J.Natn.Sci.Foundation Sri Lanka 2025 54 (1): 31 - 43
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12587

RESEARCH ARTICLE

Deep Learning

Fast and reliable identification of abnormal crowd behaviour in surveillance footage

MM Krishnan[1] , ST Ramya[2] , K Ramar[3]
[1] *Department of Computer Science and Engineering. Francis Xavier Engineering College, Vannarpettai, Tirunelveli, Tamilnadu, India.*
[2] *Department of Computer Science and Design, RMK Engineering College, Kavarapettai, Tiruvallur, Tamilnadu, India.*
[3] *Department of Computer Science and Engineering, RMK College of Engineering and Technology, Pudovoyal, Tiruvallur, Tamilnadu, India.*

Submitted: 07 March 2025; Revised: 23 February 2026; Accepted: 07 March 2026

Abstract: Detecting unusual crowd events in surveillance video is crucial due to increased crime rates in recent years. However, automatically identifying these events is challenging because feature comparison across the training and test phases is time-consuming. To address this, we introduce the Markov-nearest transition unusual event classifier (MTUEC) classifier, which classifies input frames as either normal or unusual events. The MTUEC algorithm comprises several modules: the spatial slice model, static object removal computation, spatio-temporal estimation, and the Markov-nearest transition based unusual events classifier. The MTUEC algorithm focuses on comparing features between immediate training frames and the testing frame. If the immediate training frame does not match the testing frame, the algorithm considers other training frames for comparison. This approach significantly reduces the time needed to detect unusual crowd events. To examine the effectiveness of the proposed method, we used two benchmark datasets for unusual crowd events: UMN and UCSD Ped1 and Ped2. We also compared the performance of the proposed approach with several existing algorithms for unusual event detection.

Keywords: Events classification, feature comparison, Markov-nearest transition, slice model, unusual events detection.

INTRODUCTION

In the digital era, surveillance cameras are ever-present in areas where security is needed. Surveillance camera footage is used for post-event analysis (Kim et al., 2011), such as event recognition, action recognition, unusual crowd event detection, and abnormal event detection. These analyses can be applied across various surveillance domains, such as traffic surveillance environments, institutions, and hospitals. To conduct post-event analysis, many approaches, such as machine learning (Paul et al., 2023), deep learning (Lohithashva et al., 2018), clustering (Anjum & Cavallaro, 2008), and feature-based (Gnouma et al., 2018), have been widely used. Nowadays, anomaly detection is widely used in surveillance settings, as crime rates have increased in recent years. Therefore, an anomaly detection system is needed to alert people and rescue them from abnormal humans or situations. Unusual crowd behaviour analysis is also a promising application in video technology, as crowds can pose unwanted problems for the public. Though many research works have addressed unusual events and crowd problems, those studies have not achieved high accuracy, since analyzing crowd behaviour is a challenging task (Al-Khazaleh et al., 2026; Beigh et al, 2025; Jadhav & Bartere, 2025; Biswas & Babu, 2017). Developing crowd patterns and extracting features from density content are difficult tasks, since humans may be located at various locations, split and move in different directions, or all be located in the same place. Even if many algorithms try to detect objects or crowds, they are mostly developed for common scenes with a minimum number of people. In this situation, when these algorithms are applied to

---

* Corresponding author (mukeshkrishnan91.m@gmail.com; https://orcid.org/0009-0003-6401-9552)

crowd event frames, the unusual crowd event problem cannot be handled, since a large number of humans are present in the frame sequence in reality. Therefore, those algorithms cannot identify the effective features in the frame sequence (Chen et al., 2016).

Abnormal actions can be categorized into two types in the literature review (Chong et al*.,* 2017; Chu et al*.,* 2018; Cong et al., 2013): dense and sparse. Sparse abnormal actions involve only a few individuals and are characterized by conditions such as loitering, intrusion, falling, fighting each other, etc. On the other hand, groups of people may be involved in anomalous behaviour, such as unusual crowd activities, people trying to escape during a natural disaster, etc. Detecting abnormal crowd behaviour in a surveillance environment is a crucial challenge, as analysing large volumes of video data can be expensive and time-consuming. To reduce computational complexity, this paper introduced the Markov-transition based unusual events classifier (MTUEC) using the following four modules: spatial slice model, removal of static objects computation, spatio-temporal estimation, and unusual events classifier. The paper's main contributions are summarized here.

- Spatio-temporal estimation (STE) is introduced to estimate temporal information between moving objects for efficient analysis of unusual event detection.
- MTUEC is introduced to classify instances as normal or abnormal. Also, this classifier reduces feature comparison between the training and testing phases, since it mainly uses the previous instance for comparison.
- Various unusual event approaches, such as feature-based, machine learning, and deep learning-based methods, are used to analyze the results in different ways.

Related works

To capture intricate spatiotemporal patterns in video and trajectory data, deep learning-based models have become increasingly important for anomaly and uncommon event identification. The modelling of long-term dependencies in motion behaviour has been investigated using transformer designs and recurrent neural networks. To avoid bridge collisions, Mou et al. (2026) suggested a Transformer–BiLSTM hybrid model for identifying anomalous ship trajectories. Although the method works well for learning sequential dependencies, it is not suitable for real-time deployment in large-scale surveillance systems due to its expensive computational resources and reliance on massive amounts of training data.

Another strategy to lessen reliance on labelled data is self-supervised learning. Yang et al. (2025) presented a paradigm for video anomaly detection based on self-supervised spatio-temporal proxy tasks, enabling reliable feature learning without explicit labels. Low-latency applications may be hampered by this method's long temporal sequences and computationally intensive training process, despite its enhanced generalization.

