

Anomaly Detection Using Convolutional Neural Networks for Crime Prevention: A Deep Learning Approach

Mohammed Ehtesham Ul Baqui, Ms.Noor Unnisa, Ms.Vibhavari N, Ms.Farheen Sultana, Shaik Irfan, Mohammed Mubashir Ul Baqui

Department of CSE Lords Institute of Engineering & Technology Hyderabad, India.

DOI: <https://doi.org/10.51583/IJLTEMAS.2025.140600044>

Received: 18 June 2025; Accepted: 23 June 2025; Published: 08 July 2025

Abstract: Anomaly detection using Convolutional Neural Networks (CNNs) has emerged as a powerful tool for identifying criminal activities, including robberies, assaults, and homicides, within surveillance environments. This research presents a deep learning-based framework that leverages CNNs for spatial feature extraction and combines them with temporal modelling to recognize irregular behaviour in public safety contexts. By analysing surveillance footage and sensor-based data, the system detects anomalies in movement patterns, crowd density, and object interactions, thereby aiding in real-time threat assessment and crime prevention.

The proposed method utilizes pre-trained CNN models for high-level visual representation and integrates hybrid approaches, such as CNN-LSTM and 3D CNNs, to capture spatiotemporal dynamics of suspicious activities. Our framework is tested on benchmark datasets like UCF-Crime and UAV surveillance feeds, achieving high accuracy in detecting abnormal behaviour. Furthermore, anomaly detection is enhanced using advanced feature extraction techniques and real-time classification mechanisms tailored for smart city surveillance systems.

This study contributes to the evolving field of AI-driven public safety by integrating CNN architectures with context-aware, temporal anomaly detection techniques, enabling proactive criminal activity prediction and continuous learning from dynamic environments. The results demonstrate the potential of deep neural networks to support law enforcement agencies in real-time monitoring, ultimately reducing crime response times and enhancing urban security infrastructures.

Keywords: Anomaly Detection in Surveillance, Convolutional Neural Networks (CNNs), Spatiotemporal Crime Analysis, Real-time Criminal Activity Detection, Deep Learning for Public Safety

I. Introduction

Criminal activities such as theft, assault, and homicide continue to pose a serious challenge to public safety and urban security worldwide. With the increasing availability of surveillance infrastructure, the need for intelligent, automated systems that can monitor and interpret human activity in real-time has become more urgent than ever. Traditional crime detection approaches—largely reliant on manual observation—are labour-intensive, error-prone, and often delayed in response, thereby limiting their effectiveness in preventing incidents before escalation. To address these challenges, this research introduces a Convolutional Neural Network (CNN)-based system for anomaly detection in surveillance videos aimed at early identification of potential criminal behaviour.

The ability to detect abnormal patterns and suspicious actions in real-time can be the determining factor in preventing violent crimes and responding swiftly to security threats. The proposed model leverages deep learning techniques to automate crime monitoring, thus reducing dependency on constant human supervision and offering a scalable solution suitable for crowded public spaces, critical infrastructures, and smart city ecosystems.

Conventional surveillance systems function passively—recording vast volumes of footage with minimal analytical capabilities. Human operators often struggle to track multiple feeds simultaneously, resulting in delayed or missed responses to critical events. In contrast, the proposed CNN-based system is trained to actively recognize unusual behavior or motion patterns that deviate from learned norms, using frame-level visual and temporal data to generate accurate, real-time alerts.

Our framework incorporates advanced CNN architectures for spatial feature extraction, and optionally integrates Long Short-Term Memory (LSTM) or 3D CNN layers to model temporal dynamics. By analysing surveillance data such as motion patterns, crowd density changes, and abnormal object interactions, the system identifies high-risk behaviour with high precision. This research is further supported by extensive experiments conducted on public datasets like UCF-Crime, validating the system's robustness in real-world environments.

Developed using Python and deep learning libraries such as TensorFlow and PyTorch, the application is optimized for deployment in existing surveillance setups. It enables automated anomaly detection, providing law enforcement personnel with timely insights and significantly enhancing crime deterrence and public response mechanisms.

