e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:08/Issue:11/November-2026 Impact Factor- 8.187 www.irjmets.com

SURAKSHA DRISHTI AN AI-POWERED INTELLIGENT VIDEO SURVEILLANCE SYSTEM FOR REAL-TIME THREAT DETECTION AND EMERGENCY RESPONSE

Ms. Aliza Raza Rizvi*1, Mohak Pandey*2, Krishnam Purwar*3, Suraj Shukla*4

*1Assistant Professor, Department Of Computer Science And Engineering, Babu Banarasi Das Northern India Institute Of Technology, Lucknow, India.

*2,3,4Student, Department of Computer Science and Engineering, Babu Banarasi Das Northern India Institute of Technology, Lucknow, India.

DOI: https://www.doi.org/10.56726/IRJMETS104975

ABSTRACT

Technology has become an essential part of our everyday lives, continuously transforming the way we work, communicate, and solve problems. With the rapid growth of Artificial Intelligence, Machine Learning, and Data Science, new opportunities are emerging to make processes more efficient, improve decision-making, and address real-world challenges. However, technology is not just about innovation; it is about using knowledge and creativity to develop solutions that make a meaningful difference. Continuous learning, practical experience, and adaptability are key to keeping pace with these changes and contributing effectively to the evolving digital world.

Keywords: Video Surveillance, Threat Detection, Deep Learning, YOLO11, Pose Estimation, LSTM, Action Recognition, Crowd Monitoring.

I. INTRODUCTION

Video surveillance systems have become an essential part of modern security infrastructure. CCTV cameras are widely installed in educational institutions, hospitals, railway stations, airports, shopping malls, banks and other public places. However, conventional surveillance systems primarily record video footage and depend on human operators to continuously monitor multiple screens. This creates difficulties in identifying dangerous activities at an early stage, particularly when several cameras are operating simultaneously.

With advancements in Artificial Intelligence (AI), Machine Learning (ML) and Computer Vision, intelligent surveillance systems can automatically analyse video footage and identify suspicious activities. Deep learning techniques have made it possible to detect objects, recognise human actions and analyse movement patterns from video sequences. Suraksha Drishti AI is a proposed intelligent surveillance framework designed to improve conventional CCTV monitoring by combining early threat prediction with real-time incident detection. The system aims to identify potentially dangerous behavioural patterns before an incident escalates and recognise ongoing events such as fighting, theft-related activities, weapon presence and accidental falls.

The proposed framework integrates multiple deep learning components, including YOLO11 for object detection, YOLO11-Pose for human posture estimation, Long Short-Term Memory (LSTM) networks for temporal behaviour analysis, and 3D Convolutional Neural Networks for video action recognition. Crowd analysis and tracking mechanisms will also be incorporated to monitor unusual crowd activity.

The primary objective of this research is to propose an integrated AI-based surveillance framework that can support faster security responses, reduce dependence on continuous manual monitoring and improve public safety.

II. LITERATURE REVIEW

Existing research in intelligent video surveillance has explored several approaches to object detection, human activity recognition and abnormal event identification. These approaches provide the foundation for developing automated surveillance systems.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:08/Issue:11/November-2026 Impact Factor- 8.187 www.irjmets.com

Object detection techniques such as the You Only Look Once (YOLO) family have significantly improved the ability to identify and locate multiple objects in images and video frames. YOLO11 supports object detection and human pose estimation, making it a suitable candidate for identifying people and potentially dangerous objects in surveillance footage.

Human activity recognition has also received considerable attention in computer vision research. Convolutional Neural Networks (CNNs) are commonly used to extract visual features, while recurrent neural networks are useful for analyzing sequential information. Long Short-Term Memory networks, introduced by Hochreiter and Schmidhuber, are designed to learn temporal dependencies in sequential data. These capabilities make LSTM networks relevant to analyzing human movement patterns over time.

Research by Sultani, Chen and Shah (2018) introduced a large-scale real-world surveillance anomaly detection dataset containing activities such as fighting, robbery, burglary and accidents. Their work demonstrated the importance of learning abnormal patterns from surveillance videos and highlighted the challenges associated with real-world anomaly detection.

