International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 14, Issue 1, 2026

Intelligent Video Surveillance

Y. Srinija[1], Dr.Md Asif[2]

[1]B.Tech Student, Department of Electronics and Computer Engineering, J.B. Institute of Engineering and Technology, Hyderabad, India.

[2]Associate Professor, Department of Electronics and Computer Engineering, J.B. Institute of Engineering and Technology, Hyderabad, India

asif.ecm@jbiet.edu.in

Abstract

The rapid growth of large-scale surveillance infrastructures has created an urgent requirement for automated techniques capable of analysing continuous video streams in real time. Conventional surveillance systems depend almost entirely on prolonged human observation, which is inherently inefficient and vulnerable to fatigue, delayed response, and missed critical incidents. This paper presents an intelligent video surveillance framework based on an unsupervised deep learning approach for detecting abnormal activities in video streams. The proposed system employs a Convolutional Long Short-Term Memory (Conv-LSTM) autoencoder to model normal spatio-temporal patterns in surveillance videos and to identify deviations through reconstruction error analysis. The network is trained exclusively on normal activity sequences, enabling the detection of unforeseen abnormal events without requiring explicit anomaly labels. The complete framework includes a real-time preprocessing pipeline, sequence buffering mechanism, threshold-based decision logic, and visual alert generation. Experimental evaluation conducted on a standard benchmark dataset demonstrates that the system can effectively identify abnormal behaviours such as running and object throwing while maintaining a low false-alarm rate. The results confirm the suitability of the proposed framework for practical real-time intelligent surveillance applications.

Keywords: *Intelligent video surveillance, anomaly detection, Conv-LSTM, autoencoder, unsupervised learning, spatio-temporal modelling.*

1. Introduction

The increasing deployment of cameras in public spaces, transportation hubs, educational campuses and industrial environments has led to a dramatic rise in the volume of video data generated every day. Monitoring such large volumes of video content using traditional Closed-Circuit Television (CCTV) systems remains a labour-intensive and error-prone process. Although conventional CCTV infrastructures are effective for post-incident investigation and evidence collection, they are not designed to proactively prevent incidents or assist security personnel in real time.

Human operators are required to observe multiple camera feeds for long durations. Several studies have shown that continuous visual monitoring significantly reduces operator vigilance after short periods of time, resulting in delayed reactions and overlooked events. Furthermore, as the scale of surveillance networks grows, the cost of employing sufficient personnel becomes unsustainable.

Recent advances in artificial intelligence, computer vision and deep learning have enabled the development of intelligent video surveillance systems capable of automatically interpreting visual information. These systems aim to shift surveillance from a reactive paradigm to a proactive and predictive one by detecting suspicious or abnormal activities as they occur. Modern video analytics platforms can automatically identify objects, track individuals and recognise behaviour patterns in complex environments.

Despite these advancements, anomaly detection remains a particularly challenging task. Unlike object recognition or action classification, abnormal events are inherently rare, context-dependent and often poorly defined. Collecting comprehensive labelled datasets containing all possible abnormal activities is unrealistic in real-world environments.

This work proposes an intelligent anomaly detection framework that learns normal motion and appearance patterns from video data and automatically identifies unusual events without relying on explicit anomaly labels. The system is designed to operate in real time and can be integrated with modern cloud- and edge-based infrastructures, making it suitable for large-scale and distributed deployment scenarios.

2. Problem Definition

Conventional surveillance infrastructures suffer from several fundamental limitations:

- Continuous manual monitoring is cognitively demanding and highly susceptible to fatigue and attention loss.
- Operational costs increase significantly as the number of deployed cameras grows.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

- A substantial portion of recorded video is never reviewed.
- Incident response is usually delayed, as abnormal events are often discovered after they have already occurred.

The central challenge is to design a surveillance system capable of automatically distinguishing between normal and abnormal activities without constant human supervision. Furthermore, due to the rarity and unpredictability of abnormal events, the system should not rely on exhaustive manual annotation of anomaly categories.

