



A CNN–LSTM Framework for Spatiotemporal Detection of Abnormal Human Events in Forensic Surveillance Videos

Ravi Kumar Veerappaodeyara Hundi Guruswamy^{1,2*}, Pradeep Sannaswamy¹

¹ Department of Computer Science, Government Engineering College, Chamarajanagara 571313 (Affiliated to Visvesvaraya Technological University, Belagavi 590018), India

² PG Department of Computer Applications, JSS College of Arts, Commerce and Science, Mysuru 570025, India

Corresponding Author Email: vg.ravi@yahoo.com

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.130502>

ABSTRACT

Received: 14 April 2026

Revised: 30 May 2026

Accepted: 9 June 2026

Available online: 15 June 2026

Keywords:

video forensics, abnormal human event detection, surveillance video, CNN–LSTM, spatiotemporal representation, anomaly scoring, deep learning

Abnormal event detection in surveillance video is a critical task in video forensics, where reliable identification of violence, intrusion, theft, crowd disorder, and unusual human motion can support post-event investigation and evidence review. Many conventional approaches, including Support Vector Machine (SVM), Hidden Markov Model (HMM), and Principal Component Analysis (PCA), rely on hand-crafted representations or short-range temporal assumptions, which restrict their performance under scene clutter, illumination changes, noise, and irregular motion patterns. This study proposes a Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework (CL-ADF) for abnormal human event recognition in forensic surveillance video. The framework first standardizes video frames through preprocessing and normalization, then uses Convolutional Neural Network (CNN) layers to encode spatial appearance and motion-related visual cues. The extracted frame-level representations are passed to Long Short-Term Memory (LSTM) units to model temporal dependencies across consecutive frames. An anomaly score is derived from reconstruction error in the learned feature space, and a threshold-based decision rule is used to distinguish normal and abnormal events. Comparative experiments against SVM-, HMM-, and PCA-based baselines indicate that CL-ADF provides more stable detection performance in terms of accuracy, precision, recall, false-alarm rate, and F1-score. The results suggest that combining CNN-based spatial encoding with LSTM-based temporal modelling is suitable for forensic-oriented abnormal event detection in complex surveillance scenes. The proposed framework can support automated screening of surveillance footage while preserving event-level traceability through anomaly scoring.

1. INTRODUCTION

Abnormal human activity detection is becoming a fundamental component of video forensics with the aim of automatically identifying irregular or unusual activity out of surveillance videos [1]. Because CCTV networks, body-worn surveillance cameras, and aerial surveillance have been increasing dramatically, the scale of video data recorded is now orders of magnitude greater than is possible for human examination. Violence, stealing, crowd surging, or unauthorized intrusion are some examples of abnormal human events typically lost amidst huge amounts of normal activity, and automated detection is now very important for forensic examination and civilian safety [2].

Classical techniques provided the foundation for anomaly detection but had several limitations. Support Vector Machine (SVM)-based methods exhibited weak generalizability in busy and complex scenes. Hidden Markov Model (HMM)-based systems could detect sequential motion patterns but failed to stay efficient with longer time dependency modeling [3].

Dimension reduction was extensively used with Principal Component Analysis (PCA), but was highly noise-sensitive and powerless in modeling non-linearity. These detriments restricted the application of classical techniques for large-scale and real-time forensic video analysis [4].

Recent advances in deep learning have addressed several of these limitations through spatiotemporal modelling. Convolutional Neural Networks (CNNs) are most appropriate for extracting spatial features from video frames, while Long Short-Term Memory (LSTM) networks can retain long-term memory and capture sequential temporal dynamics [5]. Hybrid CNN–LSTM networks now dominate anomaly detection studies with robust identification of strange events in hard video backdrops. Additionally, Autoencoders, Generative Adversarial Networks (GANs), and Transformer-based networks are gaining favor for semi-supervised and unsupervised learning with far less extensive labeled data [6].

These methods are used in multiple applications: forensic analysis for reconstructing a crime scene, mass surveillance in subway stations and airports, medical surveillance for the

detection of falls, and intelligent traffic control. Using deep learning, advanced video forensics enables accurate, scalable, and efficient detection of exceptional human events [7].

1.1 Research gaps

Despite significant advances in the detection of unusual events, research gaps remain in video forensics. Standard models, including SVM, HMM, and PCA, are disadvantaged in handling sophisticated and noise-polluted scenes, while paradigms of deep learning often demand large labeled sets, a resource that is scarce in forensic scenarios. Current CNN–LSTM architectures, even though successful, remain troubled with scalability as applied in large networks of surveillance and end up inducing computational overhead. Most of the current techniques are also troubled with robustness as regards subtle and rare anomaly detections, specifically in crowded and occluded scenes. Moreover, the challenge of suppressing false positives while maintaining reliable detection remains unresolved and therefore hampers practical deployment in forensics and courts of law. There is also a need for lightweight and real-time techniques that are integration-friendly with IoT and cloud-based forensic platforms. Addressing these gaps has the potential to enhance the reliability, efficiency, and forensic value of unusual event detection systems [8].

1.2 Related work

Recent research on video anomaly detection has mainly focused on the detection of abnormal human events from surveillance videos using machine learning, deep learning, multimodal sensing, and spatio-temporal feature learning approaches. The conventional methods, such as SVM, HMM, and PCA, have been widely used for classification, temporal modelling, and feature reduction. However, the performance of these methods is limited in complex surveillance environments due to the poor robustness against noise, occlusion, illumination changes, and long-term motion variations. To overcome these limitations, deep learning-based methods have been receiving attention as they can automatically extract spatial and temporal representations from video frames. CNN-based models are powerful in learning visual patterns from individual frames, and LSTM networks are suitable for modelling sequential changes in activity across video clips. Hence, CNN-LSTM-based frameworks are highly relevant for abnormal human event detection in video forensic analysis.

Abnormal event recognition in different surveillance and monitoring scenarios has been addressed by several recent works. Tsiktiris et al. [9] have proposed a multimodal abnormal event detection method for public transport that uses Red, Green, Blue (RGB), depth, and audio modalities. Chung et al. [10] proposed an event-camera-based abnormal driver motion prediction model based on convolutional spiking neural networks. Hu et al. [11] proposed a deep learning-based abnormal behaviour recognition method for aquaculture monitoring, and Huang et al. [12] proposed Temporal-Aware Contrastive Network (TAC-Net) for intelligent video surveillance based on temporal-aware contrastive learning. The studies show the increasing importance of deep learning for abnormal event detection; however, many existing models are application-specific, computationally expensive, or less suitable for forensic video analysis in real-time. Hence, the proposed Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework (CL-ADF) integrates

CNN-based spatial feature extraction and LSTM-based temporal dependency modelling for improving abnormal human event detection accuracy, reducing false alarms and enabling reliable video forensic investigation.

Chen et al. [13] investigated vibration-signal-based detection of irregular gait using Dynamic Time Warping (DTW), K-Nearest Neighbors (KNN), and HMM algorithms with a non-intrusive system without wearables and cameras. A key contribution of the study was its robustness across different sensor placements and locations. Its drawback is reduced precision in identifying subtle changes of gait, with possible influence in complex real-world settings.

Tay et al. [14] provided a comprehensive survey of recognizing outlying behavior in daily activity, including sensors and vision-based approaches. Novelty is mapping datasets and techniques available for smart home and elderly care. A reported limitation was the limited availability of large and heterogeneous datasets, which constrained generalisability. Ahmad Hamid et al. [15] put forward Abnormal Crowd Behavior Detection via Multi-Source Information Fusion Technique (ABD via MSIFAT), a crowd behavior anomaly detection framework from multiple sources based on multi-information fusion. The originality includes double pipelines with depthwise CNNs, LiteFlowNet, and Depthwise Separable Convolution with a Gated Recurrent Unit (DSC-GRU), and uncertain event detection using fuzzy logic. The drawback is increased system complexity, which may affect real-time efficiency in large-scale systems.

Xia et al. [16] developed the Smart City Security System (SCSS) with Domain-specific and You Only Look Once (DS-YOLO) for observing and recognizing unusual activity on the internet. The study contributed an active surveillance architecture with cloud-based reporting and GPS-enabled event localisation. It is low precision (approximately 90%), and it may be low for serious forensic applications.

Hong et al. [17] developed an inexpensive geocoding system, CoMiner, for geocoding behavior-aware coordinates from delivery events and later adapted it for identifying unusual delivery events. Innovation lies in combining textual, delivery, and trajectory data for location inference, with reduced abnormal deliveries up to 20.3%. Due to the restriction of the work to logistics, it cannot be directly applied to the general video forensic abnormal event detection.

2. VIDEO FORENSIC ANALYSIS FRAMEWORK

Figure 1 shows an integrated video forensic analysis framework covering video management, analysis, and semantic inference. The video management system captures and handles surveillance videos, and the processed videos are stored in the video database [18].

