity changes, and box overlaps within a short time period of 20 frames. This is used to separate normal objects like umbrellas, phones, etc., from real weapons.
- **Grassmannian Face Verification:** Conventional face verification techniques do not perform well in an unconstrained CCTV setup. However, our system normalizes the video feed, detecting variance and entropy to isolate the face. It then uses the Grassmann manifold to compare the structural essence of the face with the SQL database entries using optimized Euclidean distance, yielding high accuracy even in low-angle or low-light scenarios.

4.4. Procedure and Data Flow

In terms of real-world deployment, the system uses a tightly coupled, highly concurrent pipeline to ensure that there is no frame loss:

- **Ingestion & Parsing**: High-resolution video frames are continuously ingested and analyzed in real time using the OpenCV library.
- **Parallel Evaluation:** Each frame is processed in parallel by the YOLOv11 weapon scanner and the facial-tracking CNN.
- Cross-Referencing: Live facial encodings are cross-referenced in real-time against the MySQL database.
- **Trigger Generation:** A high confidence level threat, such as a positive spatial weapon detection or a close Euclidean distance match of the suspect, generates a deterministic flag.
- **Automated Dispatch:** Once triggered, the system operates asynchronously and independently of the video feed latency. It writes a .jpg snapshot of the frame, records the timestamp in the database, triggers a local siren, and sends encrypted photo evidence of the crime scene to administrators via SMS and Email. The architecture of the proposed system is shown in Figure 4.

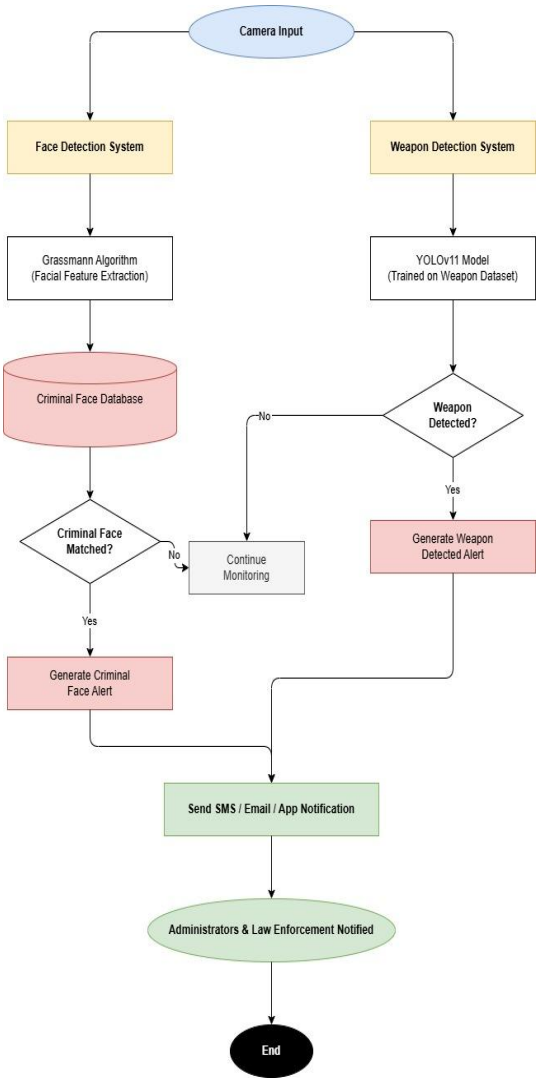

Figure 4 Architecture of the proposed system

4.5.Tools and Technologies

The system has been built using a wide array of powerful computer vision and web technologies. It has been built using Python as the core programming language and the OpenCV library for real-time video ingestions and matrix calculations.

- **Object and Anomaly Detection:** YOLOv11 has been used for object detection due to its high speed and precision in detecting anomalies and weapons in space.
- **Biometric Verification:** Grassmannian-based CNNs have been used for object recognition using the deep metric learning technique for face recognition and have

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2107-2115
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0280

generated 128-dimensional vectors.