With learning-based pattern modelling, trajectory-focused anomaly detection continues to develop. A thorough analysis of trajectory anomaly detection techniques was given by Li et al. (2025), who also highlighted new issues such scene reliance, occlusion, and trajectory fragmentation. Cho and Kang (2025) further address these problems by proposing a context-aware variational autoencoder for detecting pedestrian trajectory anomalies in metropolitan settings. VAE-based techniques are sensitive to reconstruction thresholds and require careful adjustment to prevent excessive false-positive rates, even with improved contextual modelling.

Liang et al. (2025) introduced an interaction–scene collaborative framework that represents object interactions and scene context together for traffic anomaly identification, going beyond trajectory-only representations. Although relational dynamics are well captured by this method, it is highly dependent on precise item detection and interaction modelling, both of which may suffer in cluttered or crowded environments. Similarly, Sundaram et al. (2025) used deep learning methods to analyze human behaviour in intelligent video surveillance systems, achieving excellent results at the expense of higher memory and processing demands.

In this case, spatial location information is erased by computing a histogram. Furthermore, most methods include a single mechanism to extract spatial or temporal information (Sodemann et al., 2021; Vallejo et al., 2009; Ahmed et al., 2018; Lin et al., 2015), because those algorithms work based on anyone's perception of objects, such as direction, appearance, motion, etc. The end result failed in many places due to illumination changes, which occluded the foreground object. It tends to have poor accuracy. To avoid these problems, STE is introduced to extract temporal information between moving objects for efficient detection of unusual events. Also, to reduce the computational and comparison complexity, this paper introduced the MTUEC technique by using the following four modules: spatial slice model, removal static objects computation, spatio-temporal estimation, and unusual deep learning-based techniques have been increasingly used in anomalous event classifiers in recent years to overcome the drawbacks of manually created features and explicit object tracking. Spatiotemporal autoencoders, which learn compact representations via frame reconstruction or future frame prediction, have been widely used to model typical crowd behaviour. The identification of abnormal occurrences is based on high prediction or reconstruction mistakes. Despite their effectiveness, these techniques often struggle with complex motion patterns and require substantial training data to generalize effectively across other scenarios.

This concept is further developed by GAN-based video synthesis methods, which produce realistic future frames conditioned on historical observations (Jency & Ramar, 2025). When there is a substantial difference between the synthesized frames and the real observations, abnormal events are identified. GAN-based techniques are vulnerable to mode collapse, computationally intensive, and challenging to train, which reduces their resilience in real-world surveillance settings, even with improved motion dynamics models.

In recent years, transformer-based anomaly detection (Mancy & Naith, 2025) models have attracted attention for their ability to leverage self-attention mechanisms to capture long-range temporal relationships. When modelling intricate spatiotemporal interactions across lengthy video sequences, these techniques work admirably. However, they are less appropriate for real-time or resource-constrained surveillance systems due to their huge memory requirements and quadratic computational complexity with respect to sequence length.

By building spatiotemporal graphs, graph-based models have also been proposed to explicitly depict interactions among objects, regions, or motion patterns (Ilyas & Bawany, 2025). Relational dynamics in crowded scenes can be effectively modeled using graph convolutional networks. However, these methods usually depend on precise object tracking and detection, and they perform worse in crowded settings with a lot of occlusions and fragmented observations.

The majority of current methods have one or more of the following drawbacks, notwithstanding the success of these sophisticated models: they are computationally intensive, require extensive annotated data, are sensitive to changes in illumination and occlusion, or depend on long-term trajectory consistency. Furthermore, many approaches focus on either temporal dynamics or spatial appearance separately, leading to poor robustness in real-world surveillance scenarios.

Research gap

A number of important issues still need to be addressed, despite the significant advances in unusual event detection enabled by trajectory-based techniques and more recent deep learning approaches such as spatiotemporal autoencoders, GAN-based video synthesis, transformer-based anomaly detection, and graph-based models. While deep learning methods frequently require large-scale training data, significant computational resources, and complex model architectures, trajectory-based approaches struggle in scenarios with many people due to occlusion and fragmented tracking. The application of transformer-based and graph-based techniques in real-time surveillance scenarios is further limited by their reliance on lengthy temporal sequences or dependable object interactions. Furthermore, the majority of current approaches focus on either temporal dynamics or spatial appearance separately, resulting in reduced resilience to partial occlusions, background motion, and changes in lighting. As a result, there is a research need for a reliable and effective framework for detecting abnormal events that can preserve low computational and comparison complexity appropriate for real-world surveillance systems while simultaneously capturing spatiotemporal information without explicit trajectory dependency.

MATERIALS AND METHODS

The ultimate aim of the proposed method is to reduce the computational and feature comparison complexity by using the following four modules: 1. The spatial slice model is used to estimate the frame sequence using colour, texture, and edge features. 2. Removal static objects computation: it is used to identify the foreground objects in order to reduce the feature comparison while removing unwanted static objects. 3. Spatio-temporal estimation is employed to extract the motion of foreground objects. 4. Marko-nearest transition-based unusual events classifier is introduced to reduce the feature comparison between the training and testing phases. An overview of the proposed methodology is shown in Figure 1. The proposed method used the frame sequence as the input for analyzing unusual events.

More clarity can be obtained by clearly outlining the data pretreatment and module integration pipeline, even though the suggested methodology describes the main functional modules. To capture localized movement patterns, the suggested architecture first breaks down raw surveillance footage into individual frames, which are then spatially segmented using the spatial slice model.