Furthermore, three-dimensional convolutional neural networks have been investigated for video action recognition because they can extract spatial and temporal information from consecutive video frames. These techniques can support the classification of activities that cannot be reliably identified from individual images.

Although existing studies have addressed individual aspects of intelligent surveillance, integrating object detection, human posture estimation, temporal behaviour analysis, action recognition and crowd monitoring into a unified framework remains an important area for investigation. Suraksha Drishti AI proposes to explore this integration with a particular focus on both early risk assessment and real-time incident detection. Table 1 summarises the limitations of existing approaches and the corresponding improvements proposed in Suraksha Drishti AI.

Table 1. Comparison of Existing Approaches and Proposed Suraksha Drishti AI Improvements

Existing Approach	Limitation	Proposed Suraksha Drishti AI Improvement
Traditional CCTV	Relies on continuous human monitoring, which is limited by natural cognitive fatigue. Research indicates that an operator's ability to detect critical events drops by over 45% after just 20 minutes of uninterrupted viewing, leading to reactive rather than proactive security responses.	Replaces manual continuous viewing with automated, always-on AI detection branches that only push notifications to the Security Monitoring Dashboard when a categorized risk event occurs.
Human Detection	Standard models successfully identify human presence but lack the contextual awareness to distinguish between normal activity and potentially dangerous behavior or criminal intent.	Combines human bounding box localization with a dedicated Temporal Behavior Analysis layer to track consecutive movement patterns, event duration, and abnormal activity sequences.

III. PROBLEM STATEMENT

Conventional CCTV surveillance systems primarily function as video recording platforms and require continuous human observation. Monitoring multiple cameras simultaneously can result in delayed identification of suspicious activities and emergency situations.

Existing computer vision approaches often focus on individual tasks, such as detecting objects, recognising human actions or identifying abnormal events. However, identifying an object or recognising an ongoing incident alone does not necessarily provide sufficient information about potential threats before they escalate.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:08/Issue:11/November-2026 Impact Factor- 8.187 www.irjmets.com

Another challenge is analysing human movement and behavioural patterns over time to distinguish normal activities from potentially dangerous situations. Crowd congestion, sudden falls, aggressive movements and weapon presence also require different types of visual analysis.

Therefore, there is a need for an integrated surveillance framework capable of combining visual object detection, human posture estimation, temporal behaviour analysis, activity recognition and crowd monitoring. The proposed system aims to address these challenges by providing early risk assessment and automated detection of ongoing incidents through a unified AI-based approach.

IV. PROPOSED APPROACH

The proposed Suraksha Drishti AI framework will analyse live CCTV footage using multiple deep learning models and computer vision techniques. The system will process incoming video frames and extract information about people, objects, body posture, movement patterns and surrounding activities.

YOLO11 will be investigated for detecting individuals and potentially dangerous objects, including weapons. YOLO11-Pose will extract human body keypoints to support posture and movement analysis. An LSTM-based temporal behaviour model will analyse sequential movement information to identify patterns associated with potentially dangerous situations.

For activity recognition, an R3D-18-based video classification model will be investigated to recognise activities such as fighting, falling and other abnormal events. Crowd analysis will use person detection and multi-object tracking techniques to estimate crowd density and identify unusual changes in crowd movement.

The outputs of these components will be integrated through a risk assessment and event classification mechanism. The system will maintain separate capabilities for early risk prediction and real-time incident detection. When a relevant risk or incident is identified, the proposed platform will generate an alert for security personnel.

The notification system is planned to include visual alerts, alarm activation, SMS notifications, location information where available and the storage of relevant video clips. The overall framework will be developed using Python, OpenCV and PyTorch, with Firebase and Twilio considered for notification services. The architecture of the proposed framework is shown in Fig. 1.