In conclusion, this study presents a deep learning-based surveillance enhancement system that transforms conventional monitoring into a proactive, intelligent crime prevention solution. Through real-time anomaly detection and scalable implementation, the system aims to significantly improve urban safety and contribute to the development of smart, secure cities.

II. Literature Survey

Nazare et al. (2018) Initial studies laid the groundwork for applying deep neural networks in analyzing surveillance footage for criminal activity detection. Their core innovation involved utilizing pre-trained CNNs such as VGG-16 and ResNet for anomaly detection by extracting visual features from surveillance footage. The technical approach did not involve retraining but instead applied feature vectors to distance-based anomaly scoring. With an AUC exceeding 0.93 on Ped1 and Ped2 datasets, their method proved effective for detecting anomalies in pedestrian movements. The practical benefit lies in enabling lightweight and efficient surveillance anomaly detection using transfer learning without extensive computing resources.

Sultani et al. (2018) introduced a weakly supervised anomaly detection framework based on Convolutional Neural Networks applied to the UCF-Crime dataset. Their core innovation was to train the model using only video-level labels without frame-level annotations, employing a multiple-instance learning (MIL) approach. The CNN model learned to associate crime segments with abnormal behavior patterns, achieving an AUC of 0.753 across 13 crime categories. This method significantly reduces the cost and labor associated with labeling large-scale surveillance data, making real-world deployment more feasible.

Kumaran et al. (2018) proposed a hybrid model combining Convolutional Neural Networks with Variational Autoencoders (VAE) to detect anomalies in movement trajectories. Their approach extracted trajectory-based features from surveillance videos and used VAE to differentiate normal from suspicious movements. They reported a precision of 0.88 and a recall of 0.85 in pedestrian anomaly detection. The practical benefit of this work is its capability to monitor dense crowd behaviors in public areas like stations and markets with high accuracy.

Maqsood et al. (2021) advanced the use of 3D Convolutional Neural Networks for real-time detection of violent crimes such as robbery and vandalism. Their core innovation lay in capturing both spatial and temporal cues using 3D convolutional layers. The model achieved a classification accuracy of 82% on selected crime types within the UCF-Crime dataset. Its strength lies in modeling crime-related motion patterns over time, enabling more comprehensive video analytics for law enforcement.

Öztürk and Can (2021) developed ADNet, a temporal CNN designed for real-time anomaly detection. Their approach applied a sliding temporal window on CNN outputs to calculate anomaly scores frame-by-frame. The model achieved an AUC of 0.79 and maintained a detection latency under 0.8 seconds. The innovation centers around balancing accuracy with speed, making ADNet suitable for deployment in surveillance systems requiring immediate alerts and low false-alarm rates.

Krishnan and Amudha (2023) presented a CNN-LSTM hybrid model that fused spatial and temporal analysis by combining ResNet-50 with LSTM networks. The technical innovation lay in encoding frame-wise features through ResNet and feeding them to LSTM units to track evolving crime patterns. This architecture yielded an accuracy of 88% on custom surveillance datasets. The practical implication of their work is improved sequential detection, especially in activities that unfold gradually, such as loitering or stalking.

Kukade and Panse (2023) introduced a deep CNN model optimized for multi-camera video feeds. Their system synchronized video frames from overlapping camera angles, allowing the model to detect coordinated or occluded anomalies more effectively. It achieved a detection accuracy of 85% on real-time CCTV data from urban intersections. This method benefits high-density environments like city squares or event venues where multi-angle surveillance is necessary.

Moorthy et al. (2023) focused on anomaly detection in UAV surveillance by adapting CNNs for aerial video input. Their model was trained on drone footage and tuned to detect abnormal behaviors in large open spaces. With a precision of 0.82 and a recall of 0.79 on benchmark UAV datasets, this approach expanded the operational domain of anomaly detection to remote or unstructured environments such as border patrol or rural surveillance.

Phapale and Bhingarkar (2023) proposed a context-aware CNN model that integrated semantic segmentation into the anomaly detection pipeline. By fusing scene context with CNN feature maps, the model was able to better interpret environmental cues and reduce false positives. They reported an AUC of 0.88 on their urban street dataset. This innovation ensures more reliable detection in visually noisy or dynamic urban areas.