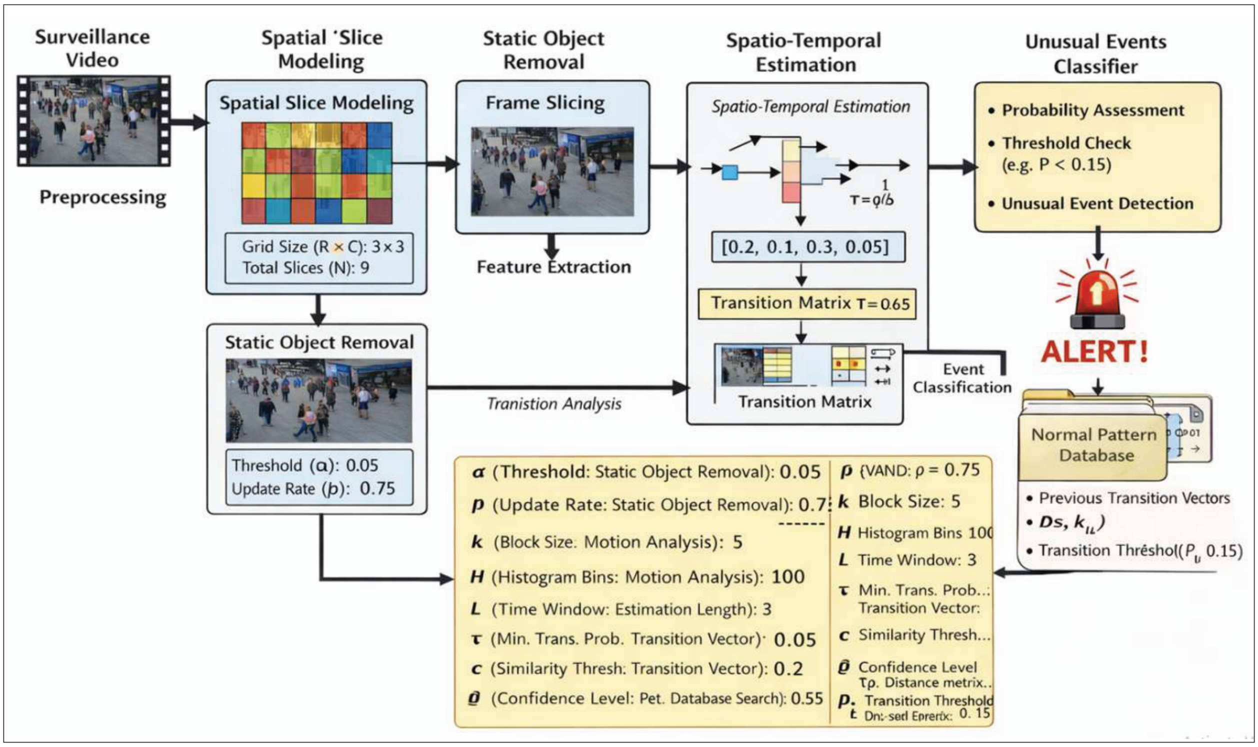

Figure 1: Proposed architecture

After that, immobile elements and background regions are removed using static object removal, ensuring that only significant dynamic regions are analyzed further. The analysis of frame-to-frame transitions from these processed slices yields spatiotemporal features, which are then modeled using a Markov-based nearest-neighbor transition technique to create compact feature vectors that depict local temporal dynamics. To differentiate between regular and abnormal occurrences, these feature vectors are then passed to the unusual occurrences classifier, which calculates transition probabilities and confidence scores. To balance detection accuracy and false alarm rates, rule-based thresholds, such as transition probability limits and confidence levels, are selected via empirical tuning on validation subsets of benchmark datasets. Future developments might include adaptive or data-driven threshold optimization, even if fixed thresholds guarantee computational effectiveness and real-time applicability. It would be easier to comprehend and give a clear visual grasp of module interactions if a flowchart showed this entire processing pipeline, from raw input frames to the final anomaly judgment.

Spatial slice model

To reduce feature comparison across frames, the frame sequence is segmented (Bird *et al.,* 2005) into equally spaced k vertical slices, where k = 1 to N, and then slice-wise feature-based normalization is carried out. The features such as colour c_i, texture t_i, and edge e_i are derived on each slice, where $i = 1$ to n. The measurement of c_i, t_i, and e_i are used to efficiently describe the spatial information whether the foreground region moves slightly or abruptly. To estimate the data point d_p, where p = 1 to N, the slices are used to extract the c_i, t_i, and e_i features. Then, the slice-wise mean m is computed by dividing each data point value by the sum of the slice-wise data point feature values, since each data point summarizes the slices. A d_p can reflect the centre of the c_i, t_i, and e_i using the m parameter. The d_p will be derived on each slice using all three features c_i, t_i, and e_i. The resultant d_p can be explained by using the eqn. (1). The frame slices are shown in Figure 2.

$$m \text{ of } d_p = \frac{c_1+c_2+.......,+c_n}{\text{number of } c_i \text{ in the } k^{th} \text{ slice}}, \quad \frac{t_1+t_2+.......,+t_n}{\text{number of } t_i \text{ in the } k^{th} \text{ slice}}, \quad \frac{e_1+e_2+.......,+e_n}{\text{number of } e_i \text{ in the } k^{th} \text{ slice}}, \qquad ...(1)$$

where $c_1, c_2, ..c_n$, t_1, $t_2,...t_n$, and e_1, $e_2,....,e_n$ are feature values of data points.