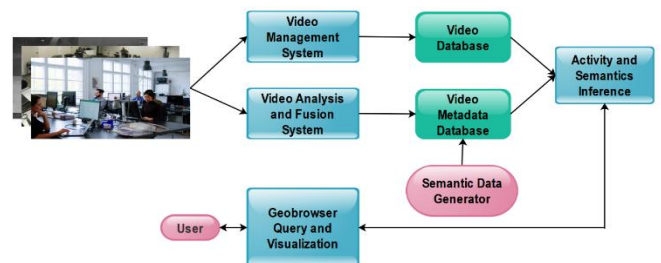


Figure 1. Flow of video management, analysis, and semantic inference system

Simultaneously, the video analysis and fusion system processes video inputs and extracts corresponding features, and the latter are stored in the video metadata database [19]. A semantic data generator converts such metadata into meaningful representations, and these are utilized in supporting activity recognition and semantic inference for recognizing unusual or surprising human events. These inferred activities are provided as output to the end user through a geo-browser query and visualization interface so that the end-user may conduct an interactive analysis, retrieval, and visualization of raw video and semantic data. The whole system enables fast and accurate video forensic investigation with the integration of raw data and semantic context so that unusual human events may be recognized faster and with better precision [20].

2.1 Spatial feature extraction

The CNN extracts spatial features from input frames, forming the basis of metadata generation in video forensic analysis [21], as expressed in Eq. (1).

$$F_{CNN} = \sigma(W_c * X + b_c) \quad (1)$$

where, F_{CNN} denotes the extracted spatial feature map, X represents the input video frame, W_c is the convolutional filter weight, b_c is the bias term, and σ refers to the activation function such as ReLU.

2.2 Temporal activity modeling

The LSTM network models temporal dependencies between frames enabled recognition of sequential abnormal events [22]. The temporal state transition is expressed in Eq. (2).

$$h_t = f(W_h h_{t-1} + W_x x_t + b) \quad (2)$$

where, h_t is the hidden state at time step t , h_{t-1} represents the previous hidden state, x_t is the input feature at the current time step, W_h and W_x are the weight matrices for hidden and input states respectively, b is the bias term, and f is the non-linear activation function such as tanh or sigmoid [23].

2.3 Anomaly score estimation

Eq. (3) is represented by the autoencoder, which reconstructs video features, and the reconstruction error is used as the anomaly score to detect suspicious events [24].

$$S(x) = \|x - \hat{x}\|^2 \quad (3)$$

where, $S(x)$ indicates the anomaly score, x represents the original input video feature, and $\hat{x}$ is the reconstructed feature produced by the autoencoder [25].

2.4 Objectives

The core of this framework is an integration of video management, metadata generation, and semantic reasoning for effective automated detection of unusual events in video forensics. Through the integration of data analysis and queryable visualization for users, the system obtains accurate, timely, and interpretable forensic knowledge [26-28].

- To preprocess surveillance video frames using resizing, normalization, and fixed-length clip formation for abnormal event detection.

- To extract spatial features from video frames using a CNN-based feature extraction module.

- To learn temporal activity patterns using LSTM and classify events as normal or abnormal using anomaly score and threshold-based decision-making.

- To evaluate the proposed CL-ADF framework using accuracy, precision, recall, false alarm rate, and F1-score, and compare it with SVM, HMM, and PCA.

3. PROPOSED METHODOLOGY OF CONVOLUTIONAL NEURAL NETWORK–LONG SHORT-TERM MEMORY BASED HUMAN ABNORMAL EVENT DETECTION

Figure 2 shows the five-stage methodology of the proposed CNN–LSTM-based abnormal event detection model. The process consists of video preprocessing, CNN-based spatial feature extraction, and LSTM-based temporal learning. Then, anomaly scoring is performed by reconstruction error, and the final decision is made by classifying the event as normal or abnormal according to a threshold value [29-31].

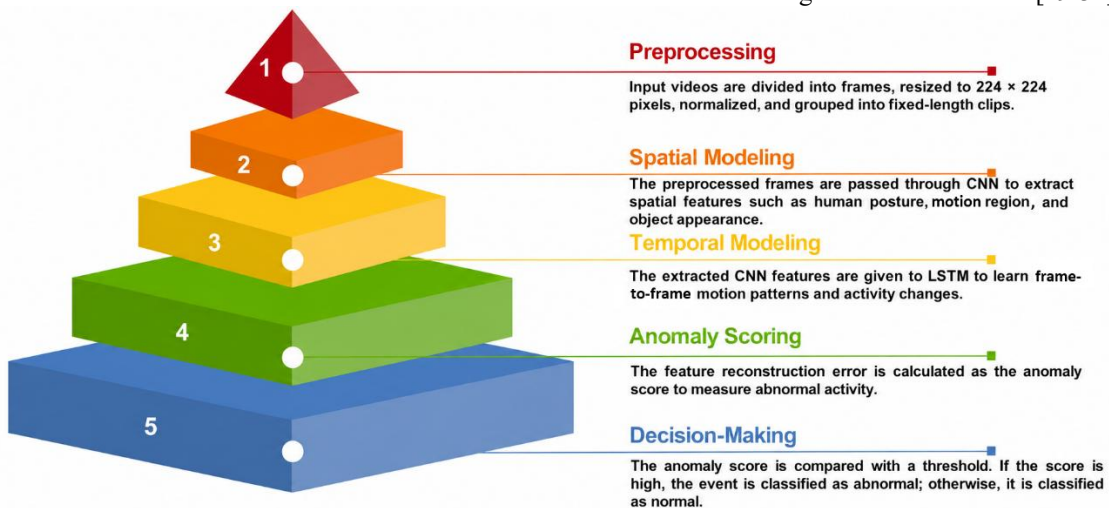


Figure 2. Layered pipeline of the proposed Convolutional Neural Network (CNN)–Long Short-Term Memory (LSTM) methodology for video forensic analysis

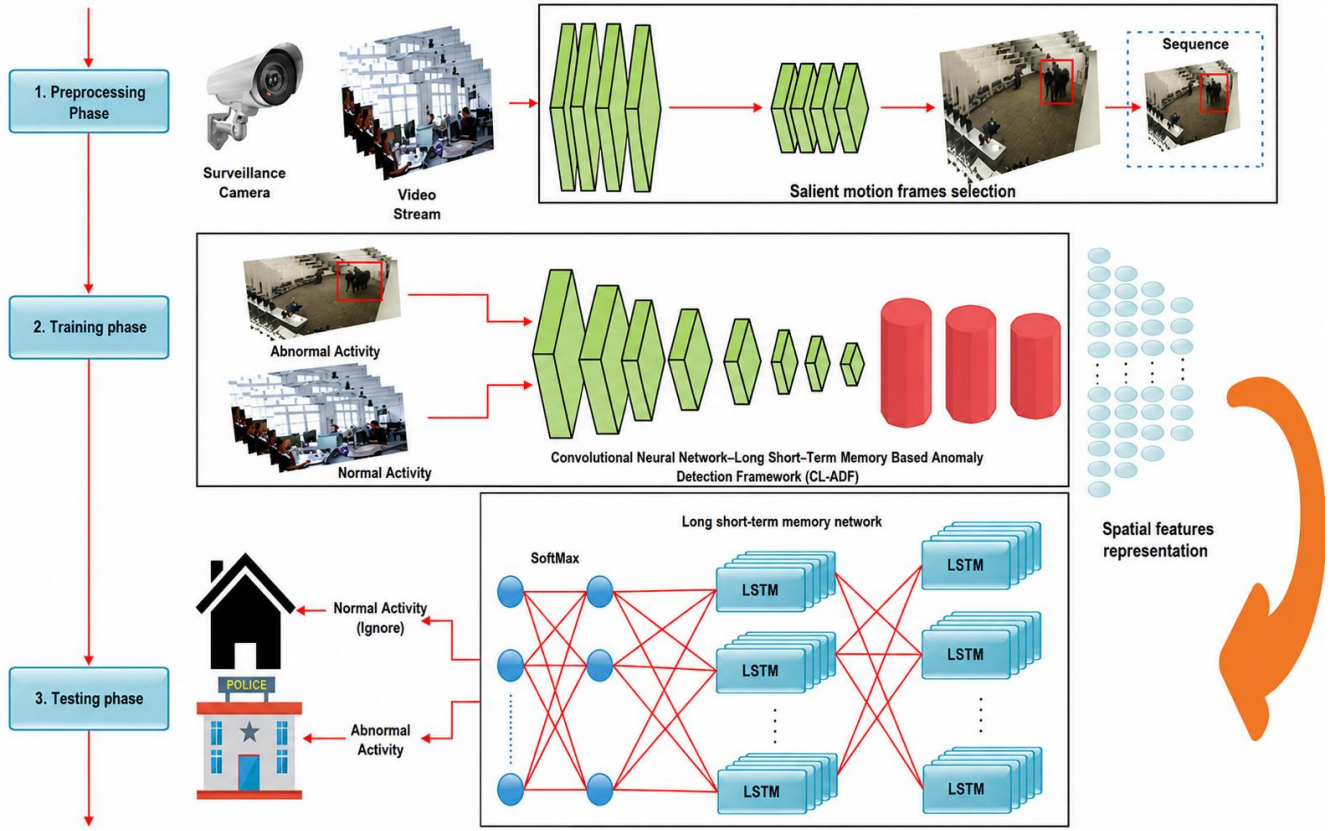


Figure 3. Architecture of the proposed Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework (CL-ADF) for human abnormal event recognition in video forensics