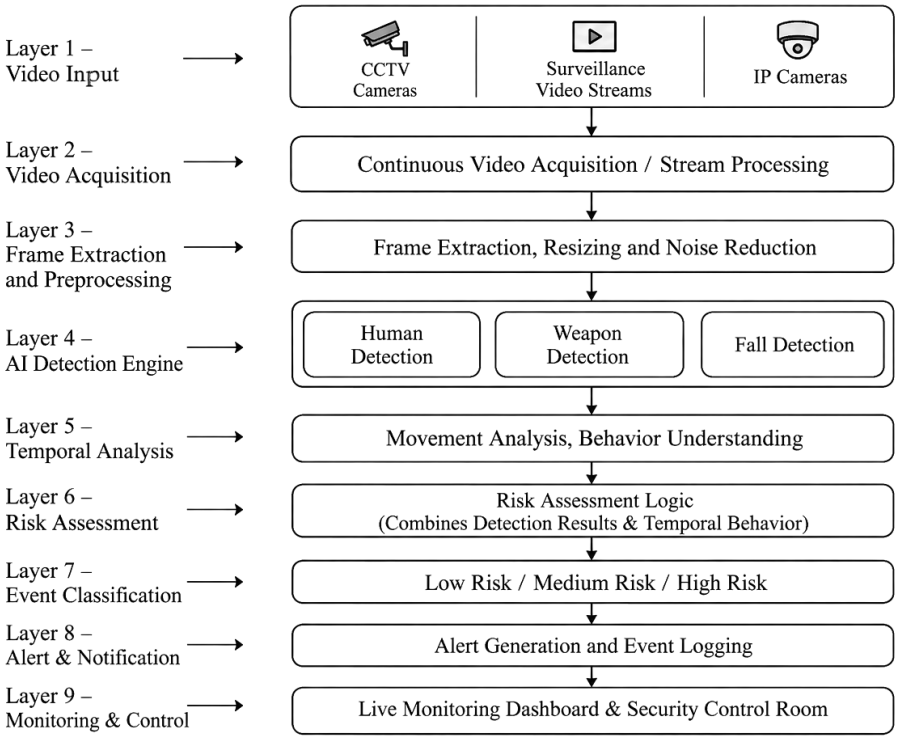

Figure 1: Overall Architecture of Suraksha Drishti AI

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:08/Issue:11/November-2026 Impact Factor- 8.187 www.irjmets.com

V. METHODOLOGY

The proposed methodology consists of video acquisition, preprocessing, object detection, pose estimation, temporal behaviour analysis, action recognition, crowd analysis, risk assessment and emergency notification.

Initially, CCTV cameras will provide continuous video input. OpenCV will be used to capture video frames, resize images and perform the required preprocessing operations. The processed frames will then be passed to the object detection module.

The YOLO11 architecture will be investigated for identifying people and dangerous objects. Task-specific training or fine-tuning will be required to recognise weapon categories accurately. The YOLO11-Pose model will extract human body keypoints, including shoulders, elbows, wrists, hips, knees and ankles. These keypoints will provide useful information about body posture and movement.

Next, sequential movement features will be provided to an LSTM-based temporal behaviour model. The model will analyse changes in body positions, movement speed, acceleration and interactions between individuals. Its objective will be to classify behavioural patterns and estimate potential risk levels. Reliable early prediction will require appropriately labelled pre-incident video sequences.

For recognising ongoing activities, an R3D-18-based video action recognition model will be investigated. It will process short video sequences to classify activities such as fighting, falling and normal human interactions. Theft-related activity recognition will also require suitable labelled video data.

Crowd monitoring will be implemented using YOLO11-based person detection and ByteTrack-based object tracking. This component will estimate the number of people in monitored areas and analyse changes in crowd density and movement.

The outputs from the different components will be combined using a risk assessment mechanism. Based on the classification results, the system will distinguish between normal activity, potentially dangerous behaviour and ongoing incidents. Appropriate alerts will then be generated through the proposed notification system.

Public datasets, including UCF-Crime and other suitable object detection, pose estimation and action recognition datasets, will be investigated for model development and evaluation. The models will be evaluated using precision, recall, F1-score, mean Average Precision where applicable, false-alarm rate and processing latency. Separate training, validation and testing data will be maintained to support reliable evaluation.