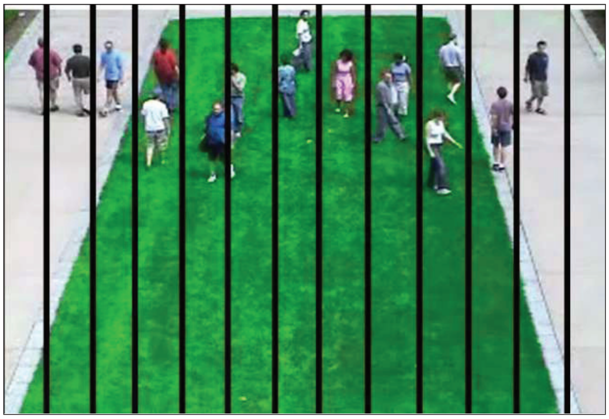

Figure 2: The Spatial Slice Model

Removal static objects

To reduce computational and feature-comparison complexity, the static objects are located on the slice and indicated as '0' in the further process, since the moving objects can only express event types such as normal and unusual events (Vennila & Balamurugan, 2023). In general, the density based unusual events detection method cannot efficiently locate the foreground objects. To overcome this challenge, this study used a foreground detection method for identifying moving objects' foreground slices, using d_p of c_i, t_i, and e_i. Let d_q be the data points of the consecutive frame where $q = 1$ to N. To identify the moving foreground regions, the d_{pi}'s were compared with the corresponding d_q. The threshold values are applied to the difference between d_p and d_q. Let α, β, and γ be the threshold values on the difference values of colour, texture, and edge respectively. If the difference meets all three thresholds α, β, and γ, it is considered a moving object slice. Let d_j be the moving object slice, where j = 1 to K. These are explained using the Eqns. (2), (3), and (4).

$$d_j \text{ for color } = \begin{cases} \text{Yes,} & \text{if } (d_p - d_q) < \alpha \\ & \text{where } p \text{ and } q = 1 \text{ to } N \\ \text{No} & \text{Otherwise} \end{cases} \qquad ...(2)$$

$$d_j \text{ for texture } = \begin{cases} \text{Yes,} & \text{if } (d_p - d_q) < \beta \\ & \text{where } p \text{ and } q = 1 \text{ to } N \\ \text{No} & \text{Otherwise} \end{cases} \qquad ...(3)$$

$$d_j \text{ for edge } = \begin{cases} \text{Yes,} & \text{if } (d_p - d_q) < \gamma \\ & \text{where } p \text{ and } q = 1 \text{ to } N \\ \text{No} & \text{Otherwise} \end{cases} \qquad ...(4)$$

To efficiently detect the moving foreground region, the slices with colour, texture, and edge features should have their thresholds set using α, β, and γ. The moving foreground slices d_j can be selected using three threshold values and three features.

Spatio-temporal estimation on spatial slice

The moving foreground region slices with d_j of the frame sequence can be completely extracted by removing static object computation. To find the temporal information between frames, the orientation is used to explain the moving objects since the speed of the d_j can have a significant influence on calculating the displacement between the d_j of consecutive frames. The process of orientation is to estimate the position of the d_j. In general, the orientation of the d_j value does not change, even when the intensity values change due to external defects. Based on this assumption, this paper used Angular Velocity (AV) to obtain the efficient speed of the d_j. It refers to how fast the orientation of the *dj* changes over time. AV is obtained using gradients in the x and y directions. It is detailed using Eqn. (5).

$$tan \ \theta = \frac{G_y}{G_x} \qquad ...(5)$$

where G_x and G_y are the gradient values on x and y directions, respectively. Then, each orientation of the d_j is computed by subtracting the corresponding slice's orientation in the current frame (f_s) and the consecutive frame (f_{s+1}). It is represented as $d\theta$. AV needs time information to calculate the change in the d_j's orientation over a given period. The time t is computed using optical flow computation. It is obtained by estimating the difference between the corresponding slice t in the current frame and the consecutive frame; it is denoted as dt. The AV is explained in Eqn. (6).

$$AV = \frac{d\theta}{dt} \qquad ...(6)$$

The collection of AV of slices on the frame sequence is explained in the Eqn. (7). Let av_h be the collections of AV where $h = 1$ to N.

$$f_s = \{av_1, av_2, \ldots\ldots av_N\} \qquad ...(7)$$

Unusual events detection

Let A be all the learned frames. A consists of normal and unusual events frames. The traditional classifier usually finds the distance between the new input frame and all learned frames (Song et al., 2019). Then the classifier finds the nearest training frame. Finally, the algorithm decides whether an event is normal or unusual based on the event's nearest distance value. By considering all learned frames for estimating the nearest distance, the classifier will incur greater computational complexity. To avoid this problem, this paper proposed MTUEC to efficiently reduce feature comparison between the training and testing phases to detect normal or unusual events using f_s.

In general, normal or abnormal events occur in continuous frames within a certain range. Based on this assumption, the MTUEC is developed to analyze the immediately prior learned frame. In order to firstly select the immediate previous learned frame to find the nearest distance, Markov-nearest transition classifier is used, in which Markov transition (Vennila & Balamurugan, 2020) is a process such that the choices of learned frames is only made on the previous learned frame. The MTUEC analyzes the distance of each d_j between training and testing instance. If the distance is less than σ, the certain d_j is considered as similar values and it is stored in the B vector. After finding all the distances d_j between training and testing phases, the MTUEC analyses the B. Let Ω be the variation value of distance among the incoming testing frame and the previous learned frame. If the distance variation is less than the Ω value in B, the MTUEC needs to check all learned frames until it gets Ω, since the normal or unusual events type may change into another type from the previous learned frame. The MTUEC procedure is discussed in the Algorithm 1.