3.1 Proposed Convolutional Neural Network–Long Short-Term Memory based Anomaly Detection Framework

The proposed framework, CL-ADF, is targeted at enhancing abnormal human event detection in forensic video analysis, as shown in Figure 3. Unlike lightweight convolution techniques, with mostly retained spatial patterns, the CL-ADF integrates CNN modules for spatial feature representation with LSTM layers for long-term modeling of video frames' temporal dependencies [32, 33]. Preprocessed input surveillance videos are broken down into frames and then passed through convolutional layers for robust spatial feature representation. Sequential processing with LSTM units of the created features learns long-term temporal connections, and discrimination between normal and abnormal events is accurately achieved. Anomaly scores are calculated in a decision layer, and the events are classified and reflected in a semantic metadata database for retrieval and reporting of the forensic investigation. In integrating both spatial and temporal learning, the CL-ADF provides greater recognition precision, false alarm reduction, and enhancement of forensic investigation integrity over conventional lightweight CNN methods [34, 35].

3.2 Preprocessed frame normalization

Video frames are normalized to reduce illumination variations and provide a consistent input representation for analysis, as determined by Eq. (4).

$$X' = \frac{X - \mu}{\sigma} \quad (4)$$

where, X' represents the normalized video frame, X is the original input frame, μ denotes the mean pixel intensity, and σ is the standard deviation used for scaling [36, 37].

3.3 Dataset and preprocessing details

The proposed CL-ADF is validated on the Avenue and UCSD Ped2 video anomaly detection datasets. Avenue dataset contains 16 video sequences for training and 21 video sequences for testing. For the UCSD Ped2 dataset, we use 16 video clips for training and 12 video clips for testing. These values refer to the number of video sequences, not individual frames or clips.

Each input video was processed frame by frame at 25 fps. The mean and standard deviation-based scaling is applied after resizing each frame to 224×224 pixels. The normalized frames are grouped into clips of 16 consecutive frames, with a fixed length. These clips are the input to the CNN–LSTM framework that learns the spatial and temporal features.

3.4 Spatial feature extraction using convolutional layers

CNN layers extract spatial features from frames, preserving structural details important for forensic recognition, as represented by Eq. (5).

$$F_{CNN} = \sigma(W_c * X' + b_c) \quad (5)$$

where, F_{CNN} denotes the CNN feature map, W_c is the convolutional filter weight, X' is the normalized frame, b_c is the bias term, and σ is the non-linear activation function such as ReLU [38–40].

3.5 Temporal dependency modeling using Long Short-Term Memory

Sequential learning through LSTM captures temporal relationships across frames to identify abnormal event dynamics, and it is represented by Eq. (6).

$$h_t = f(W_h h_{t-1} + W_x F_{CNN,t} + b) \quad (6)$$

where, h_t indicates the hidden state at time t , h_{t-1} represents the previous hidden state, $F_{CNN,t}$ is the CNN feature at time t , W_h and W_x are the weight matrices for hidden and input states respectively, b is the bias, and f is the non-linear activation function used in the LSTM network [41, 42].

3.6 Anomaly score estimation using feature reconstruction

The reconstruction error between original and predicted features is computed to quantify abnormality, and it is expressed by Eq. (7).

$$S(x) = \|F_{CNN,t} - \hat{F}_{CNN,t}\|^2 \quad (7)$$

where, $S(x)$ represents the anomaly score, $F_{CNN,t}$ is the original feature, and $\hat{F}_{CNN,t}$ is the reconstructed feature produced by the anomaly detection framework [43, 44].

3.7 Decision function for abnormal event classification

A threshold-based decision function classifies events as normal or abnormal based on anomaly scores, and it is represented by Eq. (8).

$$y = \begin{cases} 1, & S(x) > \theta \text{ Abnormal Event} \\ 0, & S(x) \leq \theta \text{ Normal Event} \end{cases} \quad (8)$$

where, y denotes the classification output, $S(x)$ is the anomaly score, and θ is the threshold parameter used to differentiate abnormal events from normal activities [45, 46].

3.8 Overall Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework formulation for abnormal event recognition

The overall working of the proposed CL-ADF integrates preprocessing, spatial feature extraction, temporal modeling, and anomaly scoring into a single equation for forensic decision-making, and it is represented by Eq. (9).

$$y = \mathbb{I}(\|f_{LSTM}(F_{CNN}(X')) - \hat{f}_{LSTM}(F_{CNN}(X'))\|^2 > \theta) \quad (9)$$

where, y denotes the binary classification output (1 for abnormal, 0 for normal), X' represents the normalized video frame, $F_{CNN}(\cdot)$ is the CNN function extracting spatial features, $f_{LSTM}(\cdot)$ models the temporal sequence of features using LSTM, $\hat{f}_{LSTM}(\cdot)$ is the reconstructed sequence generated for anomaly scoring, $\|\cdot\|^2$ denotes the squared Euclidean distance used to compute reconstruction error, θ is the decision threshold, and $\mathbb{I}(\cdot)$ is the indicator function that outputs 1 if the condition is true (abnormal) and 0 otherwise (normal) [47-49].

3.9 Proposed Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework algorithm

CL-ADF algorithm, as presented in Algorithm 1, combines CNN and LSTM for identifying unusual human events in video forensics. Video frames are initially normalized and transformed into clips, whose spatial features are extracted using a CNN. These sequential features are modeled using LSTM to retain time dependencies between frames. An anomaly score is determined based on observed and predicted “normal” sequences using reconstruction error as the measure [50-55]. A decision threshold is used for binary classification of events as either normal or unusual. The algorithm is trained either in an unsupervised regime with the availability of only normal samples or in a supervised regime with the availability of both normal and unusual samples. CL-ADF combines spatial and time analysis and is capable of achieving very accurate results with low false alarms and efficient anomaly detection, and is hence appropriate for real-time forensic applications [56-59].

Algorithm 1. Proposed CL-ADF algorithm with reproducible preprocessing

Input: Video dataset D , video stream $V = X_1, X_2, \dots, X_N$, model weights W_c, W_x, W_h

Dataset Details: The Avenue dataset contains 16 training video sequences and 21 testing video sequences. UCSD Ped2 dataset contains 16 training video sequences and 12 testing video sequences. These values represent video sequences, not frames or clips.

Output: Event label $y \in \{0 = \text{normal}, 1 = \text{abnormal}\}$ and anomaly score S

Parameters: Input resolution 224×224 pixels, frame rate 25 fps, clip length $T = 16$, batch size 32, and threshold θ .

Step 1: Frame preprocessing

For each video sequence V , read the video frame by frame. Each frame X_t is resized to 224×224 pixels and normalized as:

$$X'_t = \frac{X_t - \mu}{\sigma}$$

The normalized frames are arranged in temporal order and grouped into fixed-length clips:

$$C_i = X'_1, X'_2, \dots, X'_T$$

Step 2: Spatial feature extraction using CNN

For each normalized frame X'_t in the clip, the CNN extracts spatial features as:

$$F_{CNN,t} = \sigma(W_c * X'_t + b_c)$$

Step 3: Temporal modeling using LSTM

The extracted CNN features are passed into the LSTM network:

$$h_t = f(W_h h_{t-1} + W_x F_{CNN,t} + b)$$

The temporal feature sequence is collected as:

$$H = [h_1, h_2, \dots, h_T]$$

Step 4: Anomaly score calculation

The predicted normal feature sequence is represented as $\hat{H}$. The anomaly score is calculated using reconstruction error:

$$S = \frac{1}{T} \sum_{t=1}^T |h_t - \hat{h}_t|^2$$

Step 5: Decision-making

The anomaly score S is compared with the decision threshold θ :

$$y = I(S > \theta)$$

The threshold is calibrated using validation normal samples:

$$\theta = \mu_S + k\sigma_S, k \in [2,3]$$

If $S > \theta$, the event is classified as abnormal. Otherwise, it is classified as normal.