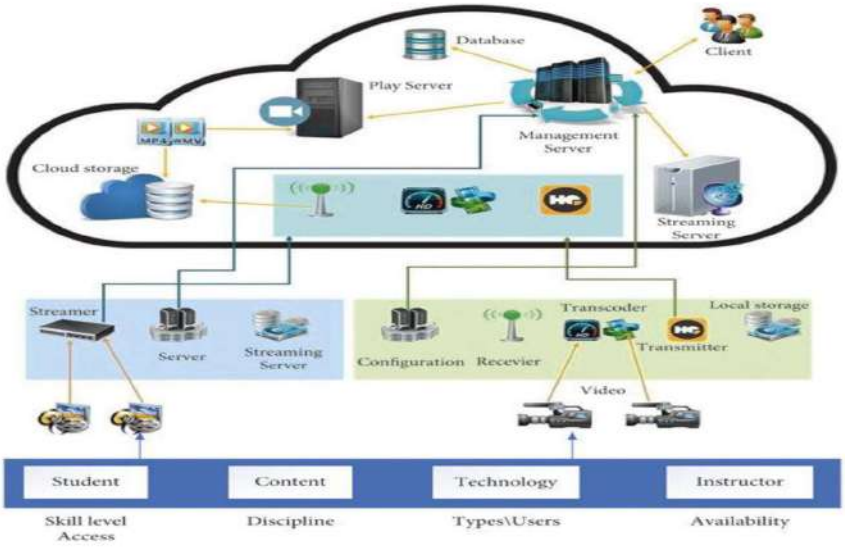
An effective anomaly detection system must therefore:

1. learn normal spatio-temporal patterns directly from video data,
2. generalise to previously unseen abnormal behaviours,
3. operate with low latency for real-time monitoring, and
4. maintain robustness under varying lighting, crowd density and background conditions.

.

3. Related Work

Early approaches to video surveillance relied primarily on background subtraction, optical flow analysis and handcrafted motion descriptors. While computationally efficient, such techniques are highly sensitive to illumination changes, camera jitter, shadows and environmental noise, and perform poorly in crowded or cluttered scenes



With the rise of deep learning, convolutional neural networks have become the dominant paradigm for visual feature extraction. Supervised learning approaches have demonstrated impressive performance in action recognition and behaviour classification. However, their reliance on large labelled datasets limits their applicability in anomaly detection, where abnormal events are rare and highly diverse.

Unsupervised and self-supervised learning techniques have therefore gained considerable attention. Autoencoder-based architectures learn compact latent representations by reconstructing input data. When trained exclusively on normal data, reconstruction errors can be exploited as an anomaly score.

Several studies have extended conventional autoencoders to video analysis by incorporating temporal modelling. Recurrent neural networks and long short-term memory units have been used to capture motion dynamics and temporal regularity in video sequences. Hybrid architectures combining convolutional layers with recurrent units have shown strong performance in modelling complex spatio-temporal patterns.

More recently, adversarial learning and predictive modelling frameworks have also been explored for video anomaly detection. However, many of these approaches introduce increased training complexity and instability.

In this work, a Conv-LSTM autoencoder architecture is adopted due to its ability to jointly model spatial appearance and temporal evolution within a unified network structure.

4. Proposed Methodology

4.1 System Overview

The proposed system follows a two-stage processing architecture:

- **Offline training stage**, in which the deep learning model is trained using only normal video sequences.

- **Online detection stage**, in which the trained model is deployed to analyse incoming video streams and detect abnormal events in real time.
  A sliding window of consecutive frames is used to preserve temporal continuity, allowing the model to analyse short video clips rather than isolated frames. This design ensures that both motion patterns and appearance variations are captured.

4.2 Data Preparation

Video frames are extracted from training videos at a fixed sampling rate. Each frame is resized to a uniform spatial resolution and converted to grayscale in order to reduce computational complexity and to focus on structural and motion-related information rather than colour variations.
Pixel intensities are normalised to a fixed range to stabilise network training and to avoid scale-related bias. The processed frames are grouped into fixed-length sequences of ten frames, forming spatio-temporal input samples.
This preprocessing pipeline ensures consistency between the training and inference stages and reduces the influence of irrelevant variations such as minor illumination changes.

4.3 Conv-LSTM Autoencoder Architecture

The core of the proposed system is a convolutional LSTM autoencoder. The encoder component consists of three-dimensional convolutional layers followed by Conv-LSTM layers. The three-dimensional convolutional layers extract spatial features while preserving temporal structure, and the Conv-LSTM layers model temporal dependencies across successive frames.
This architecture enables the network to simultaneously learn:

- spatial representations of objects and scene structure, and
- temporal relationships associated with motion patterns and behavioural dynamics.

The decoder reconstructs the original input sequence using transposed convolutional layers. During training, the network minimises the mean squared error between the input sequence and its reconstructed output.
Let X denote the input sequence and $\hat{X}$ the reconstructed sequence. The reconstruction loss is defined as

$$L = \frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{X}_i)^2$$

where N denotes the total number of pixels in the sequence.