Algorithm 1: MTUEC

```
Input: N frames, $\sigma$, and $\Omega$
Output: normal or unusual events
For $s = 1$ to $N$
Develops Spatial Slice Model
Computes Removal Static Objects
Estimates Spatio-Temporal on Slice
        For $j = 1$ to $K$
            If the distance $<\sigma$
            $B = 1$
        Else
            $B = 0$
            End
        End
        Find sum($B$)
If sum($B$)>$\Omega$
s is the unusual events frame
        $r = 1$(where $r$ is the first learned frame)
            For $r = 1$ to $N$
            If the distance $<\sigma$
                $B = 1$
            If sum($B$) >$\Omega$
                s is the unusual events frame
        Else
                $B = 0$
            End
                $k = k+1$
            End
        End
        End
End
```

The proposed MTUEC-based odd-event detection system's entire processing pipeline is depicted in the flowchart in Figure 3. To correct illumination effects, the input surveillance video is first split into individual frames and preprocessed. To record localized motion patterns, each frame is then separated into uniform spatial slices. To ensure that only dynamic forefront information is preserved, adaptive thresholding is used in conjunction with static object removal to remove background regions. Compact transition vectors are produced by applying a Markov-nearest transition model to analyze frame-to-frame transitions within a predetermined temporal window to perform spatiotemporal estimation. A transition matrix of probabilities is created using these vectors and compared to previously learn normal patterns. Lastly, an alarm is triggered when anomalous behaviour is recognized by a threshold-based decision system that classifies events as either normal or abnormal.

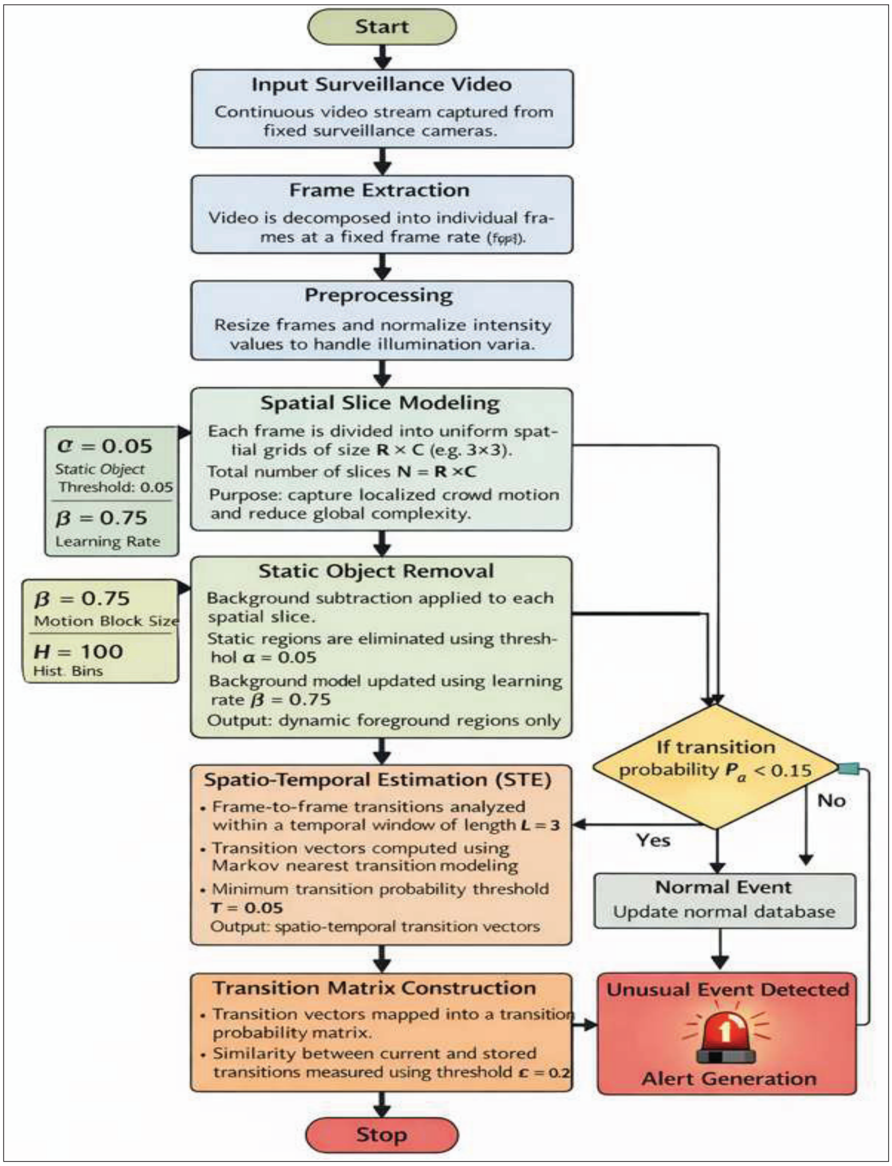

Figure 3: Flowchart of proposed model

RESULTS AND DISCUSSION

Based on the above constraints, the proposed MTUEC work was implemented, and the results were verified using MATLAB 13a. Unusual event detection benchmark datasets, such as UMN (https://UMN-dataset-anomaly-in-crowd_fig3_270790763) and UCSD Ped1 and UCSD Ped2 (http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm), were used to evaluate the performance of the proposed MTUEC. Also, the results of several unusual event detection algorithms were used to compare with the proposed work's results to estimate the effectiveness of the proposed MTUEC work.

Dataset description

UMN dataset: the dataset consists of three scenes. Different environments are captured all three scenes. All three scenes feature crowd density, followed by unusual crowd activities such as people running from the crowd, people merging, and people spreading.

USD dataset: the dataset was created using a stationary camera placed at an elevation, capturing pedestrian walkways. In the dataset, most frames contain crowd density information. The crowd density occurs across the frame sequence, spanning from sparse to dense crowds. Also, the dataset contains people in normal scenarios. On the other hand, people with high motion and non-people entries in the pedestrian walkways are observed. Non-people anomalies, such as small carts, bikers, and skaters, appear in the frame sequence. The normal and unusual events were naturally occurring during the creation of the dataset. The Ped1 consists of 34 training and 36 testing videos and Ped2 consists of 16 training and 12 testing videos. Each video consists of 200 frames. The ground-truth frames are available in the dataset, which consists of unusual events.