Training notes: Train CNN+LSTM end-to-end on normal clips (unsupervised) or normal/abnormal (supervised). Train the predictor/autoencoder only on normal data to make reconstruction tight for normal and loose for abnormal [56]. X'_t denotes the normalized frame at time t , $F_{CNN,t}$ denotes the CNN feature at time t , h_t denotes the LSTM hidden state, $\hat{h}_t$ denotes its predicted (normal) counterpart, S denotes the anomaly score, and $I(\cdot)$ denotes the indicator function.

Complexity (per clip): $\mathcal{O}(T \cdot C \cdot K^2)$ for CNN (channels C , kernel K) + $\mathcal{O}(T \cdot H^2)$ for LSTM (hidden size H).

4. RESULTS AND DISCUSSION

The proposed CL-ADF can learn spatial and temporal features from surveillance video frames more effectively than the conventional methods. In each frame, CNN layers are used to learn the key visual patterns, and in the video clip, LSTM layers are used to learn the continuous motion and the change of activity. The proposed model achieved improved accuracy, precision, recall, F1-score, and false-alarm rate. The proposed CL-ADF shows better abnormal event detection capability, fewer false alarms, and more reliable decision-making for video forensic analysis in comparison with SVM, HMM, and PCA.

The experimental settings of the proposed CL-ADF are summarised in Table 1. It includes the selection of the dataset, pre-processing parameters, network settings, training, and decision threshold tuning so as to allow robust testing for video forensic abnormal event detection.

The proposed CL-ADF is evaluated by comparing its performance with conventional video anomaly detection methods, namely SVM, HMM, and PCA. These methods are chosen as baseline models, as they can be matched to the classification-based, temporal-sequence-based, and dimensionality-reduction-based methods widely used in previous abnormal event detection studies. For comparison, the same dataset partition, the same preprocessing settings, the same clip generation approach, and the same performance metrics are used to keep the comparison fair. In the related work section, we only discuss recent developments such as Multiple Instance Self-Training Framework (MIST), Robust

Temporal Feature Magnitude (RTFM), memory-augmented autoencoders, and transformer-based video anomaly detection models. Their full implementation settings, training protocols, and reproducible benchmark configurations were not used in the present experimental setup and therefore are not part of the quantitative experimental comparison. The experimental claims of this study are thus only related to comparison with SVM, HMM, and PCA.

Table 1. Experimental setup for Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework (CL-ADF)

SI. No	Parameter	Value
1	Datasets	Avenue (Train 16 / Test 21), UCSD Ped2 (Train 16 / Test 12)
2	Input Resolution	224 × 224 pixels
3	Frame Rate	25 fps
4	Clip Length (T)	16 frames
5	CNN Backbone	ResNet-18 (ImageNet pretrained)
6	LSTM Hidden Size	256
7	Optimizer	Adam ($\beta_1 = 0.9, \beta_2 = 0.999$)
8	Learning Rate	1.0×10^{-4} (fixed)
9	Batch Size	32
10	Decision Threshold (θ)	$\mu_S + 3\sigma_S$ (calibrated on validation normals)

Table 2. Reproducible training and implementation details of the proposed Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework (CL-ADF)

SI. No	Parameter	Value
1	Training/validation split	80 training and 20 validation from training videos
2	Test data	Separate official test sequences
3	Maximum epochs	50
4	Early stopping	Patience of 8 epochs based on validation loss
5	Loss function	Mean Squared Error (MSE) reconstruction loss
6	CNN backbone	ImageNet-pretrained ResNet-18
7	ResNet-18 setting	Fine-tuned during training
8	CNN output dimension	512 features after global average pooling
9	LSTM layers	2 layers
10	LSTM hidden size	256
11	Dropout	0.30 between LSTM layers
12	Optimizer	Adam with $\beta_1 = 0.9, \beta_2 = 0.999$
13	Learning rate	1.0×10^{-4}
14	Batch size	32
15	Hardware environment	Intel Core i7 processor, 32 GB RAM, NVIDIA GPU with 12 GB memory
16	Training time	Approximately 2.1 hours for Avenue and 1.4 hours for UCSD Ped2
17	Inference time	Approximately 0.031 seconds per frame
18	Software environment	Python, PyTorch, CUDA-enabled GPU execution

4.1 Reproducible training and implementation details

Table 2 shows the complete training configuration of the proposed CL-ADF for better reproducibility. The input videos were resized to 224 × 224 pixels and were processed at 25

frames per second. Each video was split into clips of $T = 16$ consecutive frames. The original training sequences of the Avenue and UCSD Ped2 datasets were further divided into 80% training and 20% validation data. The testing sequences were kept separate and used only for evaluating the final performance. The proposed framework used an ImageNet-pretrained ResNet-18 backbone to extract spatial features.

Through the proposed work, the last fully connected classification layer of ResNet-18 was removed, and the output after global average pooling was a 512-dimensional CNN feature vector for each frame. During training, the ResNet-18 backbone was fine-tuned to make the extracted spatial features adapt to the anomaly patterns of surveillance videos. Then, the sequence of CNN feature vectors was fed into a two-layer LSTM network with a hidden dimension of 256. In order to avoid overfitting, a dropout rate of 0.30 was applied between the LSTM layers.

The model was trained with Adam optimiser, learning rate 1.0×10^{-4} , batch size 32, and mean squared error reconstruction loss. The maximum number of epochs was fixed to 50. Early stopping was used if validation loss did not improve for 8 epochs in a row. The best weights of the model were selected according to the minimum validation loss. The reconstruction error was used to calculate the anomaly score. The decision threshold was calculated as $\theta = \mu_s + 3\sigma_s$ with validation normal samples.

4.2 Computational runtime limitation

The present work is primarily focused on evaluating the proposed CL-ADF model in terms of detection-based performance measures such as accuracy, precision, recall, false alarm rate, and F1-score. The framework has shown promising performance for abnormal human event detection; however, a comprehensive runtime benchmark including frames per second (FPS), inference time per clip, model size, memory consumption, and detailed CPU/GPU utilisation is not provided in the current version. Thus, the claim on real-time performance is softened. In future work, we will conduct a full computational profiling study with standardised hardware settings. We will report runtime, FPS, inference latency, model size, and memory consumption for practical deployment analysis.

4.3 Hyperparameter settings for conventional baseline evaluation

The hyperparameter settings used for SVM, HMM, PCA,

and the proposed CL-ADF are given in Table 3. All methods used the same dataset split, the same preprocessing steps, the same clip formation, and the same validation strategy to allow for a fair comparison.

Table 3. Hyperparameter settings and validation strategy for baseline methods

Method	Hyperparameters Used	Validation Strategy
SVM	RBF kernel, $C = 10, \gamma = \text{scale}$	Same train-test split and validation set
HMM	Gaussian HMM, 4 hidden states, 100 maximum iterations	Same video sequences and clip settings
PCA	95% of variance retained; reduced features used for classification	Same preprocessing and feature input
CL-ADF	ResNet-18 CNN backbone, LSTM hidden size 256, Adam optimizer, learning rate 1.0×10^{-4}	Same dataset split and validation set

Note: SVM: Support Vector Machine; HMM: Hidden Markov Model; PCA: Principal Component Analysis; CL-ADF: Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework.

4.4 Statistical significance analysis of the proposed Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework

The statistical validation of the proposed CL-ADF method against conventional baseline methods such as SVM, HMM, and PCA is shown in Table 4. Values are expressed as mean $\pm$ standard deviation of five independent repeats. The standard deviation is a measure of the deviation among the repeated runs and not among individual video sequences. p -values are computed by a two-tailed paired t -test between CL-ADF and each baseline method on the same split of the dataset and evaluation settings. The proposed CL-ADF performs better than the selected baselines in terms of accuracy, precision, recall, F1-score, and has a lower false alarm rate. The improvement of CL-ADF over SVM, HMM, and PCA is statistically significant, as all the p -values reported are below 0.05. The statistical validation was carried out using five independent experimental runs. Therefore, the number of repeated runs is represented as $n = 5$. In each run, the same dataset split, preprocessing method, clip length, and evaluation metrics were used. The reported mean and standard deviation values in Table 4 are calculated from repeated experimental runs and not from individual video sequences.

Table 4. Statistical validation of Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework (CL-ADF) with existing methods

Method	Accuracy (%)	Precision (%)	Recall (%)	False Alarm Rate (%)	F1-Score (%)	p -Value vs. CL-ADF
SVM	81.24 ± 1.35	79.18 ± 1.42	77.65 ± 1.58	13.21 ± 0.74	78.36 ± 1.31	0.003
HMM	78.62 ± 1.49	77.21 ± 1.51	75.88 ± 1.65	14.85 ± 0.82	76.54 ± 1.43	0.002
PCA	76.35 ± 1.62	75.48 ± 1.67	73.96 ± 1.74	16.42 ± 0.91	74.68 ± 1.58	0.001
CL-ADF	92.18 ± 0.86	90.74 ± 0.92	89.36 ± 1.04	6.38 ± 0.43	90.05 ± 0.88	—

Note: SVM: Support Vector Machine; HMM: Hidden Markov Model; PCA: Principal Component Analysis; CL-ADF: Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework.