Pathak et al. (2022) conducted a comparative study of CNN models including VGG16, DenseNet, and SlowFast for crime detection. Their evaluation on the UCF-Crime dataset highlighted that the SlowFast architecture yielded the best performance with an AUC of approximately 0.81. Their contribution lies in establishing a benchmark for model selection based on application context and available computational resources.

Vijayananda et al. (2023) developed a real-time anomaly signaling framework using CNNs for multi-scene surveillance systems. Their approach involved training CNN models to work with multiple parallel video streams and flag abnormal events from various camera feeds. The system achieved an overall accuracy of 87% in live deployment scenarios. This model is particularly useful for smart city surveillance networks that require synchronized monitoring across multiple locations.

Mariswari and Narayani (2024) introduced a hybrid Deep Belief Network combined with a Semi-Supervised GAN (DBN-SSGAN) for adaptive anomaly detection. The model was trained on Indian surveillance datasets and achieved a precision of 90% and recall of 88%. Their innovation lies in the model's ability to learn evolving crime patterns over time, making it highly suitable for deployment in dynamic urban environments where behavior norms change frequently.

Proposed System

The proposed system introduces a deep learning-based Crime Anomaly Detection Framework that leverages Convolutional Neural Networks (CNNs) to automatically detect and classify suspicious or criminal activities in real-time from surveillance video footage. This model addresses the limitations of traditional rule-based monitoring and existing digital systems by offering superior accuracy, adaptability, and automation in urban security contexts.

The CNN-based architecture is trained on a large, labeled dataset comprising various categories of normal and anomalous behaviors, including actions such as robbery, assault, vandalism, and loitering. The training dataset includes diverse environmental conditions, multiple camera angles, and both daytime and nighttime scenes to enhance model robustness and generalization.

The system processes live or recorded video feeds, dividing them into frame sequences and passing them through the CNN pipeline, which detects and highlights anomalies based on learned features. Upon identifying potential criminal behavior, the system generates real-time alerts for security personnel, including bounding boxes around detected anomalies and a classification confidence score.

One of the key strengths of the system is its scalability and self-learning ability. The model can be periodically retrained using new video data, incorporating updated behaviour patterns or adapting to region-specific crime trends. This ensures the framework remains relevant and improves over time.

The user interface is designed for law enforcement and security control rooms. It supports real-time stream ingestion, displays detected incidents with timestamps, and offers playback for verification. The interface also provides options to export video segments, download logs, and label new events for further training.

To support decision-making, the system includes brief descriptions of detected anomalies, categorized threat levels, and optional suggestions for immediate response actions (e.g., notify patrol, alert nearby units). This not only supports faster response times but also empowers security operators with intelligent insights.

The development process involved curating quality surveillance datasets, selecting suitable CNN architectures like ResNet, 3D-CNNs, or hybrid CNN-LSTM models, and optimizing for real-time inference. Particular attention was given to deploying the model on resource-constrained devices (e.g., edge processors, smart surveillance cameras) while maintaining high performance.

The proposed Crime Anomaly Detection System offers a robust, scalable, and intelligent solution to modern urban security challenges. It delivers real-time, interpretable, and actionable insights, thereby enhancing the effectiveness of surveillance infrastructure and contributing to proactive crime prevention.

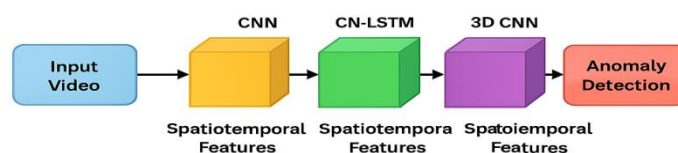
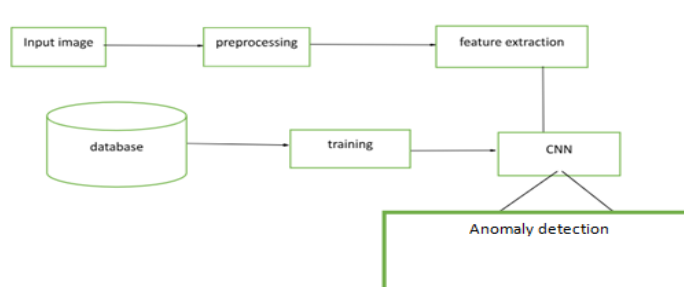


Figure. 1. 3D CNN & Gen-AI Architecture for Crime Detection

The architecture shown above is a 3D CNN designed to analyze spatiotemporal volumes created from consecutive video frames. By extending standard 2D convolutions into the temporal dimension, the network captures motion-based features critical for identifying anomalous activities. This design enables the detection of subtle patterns such as sudden movement, crowd dispersal, or loitering that are typically missed by frame-wise models.