4.4 Anomaly Detection Strategy

The model is trained exclusively with normal video data. As a result, it becomes highly specialised in reconstructing regular motion patterns and visual structures.
During deployment, each incoming video sequence is reconstructed by the trained model. Sequences that differ significantly from learned normal patterns produce higher reconstruction errors.
An anomaly score is computed using the reconstruction loss. If this score exceeds a predefined threshold, the corresponding video segment is classified as abnormal and an alert is generated.
This reconstruction-based formulation allows the system to detect previously unseen abnormal events without requiring explicit modelling of each anomaly category.

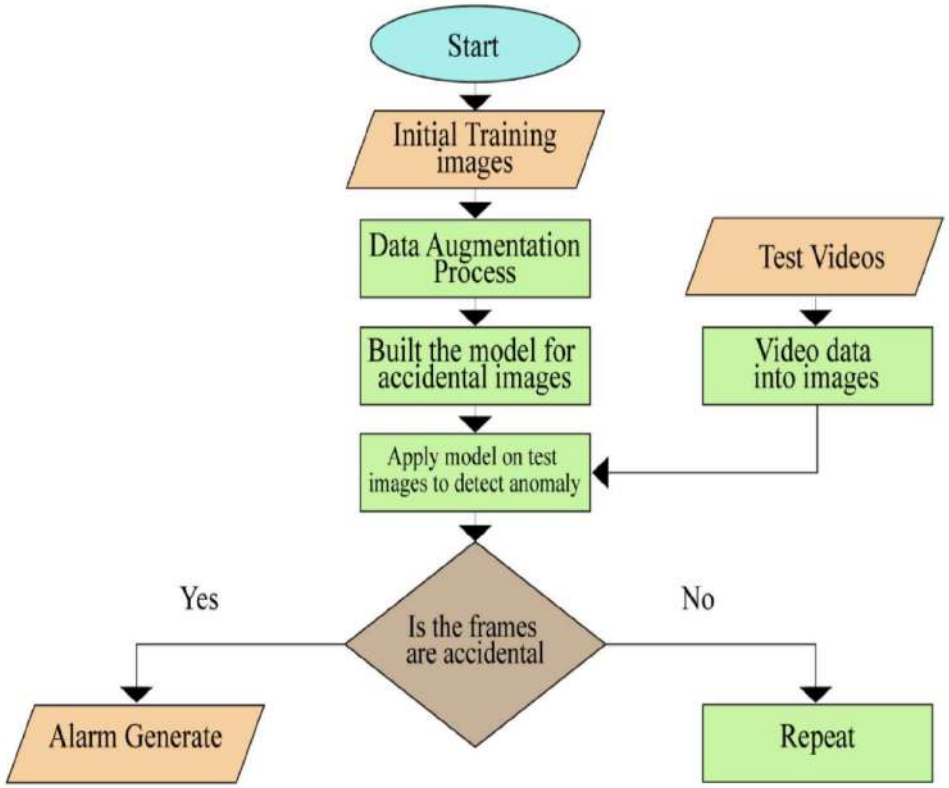
4.5 Threshold Selection

The anomaly threshold is determined empirically using a validation set containing both normal and abnormal sequences. The threshold is selected to balance detection sensitivity and false-alarm rate.
In the reported experiments, a threshold value of **0.00068** provided a suitable trade-off between missed detections and spurious alarms. The threshold was chosen by analysing the distribution of reconstruction errors and selecting a value that maximised separation between normal and abnormal samples.

4.6 Computational Complexity and Real-Time Suitability

The computational complexity of the proposed framework is dominated by the convolutional and Conv-LSTM layers during inference. To enable real-time operation, the system processes short frame sequences and employs lightweight preprocessing operations.
A sliding buffer mechanism is used to avoid redundant recomputation and to maintain continuous processing of incoming frames. The inference pipeline can be further accelerated using GPU hardware or optimised deployment frameworks.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026



5. System Implementation

The system is implemented using Python and widely adopted deep learning and computer vision libraries. The training module performs dataset preparation, sequence generation, network construction and optimisation using the Adam optimiser with mean squared error loss.

Model checkpoints are stored during training to preserve the best performing network. Early stopping is employed to prevent overfitting and unnecessary training iterations.

The real-time detection module loads the trained model and processes incoming video streams using a sliding window of ten frames. Each window is reconstructed by the model and its reconstruction error is computed. When the error exceeds the threshold, a visual alert is overlaid on the output video stream.

The implementation incorporates:

- real-time frame buffering,
- consistent preprocessing,
- error handling mechanisms for corrupted or missing frames,
- frame-rate synchronisation, and
- logging facilities for offline analysis.

This modular design facilitates maintenance and future system extension.