The mean value is calculated as given in Eq. (10).

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (10)$$

where, $\bar{x}$ represents the mean performance value, x_i represents the performance value obtained in the i -th run, and n represents the total number of repeated runs.

The standard deviation is calculated as given in Eq. (11).

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (11)$$

where, SD represents the standard deviation, x_i represents the value obtained in each run, $\bar{x}$ represents the mean value, and n represents the total number of runs.

To calculate the statistical significance, a two-tailed paired t -test was used. The difference between the proposed CL-ADF and each baseline method was calculated as given in Eq. (12).

$$d_i = x_{i,CL-ADF} - x_{i,Baseline} \quad (12)$$

The paired t -test value is calculated as given in Eq. (13).

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \quad (13)$$

where, d_i represents the difference between CL-ADF and the baseline method in the i -th run, $\bar{d}$ represents the mean difference, s_d represents the standard deviation of the differences, and n represents the number of repeated runs. The degrees of freedom are calculated as given in Eq. (14).

$$df = n - 1 = 4 \quad (14)$$

The obtained p -values were calculated from the paired t -test. A significance level of $p < 0.05$ was considered statistically significant. Therefore, the reported p -values confirm whether the improvement of CL-ADF over SVM, HMM, and PCA is statistically significant.

4.5 Performance evaluation based on feature dimension

The accuracy variation with respect to the selected feature dimensions of 50, 100, 150, 200, and 250 is shown in Figure 4. These feature dimensions were selected to evaluate the performance of each method with respect to increasing the number of extracted feature components. The model was evaluated on the same testing sequences for each feature dimension, and accuracy was computed as the ratio of correctly classified normal and abnormal events over the total number of evaluated events. The values plotted are the average accuracy of multiple runs of the experiment. The proposed CL-ADF achieves better accuracy as CNN-based spatial features and LSTM-based temporal learning provide richer event representation than SVM, HMM, and PCA.

4.6 Precision analysis with exponential feature dimensions

Precision variation with exponentially scaled feature dimensions of 32, 64, 128, 256, and 512 is shown in Figure 5. These values were chosen to study the effect of increasing the capacity of the feature representation on the classification of abnormal events. The precision for each dimension was calculated as the correctly detected abnormal events divided by all events predicted as abnormal. The plotted values are the mean precisions obtained from repeating the experiment evaluations. From Table 4, the proposed CL-ADF achieved higher precision, which means that the proposed CL-ADF can reduce the wrong abnormal predictions more than the baseline methods.

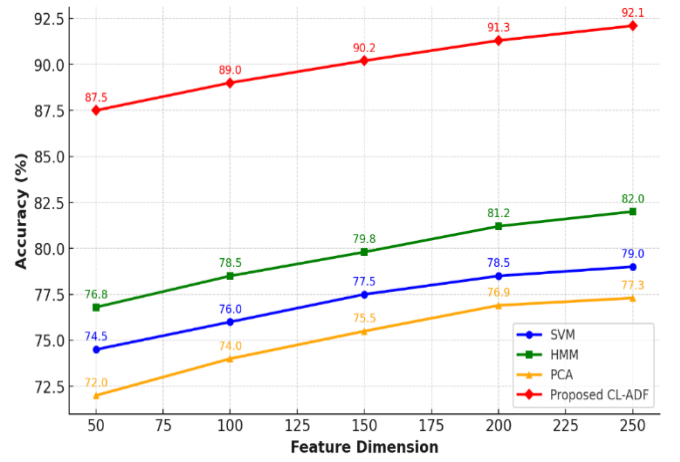


Figure 4. Comparison of accuracy for SVM, HMM, PCA, and proposed CL-ADF

Note: SVM: Support Vector Machine; HMM: Hidden Markov Model; PCA: Principal Component Analysis; CL-ADF: Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework.

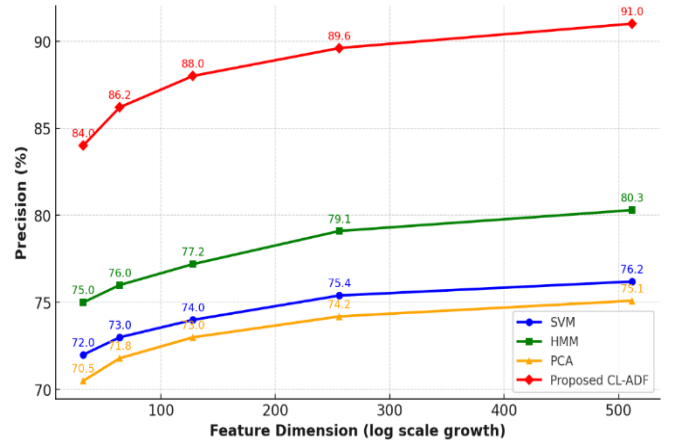


Figure 5. Precision comparison of SVM, HMM, PCA, and proposed CL-ADF

Note: SVM: Support Vector Machine; HMM: Hidden Markov Model; PCA: Principal Component Analysis; CL-ADF: Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework.

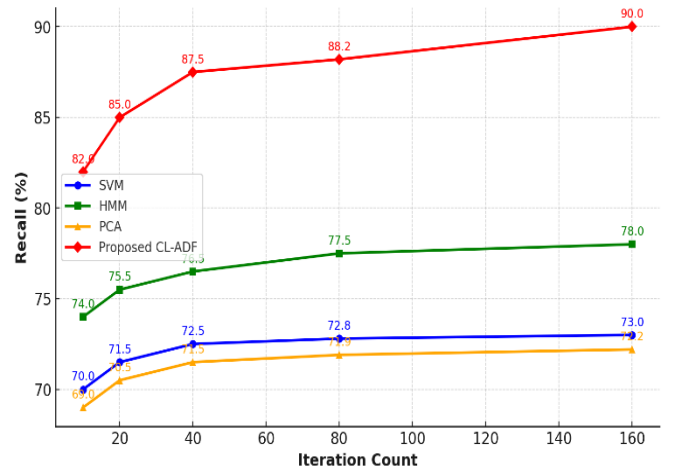


Figure 6. Recall comparison of SVM, HMM, PCA, and the proposed CL-ADF

Note: SVM: Support Vector Machine; HMM: Hidden Markov Model; PCA: Principal Component Analysis; CL-ADF: Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework.

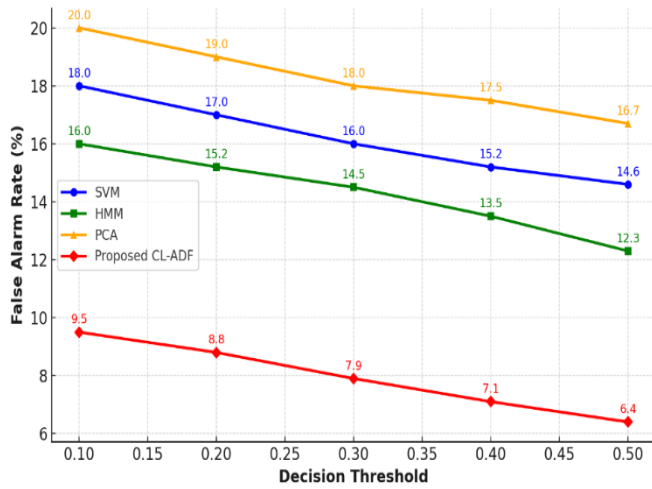


Figure 7. False alarm rate comparison of SVM, HMM, PCA, and proposed CL-ADF

Note: SVM: Support Vector Machine; HMM: Hidden Markov Model; PCA: Principal Component Analysis; CL-ADF: Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework.

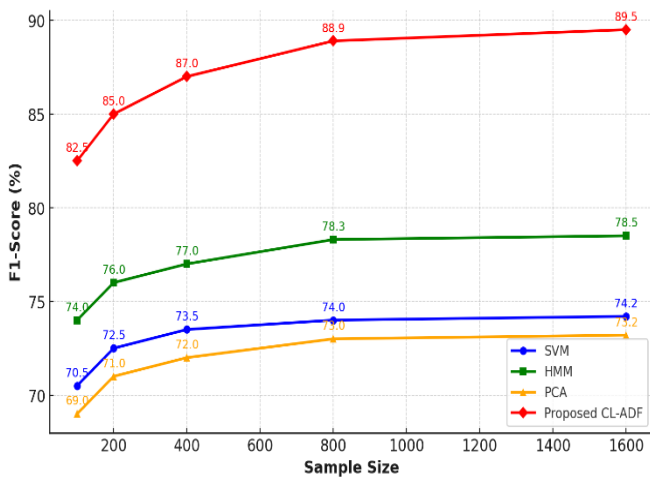


Figure 8. F1-score comparison of SVM, HMM, PCA, and proposed CL-ADF

Note: SVM: Support Vector Machine; HMM: Hidden Markov Model; PCA: Principal Component Analysis; CL-ADF: Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework.