Additional real-world data: To further validate the robustness of MTUEC, we tested the algorithm on 10 real-world surveillance videos collected from publicly available urban surveillance footage. These videos, totaling 20 minutes, included a variety of crowd behaviours such as normal pedestrian movement, sudden dispersal, and sparse-to-dense transitions.

Performance evaluation

The frame-level evaluation is used during both normal and unusual events. The Precision, Recall, F1-score, AUC, and EER measures are used to evaluate the proposed MTUEC work performance. F1-score is used to measure the test accuracy. An F1-score of 100% indicates the best test accuracy. An F1-score of 0% indicates the worst test accuracy. These measures are estimated using true positive (TP), false positive (FP), false negative (FN), true positive rate (TPR), and false positive rate (FPR). TP denotes correctly detected unusual crowd activities frames, FP denotes incorrectly detects the normal frame as unusual crowd activities frame, FN denotes the incorrectly detected unusual crowd activities frame as normal frame, the true positive rate (TPR) is defined as the proportion of correctly identified positive frames relative to the total number of actual positive frames. The false positive rate (FPR), on the other hand, represents the proportion of falsely identified positive frames out of the total number of actual negative frames. They are described in the Eqns. (12), (13), (14), (15) and (16).

$$Precision = \frac{TP}{TP + FP}, \quad ...(12)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad ...(13)$$

$$F1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad ...(14)$$

$$FPR = \frac{\text{Number of False positive Frames}}{\text{Number of negative Frames}} \quad ...(15)$$

$$TPR = \frac{\text{Number of True positive Frames}}{\text{Number of positive Frames}} \quad ...(16)$$

Table 1: Experimental result on UCSD (Ped1 and Ped2) dataset

Video No.	UCSD Ped 1			UCSD Ped 2		
	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)
Video 1	93.00	60.00	73.00	91.00	55.00	69.00
Video 2	87.00	64.00	74.00	89.00	60.00	72.00
Video 3	90.00	74.00	81.00	92.00	71.00	80.00
Video 4	93.00	71.00	80.00	96.00	74.00	84.00
Video 5	89.00	79.00	83.00	93.00	75.00	83.00

Table 2: Experimental results on UMN dataset

Scene No.	Precision (%)	Recall (%)	F1-score (%)
Scene 1	88.00	66.00	72.00
Scene 2	86.00	74.00	80.00
Scene 3	91.00	88.00	89.00

Comparative analysis

The results on UCSD Ped1, Ped2, and UMN datasets are summarized in Tables 1 and 2, showing that MTUEC outperforms existing methods in precision and recall while reducing computational complexity. Existing unusual event detection algorithms, such as IGMM

(Gnouma *et al*., 2018), PNN (Lohithashva *et al.,* 2018), CNN (Direkoglu, 2020), GLCM (Lalit & Purwar, 2022), and OPLKT-EMEHO (Rajasekaran & Sekar, 2023), are used for comparison with the proposed MTUEC. To evaluate the effectiveness of the proposed MTUEC algorithm, various approaches to unusual event detection, including machine learning-based methods, are used. The comparative analyses are shown in Tables 3, 4, and 5.