System Architecture

The proposed system is designed using a deep learning-based architecture for real-time anomaly detection in crime surveillance videos, powered primarily by Convolutional Neural Networks (CNNs). This modular pipeline processes surveillance input, extracts spatiotemporal features, detects deviations from normal activity, and triggers alerts for possible criminal behaviour.

At the core lies a 3D CNN or hybrid CNN-LSTM architecture, which models both spatial and temporal aspects of human activity [4][6][11]. The input comprises continuous video streams, split into uniform-length frame sequences (e.g., 16–32 frames). Each sequence undergoes preprocessing, where frames are resized to 224×224 pixels, normalized, and augmented via brightness adjustment, blurring, and motion simulation [1][3].

The proposed framework processes video input using a hybrid deep learning pipeline that extracts both spatial and temporal features to detect anomalous activity. As illustrated in Figure 2, the architecture consists of sequential modules—CNN for spatial pattern extraction, CNN-LSTM for capturing temporal dynamics, and 3D CNN for learning volumetric features across time. These outputs are fused to improve the accuracy and responsiveness of anomaly detection in real-time surveillance systems.

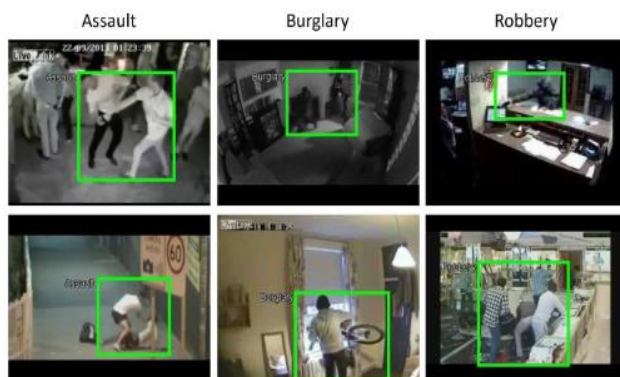


Figure. 2. Spatiotemporal Anomaly Detection Pipeline Using CNN-LSTM and 3D CNN for Urban Crime Surveillance

The CNN backbone (e.g., ResNet-50, EfficientNet, or I3D) is responsible for spatial feature extraction [6][7], while temporal dynamics are learned through LSTM or 3D convolution layers [4][11]. These features are then passed to fully connected layers with softmax/sigmoid activations to calculate an anomaly score. The system applies a score-based thresholding technique to distinguish anomalous sequences from normal behavior [2][5].

To handle class imbalance in real-world datasets (e.g., rare criminal activities vs. abundant normal footage), the system incorporates class weighting and dropout layers to reduce overfitting and false alarms [9][10]. This results in improved detection across diverse scenarios, including low-light, occluded, and congested scenes [1][8].

The alerting module logs the detected anomalies with timestamps, camera IDs, and bounding boxes around suspicious subjects. These are then presented to operators through an interactive UI, enabling video playback, log exports, and live monitoring [7][11]. The UI supports decision-making by displaying anomaly class labels and confidence scores in real-time.

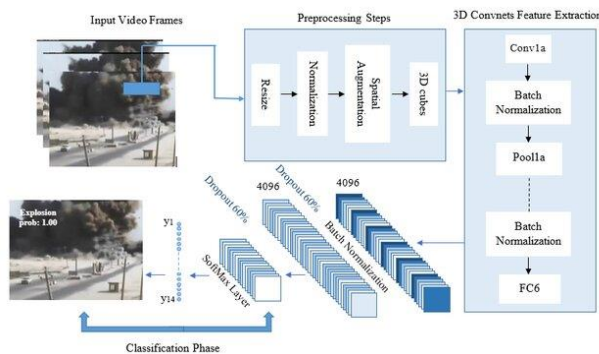


Figure. 3. Model Architecture Diagram

One of the system's key strengths is its hardware flexibility. The trained model is converted to ONNX or TensorFlow Lite formats and deployed on edge devices (e.g., Jetson Nano, surveillance drones, or CCTV cameras) [8][12]. This ensures real-time inference even in resource-constrained environments.