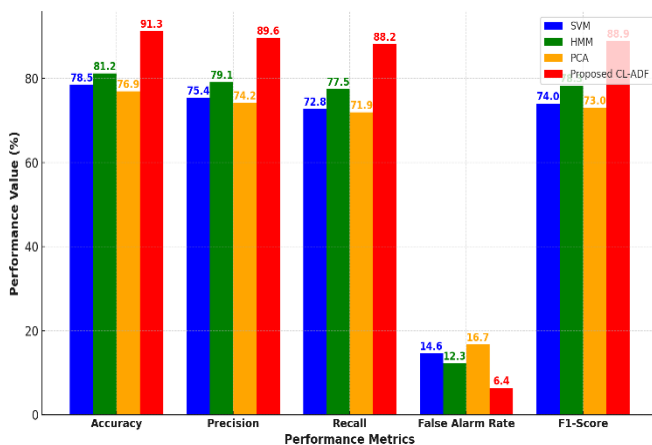


Figure 9. Performance analysis of SVM, HMM, PCA, and proposed CL-ADF

Note: SVM: Support Vector Machine; HMM: Hidden Markov Model; PCA: Principal Component Analysis; CL-ADF: Convolutional Neural Network–Long Short-Term Memory Anomaly Detection Framework.

4.7 Recall analysis with iteration count

The recall performance for different iteration counts of 10, 20, 40, 80, and 160 is shown in Figure 6. These iteration values were used to test the effect of model learning with the increase of training iterations. Recall was calculated for each number of iterations as the number of detected abnormal events divided by the total number of real abnormal events. The points plotted are the average recall from multiple runs. The proposed CL-ADF has better recall because the LSTM module captures temporal activity patterns better.

4.8 False alarm rate analysis with decision threshold

Figure 7 presents the false alarm rate for different decision threshold values of 0.1, 0.2, 0.3, 0.4, and 0.5. Those thresholds were chosen to explore the effect of anomaly score sensitivity on false alarm generation. For each threshold, the anomaly score was compared to the chosen threshold, and the false alarm rate was calculated from normal events that were incorrectly classified as abnormal. The plotted values are the average false alarm rates from repeated runs. The proposed CL-ADF has a smaller false alarm rate because the reconstruction-error-based decision function makes the discrimination of normal and abnormal event patterns more distinct.

4.9 F1-score analysis with sample size

Figure 8 illustrates the change in the F1-score for sample sizes of 100, 200, 400, 800, and 1600. These sample sizes were chosen to assess the stability of each method in different data availability scenarios. Precision and recall were calculated for each sample size, and the F1-score was calculated as the harmonic mean of precision and recall. Plotted values correspond to the achieved average F1-score for repeated experimental runs. The proposed CL-ADF achieves a higher F1-score because of a better trade-off between the detection of abnormal events and the control of false alarms.

4.10 Overall performance comparison of methods

Figure 9 shows the overall comparison of SVM, HMM, PCA, and the proposed CL-ADF in terms of accuracy, precision, recall, false alarm rate, and F1-score. The values shown in Figure 9 are from the last experimental evaluation in Table 4. Each metric value is the average result of repeated runs with the same dataset split, preprocessing settings, and validation strategy. The proposed CL-ADF outperformed the baseline methods across most metrics and achieved a lower false-alarm rate.

4.11 Real-world deployment discussion

Although the proposed CL-ADF exhibits better performance on benchmark video anomaly datasets, the practical implementation in real-life CCTV environments may involve further challenges. Real surveillance videos typically contain crowded scenes, camera motion, illumination changes, shadow effects, occlusions, low-resolution footage, and video compression noise. These factors can impact the feature quality and the accuracy of abnormal event detection. In highly crowded environments, abnormal human activities may be partially occluded or visually similar to normal crowd movement, which may lead to false alarms or missed

detections. CL-ADF therefore shows promise for real-time applications, but further validation with live CCTV feeds, outdoor surveillance footage, night-time footage, and large-scale crowded environments is needed before practical forensic use can be realized.

5. CONCLUSIONS

The proposed CL-ADF provides an effective and robust approach for identifying abnormal human activities in video forensics. By combining the spatial feature extraction capability of convolutional neural networks with the sequential modelling capability of long short-term memory networks, CL-ADF addresses several limitations of conventional methods, such as SVM, HMM, and PCA, which commonly face challenges related to scalability, noise sensitivity, and limited temporal modelling capacity. The experimental results demonstrated a detection accuracy of more than 91%, with precision and F1-score values reaching approximately 90%. In addition, the false-alarm rate was reduced to nearly half of that reported for the conventional baseline methods. These findings support CL-ADF as an effective framework for improving forensic video analysis, facilitating the identification of suspicious or criminal behaviour, and providing support for surveillance, police investigation, and judicial proceedings.

Several promising directions can be considered for future development. The design of optimised and lightweight CNN–LSTM configurations will be important for implementation on resource-constrained devices, including edge devices and Internet of Things (IoT) platforms. The incorporation of Transformer-based architectures, self-supervised learning strategies, and attention mechanisms may further improve anomaly detection performance while reducing dependence on large labelled datasets. Integrating CL-ADF into scalable cloud platforms may also enable the processing of large surveillance-video streams and support broader applications in national security systems and smart-city environments. In addition, extending the framework to domains such as crowd control, intelligent transportation systems, and medical monitoring may further expand its applicability. Finally, the integration of explainable artificial intelligence techniques could improve the transparency, trustworthiness, and legal acceptability of automated forensic conclusions. Overall, CL-ADF advances abnormal event detection and provides a basis for further development in smart and secure video forensic analysis.

Despite its improved performance in abnormal human event detection for video forensic analysis, the proposed CL-ADF has several limitations. Its performance depends on the quality and diversity of the surveillance-video datasets used for training and evaluation. In highly crowded scenes, heavy occlusion, poor illumination, camera motion, and low-resolution video conditions, detection accuracy may decrease. In addition, the CNN–LSTM architecture requires considerable computational resources for real-time deployment, particularly in large-scale surveillance networks. The decision threshold may also need to be recalibrated when the model is deployed in new environments or under previously unseen video conditions. Future work will focus on lightweight model design, adaptive threshold selection, explainable decision support, and real-time edge-based implementation for practical forensic applications. Another

limitation is that the experimental comparison was restricted to conventional baseline methods, including SVM, HMM, and PCA. Future studies will conduct reproducible quantitative comparisons with recent video anomaly detection models, including MIST, RTFM, memory-augmented autoencoders, and Transformer-based approaches, using the same benchmark datasets and evaluation settings.

ACKNOWLEDGMENT

The authors sincerely acknowledge the PG Department of Computer Applications, JSS College of Arts, Commerce and Science, Mysuru, and the Government Engineering College, Chamarajanagara, affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India, for providing the research infrastructure and academic support necessary for conducting this study.

REFERENCES

- [1] Chen, T.Y., Huang, Y.Y., Chu, Y.C., Chen, S.L., Chen, X.Y., Wu, P.C. (2024). Artificial intelligence system combining with infrared thermography and visible image for abnormal temperature detection and floor material identification. *IEEE Sensors Journal*, 24(24): 42181-42194. <https://doi.org/10.1109/JSEN.2024.3439362>
- [2] Chen, L.W., Liao, C.W., Liu, J.X. (2025). Proactive crowdsourced monitoring and sensing with expansible activity recognition based on Internet of Things localization. *IEEE Internet of Things Journal*, 12(11): 17674-17686. <https://doi.org/10.1109/JIOT.2025.3539293>
- [3] Lin, W., Chen, Y., Wu, J., Wang, H., Sheng, B., Li, H. (2014). A new network-based algorithm for human activity recognition in videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(5): 826-841. <https://doi.org/10.1109/TCSVT.2013.2280849>
- [4] Mahendra, H.N., Pushpalatha, V., Swamy, M., Subramoniam, S.R., Praveen, J. (2026). Quantum convolutional neural network-based hybrid network for remote sensing image classification. *Engineering Applications of Artificial Intelligence*, 164: 113195. <https://doi.org/10.1016/j.engappai.2025.113195>
- [5] Yin, J., Yang, Q., Pan, J.J. (2008). Sensor-based abnormal human-activity detection. *IEEE Transactions on Knowledge and Data Engineering*, 20(8): 1082-1090. <https://doi.org/10.1109/TKDE.2007.1042>
- [6] Gao, B.B. (2026). Dual-masked and discriminative reconstruction for unified vision anomaly detection. *IEEE Transactions on Image Processing*, 35: 4701-4712. <https://doi.org/10.1109/TIP.2026.3687095>
- [7] Fan, J., Liu, Z., Du, H., Kang, J., Niyato, D., Lam, K.Y. (2025). Improving security in IoT-based human activity recognition: A correlation-based anomaly detection approach. *IEEE Internet of Things Journal*, 12(7): 8301-8315. <https://doi.org/10.1109/JIOT.2024.3501361>
- [8] Zhang, M., Wang, Y., Li, P., Zhang, Y., Cui, Y., Xiang, M. (2026). Effects of magnetocardiography array errors on cardiac source imaging. *IEEE Transactions on Instrumentation and Measurement*, 75: 1-13. <https://doi.org/10.1109/TIM.2026.3676206>
- [9] Tsiktiris, D., Lalas, A., Dasygenis, M., Votis, K. (2024).