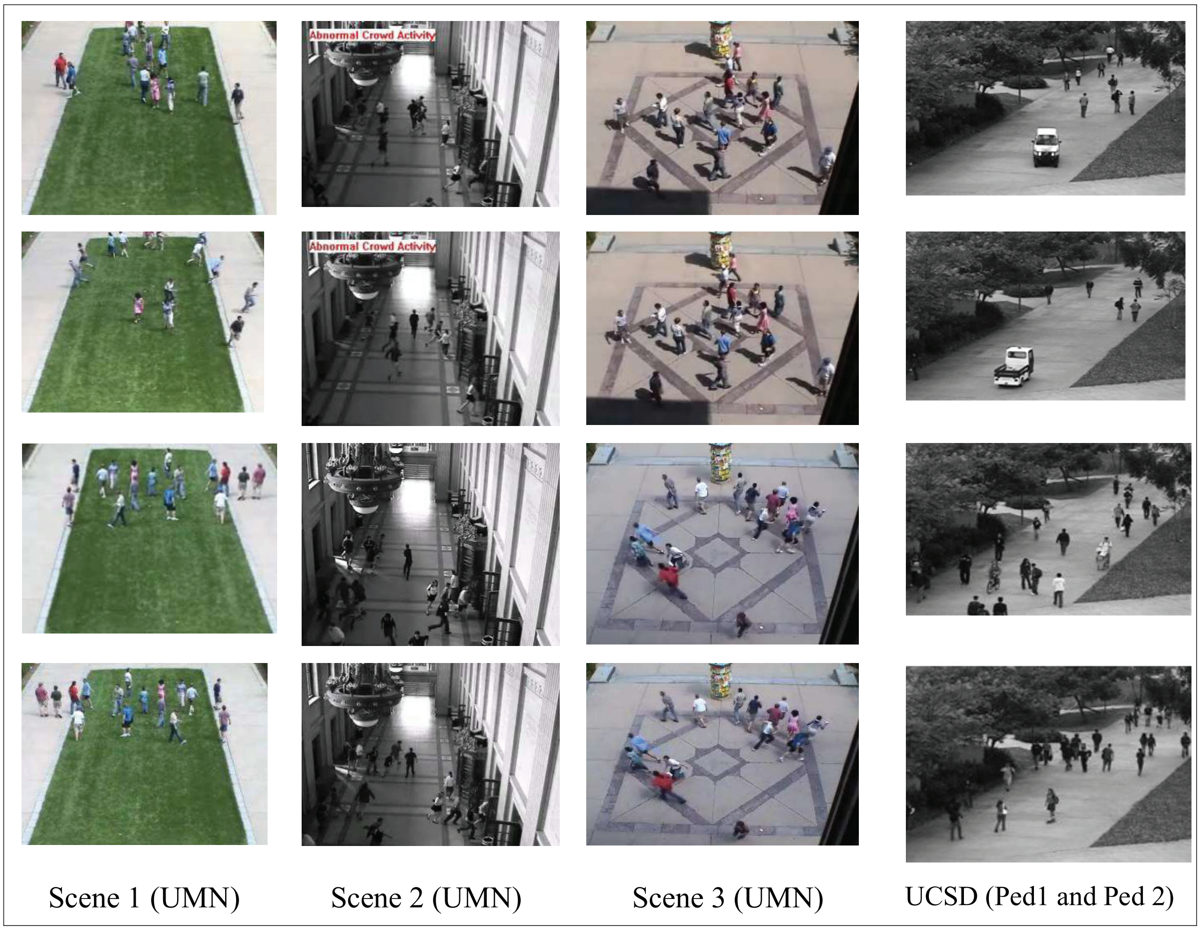

Figure 4: Examples of unusual crowd events and unusual moving objects

Table 3: Comparative analysis on UMN (Scene 1) dataset

Algorithms	Years	AUC	EER
IGMM	2018	79.20%	41.01%
PNN	2019	82.40%	15.00%
CNN	2020	85.00%	17.00%
GLCM	2022	90.00%	12.00%
OPLKT-EMEHO	2023	92.00%	13.00%
Proposed MTUEC		93.50%	14.00%

Table 4: Comparative analysis on UMN (Scene 2) dataset

Algorithms	Years	AUC	EER
IGMM	2018	81.25%	24.00%
PNN	2019	84.20%	13.50%
CNN	2020	86.00%	14.00%
GLCM	2022	91.10%	13.20%
OPLKT-EMEHO	2023	93.40%	13.10%
Proposed MTUEC		94.00%	15.00%

Table 5: Comparative analysis on UMN (Scene 3) dataset

Algorithms	Years	AUC	EER
IGMM	2018	85.00%	18.20%
PNN	2019	87.10%	14.20%
CNN	2020	88.20%	13.10%
GLCM	2022	92.00%	12.00%
OPLKT-EMEHO	2023	94.00%	12.00%
Proposed MTUEC		95.10%	13.30%

Table 6: Performance on real-world surveillance Footage

Scenario	Precision (%)	Recall (%)	F1-score (%)	Observations
Normal crowd movement	92.0	87.0	89.4	High accuracy in detecting normal events
Sudden unusual movement	90.5	85.0	87.6	Successfully detected running, dispersal, and merging
Sparse-to-dense transition	85.0	72.0	78.0	Some missed anomalies during gradual crowd build-up
Minimal crowd movement	80.0	68.0	73.2	Difficulty in detecting anomalies in static crowd
Small object anomalies (*e.g.*, bikers, carts)	88.0	75.0	81.0	Some misclassification due to size and speed

From Tables 3, 4, and 5, the proposed MTUEC algorithm yields better performance than the existing algorithms across measures such as precision, recall, F1-score, AUC, and EER. To reduce computational and feature-comparison complexity, the proposed MTUEC algorithm is developed using the following modules: spatial slice model, removal of static objects computation, spatio-temporal estimation, and Marko-nearest transition based unusual events classifier. The above modules help reduce computational and feature-comparison complexity because the features of the frame sequence are computed on the slice, not on each frame element. Then, the MTUEC analyzes the feature comparison between the immediate training frames with the testing frame. If the immediate training frame does not match with testing frame only, the MTUEC algorithm considers the other training frame for comparison. The MTUEC algorithm helps detect normal or unusual events effectively at the first iteration, since the frame sequence contains them continuously within a specific range. Though the MTUEC algorithm works in almost all cases, such as unusual crowds with more moving objects and unusual activities with more multiple objects suddenly moving from the crowd, it does not work well in cases of crowds with fewer moving objects and crowd spreading with fewer moving objects. The example of unusual events in the frame is shown in Figure 4.

To further validate the robustness of MTUEC, we tested the algorithm on 10 real-world surveillance videos collected from websites. These 20-minute videos spanned a variety of crowd behaviours, allowing us to assess the model's adaptability beyond standard benchmark datasets. The model successfully detected anomalies in most cases, particularly sudden, unusual movements and small-object anomalies. However, some limitations were observed in scenarios with minimal crowd movement or during sparse-to-dense transitions, resulting in a slight drop in recall. The detailed evaluation is presented in Table 6, highlighting both strengths and areas for improvement.

To further validate the robustness of MTUEC, we tested the algorithm on a small set of real-world surveillance footage (outside the training datasets). The model successfully detected anomalies in most cases, particularly sudden, unusual movements and small-object anomalies. However, some limitations were observed in scenarios where crowd movement was minimal or when a sparse-to-dense transition occurred, leading to a slight drop in recall performance. The detailed evaluation of MTUEC on real-world surveillance footage is presented in Table 6, highlighting both strengths and areas for improvement.