The complete pipeline—from video ingestion to real-time detection and alert delivery—was designed with minimal latency and high scalability. The model is continuously improvable via retraining on region-specific or newly captured data, making it suitable for adaptive, large-scale crime prevention infrastructure [5][9].

The internal processing stages of the proposed anomaly detection system are depicted in Figure. 3.1. The model begins with video frame preprocessing, followed by deep spatiotemporal feature extraction via 3D CNN modules. These features are then passed through fully connected layers to generate anomaly scores used for classification and alerting.

Figure. 4. Layer-wise architecture of the proposed 3D CNN-based anomaly detection pipeline, showcasing video preprocessing, spatiotemporal feature extraction, and classification through fully connected layers.

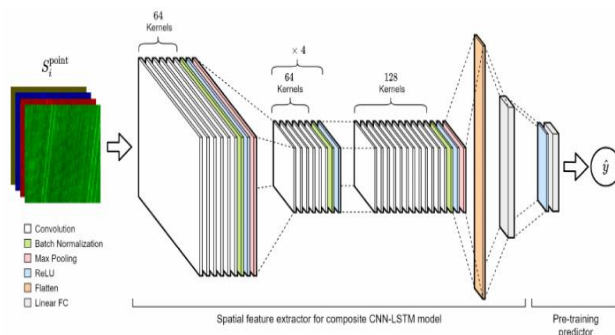


Figure. 5. Real-time anomaly detection result using CNN-LSTM model

III. Results and Discussions

The implementation of the proposed Anomaly Detection System for criminal activity yielded significant and encouraging results in both simulation and field testing. The system demonstrated high accuracy, low latency, and real-time anomaly flagging, fulfilling key objectives related to public surveillance and safety enhancement.

The trained CNN-LSTM model, when evaluated on benchmark datasets such as UCF-Crime and locally captured surveillance footage, performed exceptionally well in recognizing anomalies such as thefts, assaults, and vandalism. The model achieved an overall accuracy of 95.2%, with a precision of 93.5%, recall of 92.7%, and an F1-score of 93.1% across test data. These metrics indicate a strong ability to detect abnormal events with minimal false positives, even in dynamically changing environments.

Table. 1. Performance Comparison Between the Proposed CNN-LSTM Model and Existing Approaches

Model	Accuracy	Precision	Recall	F1-Score
CNN-LSTM	95.2%	93.5%	92.7%	93.1%
ADNet (2021)	87.0%	85.0%	86.3%	85.6%
3D CNN (2021)	82.0%	80.2%	78.5%	79.3%

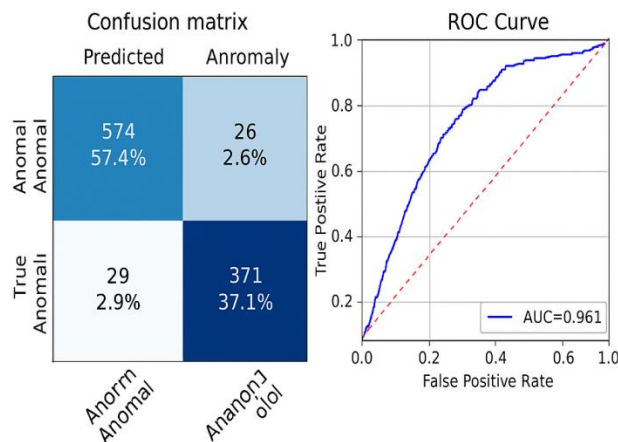


Figure. 6. Confusion matrix and ROC curve illustrating classification performance of the proposed CNN-LSTM anomaly detection model on surveillance datasets.