The proposed methodology will be implemented in phases, followed by integration and testing under different surveillance scenarios. Actual performance results will be reported after experimental validation.

VI. CONCLUSION

Suraksha Drishti AI proposes an integrated intelligent surveillance framework designed to improve the capabilities of conventional CCTV systems. By combining object detection, human pose estimation, temporal behaviour analysis, video action recognition and crowd monitoring, the proposed system aims to support both early threat prediction and real-time incident detection.

The integration of YOLO11, YOLO11-Pose, LSTM and R3D-18-based architectures provides a foundation for analysing visual information and human activities over time. The proposed notification mechanism will further support faster communication of potential security threats to responsible personnel.

The framework is intended to reduce dependence on continuous manual monitoring and improve the efficiency of surveillance operations in public and private environments. Future work will focus on model development, dataset preparation, experimental evaluation, performance optimisation and real-world deployment. The effectiveness of early threat prediction and incident detection will need to be established through systematic testing. The planned phases of this future work are outlined in Fig. 2.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:08/Issue:11/November-2026 Impact Factor- 8.187 www.irjmets.com

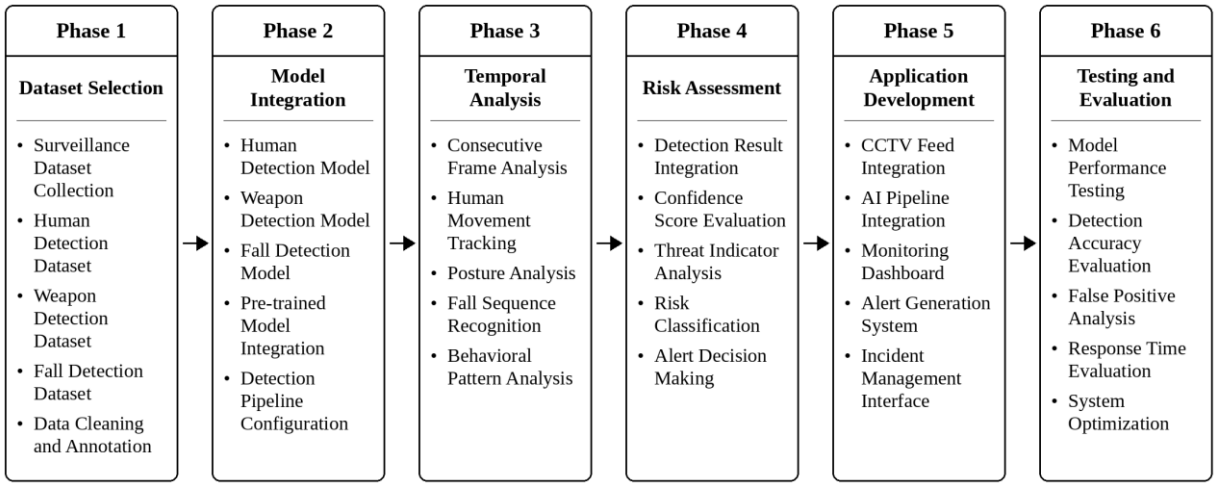

Figure 2: Future Implementation Roadmap of Suraksha Drishti AI

VII. REFERENCES

[1] Jocher, G., and Qiu, J. (2024). Ultralytics YOLO11. Ultralytics Documentation. https://docs.ultralytics.com/models/yolo11/

[2] Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[3] Sultani, W., Chen, C., and Shah, M. (2018). Real-World Anomaly Detection in Surveillance Videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6479–6488. https://doi.org/10.1109/CVPR.2018.00678

[4] Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., and Paluri, M. (2018). A Closer Look at Spatiotemporal Convolutions for Action Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[5] Lin, T. Y., et al. (2014). Microsoft COCO: Common Objects in Context. European Conference on Computer Vision (ECCV), 740–755.

[6] Zhang, Y., et al. (2022). ByteTrack: Multi-Object Tracking by Associating Every Detection Box. European Conference on Computer Vision (ECCV).