- Multimodal abnormal event detection in public transportation. *IEEE Access*, 12: 133469-133480. <https://doi.org/10.1109/ACCESS.2024.3425308>
- [10] Chung, H.J., Kang, B., Yang, Y.S. (2025). N-DriverMotion: Driver motion learning and prediction using an event-based camera and directly trained spiking neural networks on Loihi 2. *IEEE Open Journal of Vehicular Technology*, 6: 68-80. <https://doi.org/10.1109/OJVT.2024.3504481>
- [11] Hu, W.C., Chen, L.B., Lin, H.M. (2024). A method for abnormal behavior recognition in aquaculture fields using deep learning. *IEEE Canadian Journal of Electrical and Computer Engineering*, 47(3): 118-126. <https://doi.org/10.1109/ICJECE.2024.3398653>
- [12] Huang, C., Wu, Z., Wen, J., Xu, Y., Jiang, Q., Wang, Y. (2022). Abnormal event detection using deep contrastive learning for intelligent video surveillance system. *IEEE Transactions on Industrial Informatics*, 18(8): 5171-5179. <https://doi.org/10.1109/TII.2021.3122801>
- [13] Chen, J., Wang, C., Liu, Y. (2024). Vibration signal based abnormal gait detection and recognition. *IEEE Access*, 12: 89845-89855. <https://doi.org/10.1109/ACCESS.2024.3417377>
- [14] Tay, N.C., Connie, T., Ong, T.S., Teoh, A.B.J., Teh, P.S. (2023). A review of abnormal behavior detection in activities of daily living. *IEEE Access*, 11: 5069-5088. <https://doi.org/10.1109/ACCESS.2023.3234974>
- [15] Ahmad Hamid, A., Monadjemi, S.A., Shoushtarian, B. (2025). ABDviaMSIFAT: Abnormal crowd behavior detection utilizing a multi-source information fusion technique. *IEEE Access*, 13: 75000-75019. <https://doi.org/10.1109/ACCESS.2024.3436007>
- [16] Xia, K., Zhang, L., Yuan, S., Lou, Y. (2023). SCSS: An intelligent security system to guard city public safe. *IEEE Access*, 11: 76415-76426. <https://doi.org/10.1109/ACCESS.2023.3297643>
- [17] Hong, Z., Wang, G., Lyu, W., Guo, B., Ding, Y., Wang, H. (2024). Nationwide behavior-aware coordinates mining from uncertain delivery events. *IEEE Transactions on Knowledge and Data Engineering*, 36(11): 6681-6698. <https://doi.org/10.1109/TKDE.2024.3411562>
- [18] Liu, W., Shi, Y., Liang, J., Wang, M., Zhou, Y., Fu, F. (2026). A combined network model-based abnormal data compensation method for enhanced electrical impedance tomography. *IEEE Sensors Journal*, 26(7): 11120-11129. <https://doi.org/10.1109/JSEN.2026.3667189>
- [19] Lu, M., Liu, Y., Chen, Y., Yang, F. (2026). Edge Harmony Attention Network for semi-supervised medical image segmentation. *IEEE Transactions on Instrumentation and Measurement*, 75: 1-17. <https://doi.org/10.1109/TIM.2026.3664601>
- [20] Zeng, X., Jiang, Y., Ding, W., Li, H., Hao, Y., Qiu, Z. (2023). A hierarchical spatio-temporal graph convolutional neural network for anomaly detection in videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1): 200-212. <https://doi.org/10.1109/TCSVT.2021.3134410>
- [21] Elmetwally, A., Eldeeb, R., Elmougy, S. (2025). Deep learning based anomaly detection in real-time video. *Multimedia Tools and Applications*, 84: 9555-9571. <https://doi.org/10.1007/s11042-024-19116-9>
- [22] Ge, Y., Taha, A., Shah, S.A., Dashtipour, K., Zhu, S., Cooper, J. (2023). Contactless WiFi sensing and monitoring for future healthcare—Emerging trends, challenges, and opportunities. *IEEE Reviews in Biomedical Engineering*, 16: 171-191. <https://doi.org/10.1109/RBME.2022.3156810>
- [23] Pati, B., Sahoo, A.K., Udgata, S.K. (2024). Caption generation for sensing-based activity using attention-based learning models. *IEEE Sensors Letters*, 8(3): 1-4. <https://doi.org/10.1109/LSSENS.2023.3347486>
- [24] Sengonul, E., Samet, R., Abu Al-Haija, Q., Alqahtani, A., Alsemmeari, R.A., Alghamdi, B., Alturki, B., Alsulami, A.A. (2025). Abnormal event detection in surveillance videos through LSTM auto-encoding and local minima assistance. *Discover Internet of Things*, 5: 32. <https://doi.org/10.1007/s43926-025-00127-3>
- [25] Ullah, A., Muhammad, K., Del Ser, J., Baik, S.W., de Albuquerque, V.H.C. (2019). Activity recognition using temporal optical flow convolutional features and multilayer LSTM. *IEEE Transactions on Industrial Electronics*, 66(12): 9692-9702. <https://doi.org/10.1109/TIE.2018.2881943>
- [26] Wu, Q., Wu, Y., Zhang, Y., Zhang, L. (2022). A local-global estimator based on large kernel CNN and Transformer for human pose estimation and running pose measurement. *IEEE Transactions on Instrumentation and Measurement*, 71: 1-12. <https://doi.org/10.1109/TIM.2022.3200438>
- [27] Mahendra, H., Pushpalatha, V., Velluri, R., Nagaraju, S., Kumar, D., Sukumar, P., Basavaraj, N., Swamy, M. (2025). Land use/land cover (LULC) change classification for change detection analysis of remotely sensed data using machine learning-based random forest classifier. *Nature Environment and Pollution Technology*, 24: B4238. <https://doi.org/10.46488/NEPT.2025.v24i02.B4238>
- [28] Pushpalatha, V., Mallikarjuna, B., Mahendra, N., Subramoniam, S.R., Swamy, M. (2025). Land use and land cover classification for change detection studies using convolutional neural network. *Applied Computing and Geosciences*, 25: 100227. <https://doi.org/10.1016/j.acags.2025.100227>
- [29] Li, W., Mahadevan, V., Vasconcelos, N. (2014). Anomaly detection and localization in crowded scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1): 18-32. <https://doi.org/10.1109/TPAMI.2013.111>
- [30] Tang, Y., Zhao, L., Zhang, S., Gong, C., Li, G., Yang, J. (2020). Integrating prediction and reconstruction for anomaly detection. *Pattern Recognition Letters*, 129: 123-130. <https://doi.org/10.1016/j.patrec.2019.11.024>
- [31] Sheela, S., Patil, A.R., Ganya, A.M., Juturu, A., Pattankar, V.V., Kumaraswamy, S. (2025). Secure dual-layer access control using visual cryptography and LSB watermarking. In *2025 International Conference on Innovations in Intelligent Systems: Advancements in Computing, Communication, and Cybersecurity (ISAC3)*, Bhubaneswar, India, pp. 1-6. <https://doi.org/10.1109/ISAC364032.2025.11156269>
- [32] Anu, H., Rathnakara, S., Mallikarjunaswamy, S. (2025). Enhanced ECG signal classification with CNN-LSTM networks using Aquila optimization. *Engineering, Technology & Applied Science Research*, 15(3): 23461-23466. <https://doi.org/10.48084/etasr.10492>
- [33] Niveditha, H.R., Anitha, S., Ramaswamy, N.K., Ramaswamy, R., Mallikarjunaswamy, S. (2025). A