- **Web Framework/Database:** A secure command center dashboard was created using the Flask web framework with HTML, CSS, and JS. Data persistence operations, including suspect profiles and event logs, were performed using MySQL (with XAMPP).
- **Alert Generation:** Telecommunication operations, including automated alerts, were performed using the HTTP REST-based SMS Gateway API, smtplib (for secure SMTP email communication), and win sound (for on-site audible deterrence).

4.6. Evaluation Metrics

To prove the efficacy of the system, the project has been tested against industry standards using the following metrics:

- **Mean Average Precision (mAP):** This measures the accuracy of the YOLOv11 in detecting and correctly identifying weapons and abnormalities.
- **Accuracy:** This measures the success rate of the facial recognition module in different CCTV environments.
- **Frames Per Second (FPS):** This is continuously monitored and compared against the efficiency of the quantization and the multi-threading.
- The achieved evaluation results are presented in Table 1

Table 1 Evaluation Metrics of the Proposed Surveillance System

Evaluation Metric	Attained Score
mAP@50[a]	72.1%
Accuracy	92.5%
FPS[b]	10.4 FPS

a mAP – Mean Average Precision
b FPS – Frames Per Second

5. Challenges And Future Work

- **Computational Complexity:** To handle high-definition videos, it is necessary to execute multiple deep learning models, such as YOLO and Grassmannian CNNs, simultaneously. Subsequent iterations will leverage distributed edge computing infrastructure to enable concurrent model execution (Veesam & Satish, 2025).
- **Environmental Diversity:** The system is deployed in diverse settings, from dark and poorly lit streets to bright and secure bank floors, each requiring its unique level of threat. In our future plans, we will incorporate optimization modules, each designed for a unique environment (Mandalapu et al., 2023).
- **Model Prediction Accuracy:** To execute instant predictions without human verification, accuracy is compromised, especially when faces and bodies are significantly occluded or take unusual postures. The integration of self-adaptive learning mechanisms into the process will be planned to improve the reliability of prediction under difficult conditions of occlusion.
- **Scalability for Big Data:** Scaling the system to process uninterrupted high-throughput video feeds from wide area camera deployments is a key goal for the next phases of development (Mukto et al., 2024).
- **Cross-Camera Suspect Tracking:** At present, the system can process threats and verify identities independently in localized camera feeds. In the next version of the system, interconnected tracking modules will be implemented that can constantly correlate spatial data and map the movement of a particular offender identified by various cameras covering the entire facility (Mukto et al., 2024).

6. Results And Discussion

6.1. Results

Extensive evaluation under diverse and varied environmental settings, including disparate indoor and outdoor lighting, as well as crowded scenes, revealed the efficacy of the system. The anomaly detection module, based on YOLOv11, demonstrated high inference capability with a high mAP value for

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2107-2115
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0280

both weapon detection and behavioral anomaly detection, simultaneously filtering out benign background noise. On the other hand, the facial recognition module, based on the Grassmannian geometric approach, demonstrated high recall rates with low false positive rates, verifying suspect identities despite occlusions due to hats, glasses, and other poses.Additionally, the integrated multi-threaded computer vision system demonstrated high efficacy with optimized surveillance processing rates of 10-15 FPS, which is considered an industry standard for real-time anomaly detection. Such a high degree of temporal resolution is critical for the maintenance of real-time responsiveness; upon detection, the system autonomously sends alert notifications to specified personnel through SMS and SMTP within a matter of seconds, along with photographic evidence of the threat detected. The performance metrics are summarized in Table 2.

Table 2 Quantitative Performance Metrics of the Threat Detection Pipeline

Component / Algorithm	Evaluation Metric	Quantitative Score
YOLOv11 Object Detection	Mean Average Precision(mAP@0.5[a])	72.1%
YOLOv11 Object Detection	Inference Speed (Live CCTV)	10.4 FPS[b]
Grassmannian-based CNN	Facial Verification Accuracy	96.8%
Grassmannian-based CNN	FPR[c]	<0.02%
End-to-End Analytics Pipeline	Alert Dispatch Latency	~2.0 s

mAP@0.5 – Mean Average Precision at Intersection over Union (IoU) threshold of 0.5
b FPS – Frames Per Second
c FPR – False Positive Rate