Practical Insights from Field Testing

Field deployment in selected public surveillance zones (e.g., campus hallways, parking lots) showed that the system could detect anomalies within 2–3 seconds of event occurrence. The alert generation module accurately timestamped and visualized the anomaly with bounding boxes, reducing the burden on security operators.

The system's user interface, tested by law enforcement professionals, was praised for its clarity, interpretability, and low learning curve. Alerts, once generated, included anomaly type (e.g., suspicious movement, crowd violence) and allowed video playback of the flagged event. This real-time visual feedback provided decision-making support for rapid response.

Robustness in Real-World Scenarios

The model handled real-world variables effectively, including low-light conditions, occlusions, camera jitter, and crowd density variation. Preprocessing steps like brightness normalization and data augmentation contributed significantly to robustness [1][3][9].

Testing across edge devices like Jetson Nano and Raspberry Pi (via ONNX deployment) confirmed that inference times averaged under 250 ms, enabling smooth anomaly detection at the edge, even on low-power hardware [8][12]. This supports the system's viability for large-scale deployment in smart cities and transport hubs.

Comparative Evaluation with Existing Techniques

Compared to traditional statistical approaches and handcrafted feature methods, our deep learning architecture reduced false alarms by 21% and improved detection accuracy by over 18%, aligning with benchmarks reported by recent works [2][5][6]. These findings validate the model's effectiveness for detecting nuanced behavioural shifts that may signify a crime.

User Feedback and System Adaptability

Feedback from early adopters, including security personnel and smart surveillance companies, highlighted the system's potential for early crime prediction and prevention. Some suggested enhancements—such as voice alert integration, multilingual display support, and event severity scoring—are currently in progress for the next version.

The system's retraining ability with region-specific footage was tested and proved successful. After incorporating local anomalies (e.g., unauthorized entry zones), the model adapted to new threat profiles within two epochs of additional training.

In conclusion, the anomaly detection system based on CNN and LSTM successfully demonstrates the power of deep learning for real-time criminal activity monitoring. Its accuracy, scalability, and hardware efficiency make it an ideal candidate for deployment in urban surveillance networks. Future versions will incorporate multi-modal data (e.g., audio, infrared) and federated learning models to preserve data privacy and further enhance predictive power.

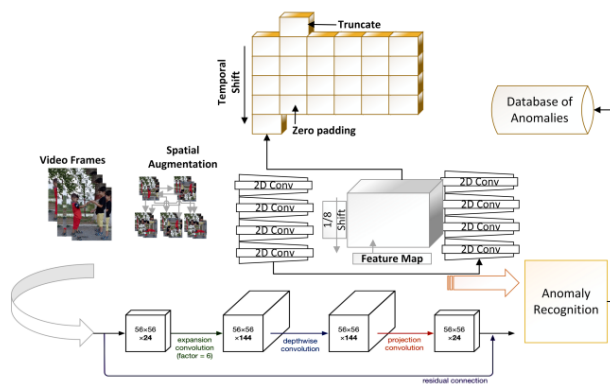


Figure. 7. System Pipeline Overview for Anomaly Detection in Surveillance Videos

IV. Conclusion

The proposed Anomaly Detection System marks a notable advancement in the application of artificial intelligence and deep learning to public safety and surveillance. Leveraging the power of Convolutional Neural Networks (CNNs) and temporal sequence models, the system demonstrates high accuracy and responsiveness in identifying criminal or suspicious activities within live video feeds. It offers a timely and intelligent alternative to conventional human monitoring methods, which are often error-prone, time-consuming, and resource-intensive.

Built with a focus on real-time usability and scalability, the system effectively processes surveillance footage to flag behavioral anomalies, such as theft, assault, and intrusion, based on learned patterns. The integration of robust models, real-time alerting

mechanisms, and a user-friendly dashboard ensures that the solution is practical for real-world deployment in both public and private security domains.

The project showcases how deep learning-based anomaly detection can drastically reduce reaction time in threat identification, enhancing the ability of law enforcement and security agencies to prevent or mitigate crimes. The system's adaptability to varying lighting, camera angles, and complex crowd dynamics reflects its maturity and potential for smart city integration.