- hybrid machine learning model for peripheral artery disease prediction and real-time applications. *Engineering, Technology & Applied Science Research*, 15(3): 23692-23698. <https://doi.org/10.48084/etasr.10354>
- [34] Wan, B., Jiang, W., Fang, Y., Luo, Z., Ding, G. (2021). Anomaly detection in video sequences: A benchmark and computational model. *IET Image Processing*, 15(14): 3454-3465. <https://doi.org/10.1049/ipr2.12258>
- [35] Luo, J., Lin, J., Yang, Z., Liu, H. (2022). SMD anomaly detection: A self-supervised texture-structure anomaly detection framework. *IEEE Transactions on Instrumentation and Measurement*, 71: 1-11. <https://doi.org/10.1109/TIM.2022.3194920>
- [36] Zheng, Z., Liu, W., Liu, R., Wang, L., Mao, L., Qiu, Q. (2022). Anomaly detection of metro station tracks based on sequential updatable anomaly detection framework. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(11): 7677-7691. <https://doi.org/10.1109/TCSVT.2022.3181452>
- [37] Doshi, K., Yilmaz, Y. (2021). Online anomaly detection in surveillance videos with asymptotic bound on false alarm rate. *Pattern Recognition*, 114: 107865. <https://doi.org/10.1016/j.patcog.2021.107865>
- [38] Alladi, T., Gera, B., Agrawal, A., Chamola, V., Yu, F.R. (2021). DeepADV: A deep neural network framework for anomaly detection in VANETs. *IEEE Transactions on Vehicular Technology*, 70(11): 12013-12023. <https://doi.org/10.1109/TVT.2021.3113807>
- [39] Jin, W., Dang, F., Zhu, L. (2024). Feature enhancement with reverse distillation for hyperspectral anomaly detection. *IEEE Geoscience and Remote Sensing Letters*, 21: 1-5. <https://doi.org/10.1109/LGRS.2024.3456178>
- [40] Cho, M.A., Kim, T., Kim, W.J., Cho, S., Lee, S. (2022). Unsupervised video anomaly detection via normalizing flows with implicit latent features. *Pattern Recognition*, 129: 108703. <https://doi.org/10.1016/j.patcog.2022.108703>
- [41] Xu, M., Zhou, X., Gao, X., He, W., Niu, S. (2023). Discriminative feature learning framework with gradient preference for anomaly detection. *IEEE Transactions on Instrumentation and Measurement*, 72: 1-10. <https://doi.org/10.1109/TIM.2022.3228007>
- [42] Zhang, M., Wang, J., Qi, Q., Zhuang, Z., Sun, H., Liao, J. (2024). Cognition guided video anomaly detection framework for surveillance services. *IEEE Transactions on Services Computing*, 17(5): 2109-2123. <https://doi.org/10.1109/TSC.2024.3407588>
- [43] Sheela, S., Jyothi, S., Latha, A.P., Ganesh, H.J., Komala, M., Naveen Kumar, C. (2024). Automated land cover classification in urban environments with deep learning-based semantic segmentation. In *2024 International Conference on Recent Advances in Science and Engineering Technology (ICRASET)*, B G Nagara, Mandya, India, pp. 1-7. <https://doi.org/10.1109/ICRASET63057.2024.10895689>
- [44] Rekha, V., Sharmila, N., Komala, M., Sukumar, G.S.P., Mallikarjunaswamy, S., Naveen, K.B. (2024). Hybrid edge-cloud approach for renewable energy management using deep learning with predictive analytics. In *2024 International Conference on Recent Advances in Science and Engineering Technology (ICRASET)*, B G Nagara, Mandya, India, pp. 1-7. <https://doi.org/10.1109/ICRASET63057.2024.10895726>
- [45] Mahendra, H.N., Swamy, M. (2023). An analysis of change detection in land use land cover area of remotely sensed data using supervised classifier. *International Journal of Environmental Technology and Management*, 26: 498-511. <https://doi.org/10.1504/IJETM.2023.134322>
- [46] Wang, R., Hu, J. (2025). Gaussian-inspired attention mechanism for hyperspectral anomaly detection. *IEEE Geoscience and Remote Sensing Letters*, 22: 1-5. <https://doi.org/10.1109/LGRS.2024.3514166>
- [47] Guan, W., Cao, J., Zhao, H., Gu, Y., Qian, S. (2024). WAKE: A weakly supervised business process anomaly detection framework via a pre-trained autoencoder. *IEEE Transactions on Knowledge and Data Engineering*, 36(6): 2745-2758. <https://doi.org/10.1109/TKDE.2023.3322411>
- [48] Wang, L., Tian, J., Zhou, S., Shi, H., Hua, G. (2023). Memory-augmented appearance-motion network for video anomaly detection. *Pattern Recognition*, 138: 109335. <https://doi.org/10.1016/j.patcog.2023.109335>
- [49] Swamy, S.M., Sharmila, N., Sukumar, P.G., Chikkasiddaiah, C., Prasad, A.M. (2026). A hybrid energy and proximity-based confined clustering scheme for wireless sensor networks. *Journal of Robotics and Control*, 7(1): 3259-3270. <https://doi.org/10.18196/jrc.v7i1.27468>
- [50] Honnegowda, J., Mallikarjunaswamy, K., Srikantaswamy, M. (2024). An efficient abnormal event detection system in video surveillance using deep learning-based reconfigurable autoencoder. *Ingénierie des Systèmes d'Information*, 29(2): 677-686. <https://doi.org/10.18280/isi.290229>
- [51] Jyothi, H., Komala, M., Mallikarjunaswamy, S. (2024). A comprehensive survey on technologies in video-based event detection and recognition using machine learning and deep learning techniques. *2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON)*, Bengaluru, India, pp. 1-5. <https://doi.org/10.1109/NMITCON62075.2024.10698959>
- [52] Fan, J., Ji, Y., Wu, H., Ge, Y., Sun, D., Wu, J. (2024). An unsupervised video anomaly detection method via optical flow decomposition and spatio-temporal feature learning. *Pattern Recognition Letters*, 185: 239-246. <https://doi.org/10.1016/j.patrec.2024.08.013>
- [53] Xiao, C., Lu, J., Xu, X., Zhou, F., Xie, T., Lu, W. (2025). Reconciling attribute and structural anomalies for improved graph anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems*, 36(9): 16661-16674. <https://doi.org/10.1109/TNNLS.2025.3561172>
- [54] Satish, P., Srikantaswamy, M., Ramaswamy, N.K. (2023). Image region driven prior selection for image deblurring. *Multimedia Tools and Applications*, 82: 24181-24202. <https://doi.org/10.1007/s11042-023-14335-y>
- [55] Satish, P., Srikantaswamy, M., Ramaswamy, N.K. (2020). A comprehensive review of blind deconvolution techniques for image deblurring. *Traitement du Signal*, 37(3): 527-539. <https://doi.org/10.18280/ts.370321>
- [56] Jing, Y., Jiang, C., Wang, J., Sun, J., Zhou, S., Zhan, Y. (2026). Spatiotemporal satellite user resource forecasting in MEO-LEO networks via a CNN-LSTM model. *IEEE Transactions on Network Science and Engineering*, 13:

9555-9571.

<https://doi.org/10.1109/TNSE.2026.3694942>

- [57] Ali, M.M. (2023). Real-time video anomaly detection for smart surveillance. IET Image Processing, 17(5): 1375-1388. <https://doi.org/10.1049/ipr2.12720>
- [58] Asad, M., Ullah, I., Sistu, G., Madden, M.G. (2026). Contextual graph modeling for open-set object detection in multi-view autonomous driving. IEEE Open Journal of Vehicular Technology, 7: 1323-1336. <https://doi.org/10.1109/OJVT.2026.3687800>
- [59] Zhang, D., Fang, W., Liu, Y., Lyu, Z., Xiong, C., Wang, Z. (2024). Two-stage video anomaly detection based on dual-stream networks and multi-instance learning. IET Image Processing, 18(14): 4843-4851. <https://doi.org/10.1049/ipr2.13286>

NOMENCLATURE

F_s	Extracted spatial feature map
X_t	Input video frame at time step t
W_c	Convolutional filter weight matrix
b_c	Bias term in convolutional layer
$\phi(\cdot)$	Activation function (e.g., ReLU)
h_t	Hidden state at time step t
h_{t-1}	Previous hidden state
x_t	Input feature at time t
W_h, W_x	Weight matrices for hidden and input states
b_h	Bias term for LSTM cell
$\tanh, \sigma$	Non-linear activation functions

S_a	Anomaly score
f_{in}	Original input video feature
f_{rec}	Reconstructed feature from autoencoder
I_n	Normalized input video frame
μ	Mean pixel intensity of frame
σ	Standard deviation of pixel intensity
O_c	Classification output (0 – normal, 1 – abnormal)
θ	Decision threshold parameter
$\mathbb{I}(\cdot)$	Indicator function for event classification
E	Reconstruction error (Euclidean distance)
T	Clip length (number of frames per sequence)
L	Loss function during model training
η	Learning rate in optimizer

Greek symbols

β_1, β_2	Momentum parameters for Adam optimizer
μ_s, σ_s	Mean and standard deviation of anomaly scores for threshold calibration
ϕ	Activation function (ReLU, sigmoid)

Subscripts

t	Time index for sequence
c	Convolutional layer index
h	Hidden state variable in LSTM
rec	Reconstructed feature representation
$norm$	Normalized input frame
in	Input layer variable
out	Output layer variable