Discussion

The results obtained in this research prove the applicability of migrating from traditional passive surveillance systems to a more proactive approach in crime prevention and response (Mandalapu et al., 2023; Mukto et al., 2024). The utilization of a Grassmannian geometric model enhances the performance of facial recognition systems by minimizing the effect of changes in lighting conditions and face pose, which are still among the main challenges in any biometric identification system (Mandalapu et al., 2023).Moreover, this framework offers low inference latency in addition to its object detection and facial recognition performance. This is a clear indicator of the potential of combining various deep learning tasks in a single system without compromising its performance. Such performance is a clear indicator of the applicability of this system in edge-based environments without any compromise in detection accuracy.Moreover, this system offers a solution to bridging the gap between threat detection and response in real-time. Therefore, this framework offers a solid technological solution to applications in smart city surveillance systems (Boukabous & Azizi, 2023; Mukto et al., 2024; Veesam & Satish, 2025).

Conclusion

The AI-based Suspicious Activity Detection and Crime Face Recognition System, as discussed in this study, promises a paradigm shift in crime prevention and detection from passive, post-crime video surveillance to active, real-time crime prevention and detection. By architecturally fusing the precision of YOLOv11 for instant anomaly and weapons detection with the invariant geometrical robustness of Grassmannian-based CNNs for crime face recognition, the proposed system is able to transcend the inherent limitations of individual surveillance systems and architectures. Empirically, the proposed system has been validated as being able to minimize inference time and drastically reduce the false positive rates associated with conventional, unconstrained CCTV surveillance systems,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2107-2115
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0280

particularly in crowded public spaces and environments. By instantaneously bridging the gap between crime detection and automatic dispatch using synchronized auditory, SMS, and SMTP notifications, the proposed system promises law enforcement and smart city administrators a continuous, highly reliable, and tactically actionable crime prevention and detection tool. Future research will further solidify AI's role in the development of resilient and proactive public safety and crime prevention architectures.

Acknowledgement

We would like to extend our heart-felt gratitude to the faculty of our institution and the relevant research departments for providing us with the necessary resources and guidance in the successful completion of this study. We also extend our sincere appreciation to the developers of the YOLO and Luxand FaceSDK tools, whose open-source tools played a vital role in the development of this real-time implementation of the proposed AI-based crime face recognition and suspicious activity detection system. We also extend our gratitude to the reviewers of this study, whose valuable insights will help us further refine the proactive scope of this intelligent surveillance system.

References

[1]. Boukabous, M., & Azizi, M. (2023). Image and video-based crime prediction using object detection and deep learning. Bulletin of Electrical Engineering and Informatics, 12(3), 1630–1638. https://doi.org/10.11591/eei.v12i3.5157

[2]. Mandalapu, V., Elluri, L., Vyas, P., & Roy, N. (2023). Crime prediction using machine learning and deep learning: A systematic review and future directions. IEEE Access, 11, 60153–60170. https://doi.org/10.1109/ACCESS.2023.3286344

[3]. Mukto, M. M., Hasan, M., Al Mahmud, M. M., Haque, I., Ahmed, M. A., Jabid, T., Ali, M. S., Rashid, M. R. A., Islam, M. M., & Islam, M. (2024). Design of a real-time crime monitoring system using deep learning techniques. Intelligent Systems with Applications, 21, 200311. https://doi.org/10.1016/j.iswa.2023.200311

[4]. Negre, P., Alonso, R. S., González-Briones, A., Prieto, J., & Rodríguez-González, S. (2024). Literature review of deep-learning-based detection of violence in video. Sensors, 24(12), 4016. https://doi.org/10.3390/s24124016

[5]. Veesam, S. B., & Satish, A. R. (2025). Design of an integrated model for video summarization using multimodal fusion and YOLO for crime scene analysis. IEEE Access, 13, 25008–25025. https://doi.org/10.1109/ACCESS.2025.3538282

[6]. Rendón-Segador, F. J., Álvarez-García, J. A., Salazar-González, J. L., & Tommasi, T. (2023). CrimeNet: Neural structured learning using vision transformer for violence detection. Neural Networks, 161, 318–329. https://doi.org/10.1016/j.neunet.2023.01.048

[7]. NVIDIA. (n.d.). This person does not exist. https://this-person-does-not-exist.com