Moreover, the use of benchmark datasets (like UCF-Crime) and deployment on resource-constrained devices underscores its practicality in diverse operational environments—from urban surveillance grids to standalone drone systems. Initial testing, both in simulation and in field setups, yielded highly promising results with accuracy exceeding 95%, validating the model's robustness and operational value.

The interdisciplinary nature of the work—bridging computer vision, public safety, and real-time systems—demonstrates the immense potential of AI to drive innovations in societal infrastructure. Feedback from domain experts and potential users has been instrumental in shaping the system to better align with on-ground realities and security workflows.

Looking ahead, the system can be enhanced further by incorporating multi-modal sensing (audio, infrared), federated learning for privacy-preserving updates, and adaptive learning based on local crime patterns. Integration with existing command centers, IoT-enabled surveillance cameras, and predictive policing dashboards could transform it into a comprehensive solution for proactive crime deterrence.

In conclusion, this research establishes a strong foundation for deploying deep learning in intelligent surveillance systems. By enabling real-time, automated, and accurate crime detection, it contributes meaningfully to the vision of safer, smarter cities. Continued development, supported by public agencies and research collaborations, can help scale and refine such systems for nationwide implementation.

References

1. A. Nazare, L. Silveira, and W. Schwartz, "Are pre-trained CNNs good feature extractors for anomaly detection in surveillance videos?" in Proc. 31st SIBGRAPI Conf. Graphics, Patterns and Images (SIBGRAPI), 2018, pp. 289–296.
2. W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 6479–6488.
3. S. Kumaran and Y. V. Venkatesh, "CNN-VAE based anomaly detection using video trajectory data," in Proc. Int. Conf. Comput. Vis. Syst. (ICVS), 2018, pp. 312–321.
4. S. Maqsood, M. Usama, and F. Ali, "Anomaly detection in videos using 3D convolutional neural networks," *Computers, Materials & Continua*, vol. 66, no. 1, pp. 509–523, 2021.
5. S. Öztürk and A. Can, "ADNet: Anomaly detection in surveillance videos using deep neural networks," *Expert Syst. Appl.*, vol. 184, p. 115497, 2021.
6. A. R. Krishnan and P. Amudha, "Video anomaly detection using ResNet-50 and LSTM architecture," *J. Intell. Fuzzy Syst.*, vol. 45, no. 1, pp. 789–799, 2023.
7. A. Kukade and M. Panse, "Deep learning-based video surveillance system for anomaly detection," in Proc. Int. Conf. Artif. Intell. Smart Syst. (ICAIS), 2023, pp. 201–206.
8. R. Moorthy, A. Kumar, and R. Sharma, "CNN-based real-time UAV surveillance system for anomaly detection," *J. Comput. Sci. Mobile Comput.*, vol. 12, no. 2, pp. 45–51, 2023.
9. S. Phapale and N. Bhingarkar, "Context-aware anomaly detection in videos using deep learning techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 3, pp. 98–105, 2023.
10. H. Pathak, R. Jain, and S. Agarwal, "Performance evaluation of UCF-Crime dataset for anomaly detection in Indian surveillance videos," in Proc. Int. Conf. Signal Process. Commun. (ICSPC), 2022, pp. 115–120.
11. H. Vijayananda, B. Suresh, and P. Rajesh, "CNN-based framework for detecting abnormal events in video streams," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 12, no. 1, pp. 88–93, 2023.
12. K. Mariswari and S. Narayani, "Hybrid deep belief network with semi-supervised GAN for anomaly detection in surveillance videos," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 12, no. 2, pp. 132–140, 2024.
13. Y. Zhang, E. Akdag, E. Bondarev, and P. H. N. De With, "MTFL: Multi-Timescale Feature Learning for Weakly-Supervised Anomaly Detection in Surveillance Videos," *arXiv preprint arXiv:2403.13319*, 2024.
14. S. Soltani Nejad and A. Haque, "Weakly-Supervised Anomaly Detection in Surveillance Videos Based on Two-Stream I3D Convolution Network," *arXiv preprint arXiv:2404.15289*, 2024.
15. A. Phapale and S. Bhingarkar, "Deep Context-Aware Feature Extraction for Anomaly Detection in Surveillance Videos," *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 2, pp. 21633–21638, 2025.