6. Experimental Evaluation

6.1 Dataset and Experimental Setup

The system is evaluated using a public benchmark dataset designed for video anomaly detection. The training set contains only normal pedestrian activities, while the testing set includes a variety of staged abnormal events such as running, loitering and object throwing.

All training sequences consist exclusively of normal behaviour. No anomaly samples are used during training. Testing is performed on previously unseen video clips.

The model is trained for several epochs with a batch size of one sequence. Input frames are resized to a uniform resolution and normalised before being passed to the network.

6.2 Evaluation Protocol

The system is evaluated qualitatively and behaviourally by analysing detection responses across all testing videos. Reconstruction losses are recorded frame-sequence wise and compared against the selected threshold.

The evaluation focuses on:

- ability to detect prominent abnormal events,
- consistency of alert generation,
- robustness under varying crowd density, and
- stability during long video streams.

6.3 Results

The proposed system successfully detects the majority of prominent abnormal events in the test sequences. Sudden changes in motion patterns, such as abrupt running or object throwing, generate noticeably higher reconstruction errors and are consistently identified.
A small number of false alarms are observed, mainly in scenes with unusually dense crowds and complex interactions. These scenarios introduce motion patterns that are under-represented in the training data.
A limited number of subtle anomalies, such as slightly faster walking, remain difficult to detect because they do not strongly deviate from learned normal patterns.

6.4 Discussion

The experimental results demonstrate that reconstruction-based anomaly detection using a Conv-LSTM autoencoder is an effective strategy for surveillance video analysis.
The principal strength of the proposed approach lies in its independence from manually labelled anomaly categories. The system is capable of identifying previously unseen abnormal behaviours by learning only from normal data.
However, the definition of normality is entirely driven by the training distribution. Environments exhibiting highly dynamic or unpredictable behaviour patterns may require more diverse training data to achieve stable performance.

7. System Testing and Validation

The software implementation is validated through unit, integration, functional and scenario-based testing.
Individual preprocessing and loss calculation modules are tested independently. Integration testing confirms compatibility between the training and detection modules.
End-to-end testing verifies correct alert generation during abnormal events and stable behaviour during prolonged video processing. Additional experiments involving low-light conditions and video noise confirm reasonable robustness under moderate environmental variations.
Stress testing with long video sequences demonstrates that the system maintains consistent memory usage and inference stability.

8. Limitations

Despite its effectiveness, the proposed system exhibits several limitations. The learned notion of normality is strongly dependent on the training environment. Deploying the model in a different setting generally requires retraining.
The use of a fixed anomaly threshold restricts adaptability to gradual environmental changes such as lighting transitions or seasonal variations. Moreover, the system currently provides only binary anomaly decisions without localising the abnormal region or classifying the event type.
The framework also does not explicitly address adversarial perturbations or deliberate camera obstructions.

9. Conclusion

This paper presented an intelligent video surveillance framework for unsupervised anomaly detection based on a Conv-LSTM autoencoder. By learning spatio-temporal representations of normal activities, the proposed system is capable of identifying abnormal events in real time without requiring labelled anomaly samples.
Experimental evaluation demonstrates reliable detection performance and a low false-alarm rate on benchmark surveillance videos. The proposed approach provides a practical and scalable foundation for proactive surveillance systems capable of reducing human monitoring workload and improving situational awareness.

10. Future Work

Future extensions of this work will focus on several directions.
First, more advanced spatio-temporal architectures such as three-dimensional convolutional networks and transformer-based models can be explored to improve representation capacity and long-range temporal modelling.
Second, dynamic thresholding strategies based on online statistical analysis of reconstruction error can be introduced to adapt to changing environmental conditions.
Third, a secondary classification and localisation stage can be integrated to identify the type and spatial location of detected anomalies, enabling richer situational awareness and decision support.
Finally, optimised deployment on edge devices combined with cloud-based monitoring and management platforms would enable large-scale, low-latency and privacy-aware intelligent surveillance infrastructures.

References

[1] Y. S. Chong and M. S. Ryoo, “Abnormal event detection in videos using spatiotemporal autoencoder,” *Proc. ICCV Workshops*, 2017.
[2] X. Shi et al., “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,” *Advances in Neural Information Processing Systems*, 2015.
[3] M. Hasan et al., “Learning temporal regularity in

ISSN 2347–3657
Volume 14, Issue 1, 2026

video sequences," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[4] B. R. Kiran, D. M. Thomas, and R. Parakkal, "An overview of deep learning-based methods for video anomaly detection," *International Journal of Computer Science and Engineering*, 2018.
[5] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, 2012.
[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, 2015.
[7] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 201