Real-time feasibility

The suggested approach significantly reduces the computational burden of exhaustive frame-wise feature evaluations by introducing a novel and useful Markov-nearest frame comparison algorithm. The suggested modular design, which includes spatial slicing and lightweight spatio-temporal estimation using angular velocity, provides an innovative and effective solution ideal for real-time surveillance applications, in contrast to deep learning-centric approaches that rely on intricate architectures and substantial training. Low-latency processing is enabled by a focus on local frame transitions and low memory utilization, making the method especially appealing for contexts with limited resources. A thorough examination of deployment-related aspects, such as per-frame processing time, memory footprint, adaptability to camera resolution, and robustness under changing illumination and crowd density, remains a crucial area for future research, even though the system is theoretically applicable to public surveillance scenarios like transit hubs and congested public areas. Including these assessments would increase the suggested framework's viability and preparedness for the real world.

Positive impacts

Compared with current monitoring systems, the proposed MTUEC framework offers several operational advantages that enhance its practicality. The method greatly reduces computational load by using Markov-nearest transitions to reduce exhaustive frame-wise feature comparisons. This allows for faster incident identification and minimizes delay in real-time monitoring applications. Instead of continuously monitoring several video feeds, security staff can be instantly alerted to unexpected events through reduced processing complexity, thereby directly reducing operator workload. The suggested method can be incorporated into existing CCTV and smart surveillance systems that use conventional CPU-based infrastructure, unlike deep learning-based surveillance software, which requires expensive GPUs, substantial memory footprints, and offline training. Additionally, the system's modular design and lightweight spatio-temporal estimation enable scalability across multiple cameras, making it suitable for deployment in expansive public areas such as shopping malls, train stations, and airports. Together, these features strengthen the suggested method's deployment potential and practical significance by positioning it as an affordable, low-latency substitute for existing surveillance technologies.

Ethical considerations and privacy implications

There are significant privacy, ethical, and societal issues with the use of automated systems for anomalous event detection in public monitoring settings. Although the suggested framework uses motion-based spatiotemporal features rather than biometric identifiers, caution must be taken to ensure compliance with data protection laws such as the general data protection regulation (GDPR). Video data should be anonymised, access-controlled, and retained for a limited period in practical deployments, in compliance with institutional and regulatory standards. Furthermore, automated anomaly detection systems are prone to false positives, which can lead to erroneous warnings or skewed perceptions of typical crowd behaviour across diverse cultural and environmental contexts. These biases may result from scene-dependent dynamics or motion patterns unique to a dataset. The suggested system should be used as a decision-support tool rather than as a fully autonomous authority to reduce these risks, and the alarm verification process should include human oversight. To improve ethical robustness and responsible deployment, future research will concentrate on integrating adaptive thresholding, bias-aware evaluation across various datasets, and privacy-preserving processing techniques.

CONCLUSION

This paper introduced the Markov-nearest transition unusual event classifier (MTUEC) algorithm, aimed at reducing the computational complexity and feature comparison time during the detection of unusual crowd events. The MTUEC algorithm was developed using several integrated modules, including the spatial slice model, static object removal, spatio-temporal estimation, and the Markov-nearest transition-based classifier. Our experimental results demonstrate that the MTUEC algorithm performs effectively in most scenarios, especially in cases involving unusual crowds with numerous moving objects and sudden changes in crowd movement patterns.

However, the algorithm exhibits limitations in scenarios with sparse crowds or fewer moving objects, where detecting unusual events becomes more challenging. Additionally, its performance may be affected by noisy or low-quality video inputs. Future work could focus on enhancing the algorithm's robustness to sparse crowds and its adaptability across diverse surveillance settings. Furthermore, the proposed method shows potential for application in various real-world surveillance tasks, such as monitoring public spaces to help prevent crowd collisions and ensuring safety during mass gatherings.

REFERENCES

Ahmed, S. A., Dogra, D. P., Kar, S., & Roy, P. P. (2018). Trajectory-based surveillance analysis: A survey. *IEEE transactions on circuits and systems for video technology*, *29*(7), 1985-1997. https://doi.org/10.1109/TCSVT.2018.2857489

Al-Khazaleh, S., Badwan, N., & Saleh, Q. (2026). Digital economy prosperity analysis based on the Grey–Markov model: Empirical evidence from MENA region. *International Journal of Islamic and Middle Eastern Finance and Management, 19*, 269–300. https://doi.org/10.1108/IMEFM-07-2024-0348

Anjum, N., & Cavallaro, A. (2008).Multifeature object trajectory clustering for video analysis. *IEEE Transactions on Circuits and Systems for Video Technology*, *18*(11), 1555-1564. https://doi.org/10.1109/TCSVT.2008.2005603

Beigh, T. M., Venkatesan, V. P., & Arumugam, J. (2025). Predictive summarization framework for resource-constrained device surveillance videos. *Pattern Recognition*, 112274. https://doi.org/10.1016/j.patcog.2025.112274

Bird, N. D., Masoud, O., Papanikolopoulos, N. P., & Isaacs, A. (2005). Detection of loitering individuals in public transportation areas. *IEEE Transactions on Intelligent Transportation Systems*, *6*(2), 167-177. https://doi.org/10.1109/TITS.2005.848370

Biswas, S., & Babu, R. V. (2017). Anomaly detection via short local trajectories. *Neurocomputing*, *242*, 63-72. https://doi.org/10.1016/j.neucom.2017.02.058

Chen, H., Zhao, X., Wang, T., Tan, M., & Sun, S. (2016, June). Spatial-temporal context-aware abnormal event detection based on incremental sparse combination learning. In *2016 12th World Congress on Intelligent Control and Automation (WCICA)* (pp. 640-644). IEEE. https://doi.org/10.1109/WCICA.2016.7578533

Cho, J., & Kang, Y. (2025). Context-Aware Anomaly Detection of Pedestrian Trajectories in urban back streets using a variational autoencoder. *ISPRS International Journal of Geo-Information*, *14*(11), 438. https://doi.org/10.3390/ijgi14110438

Chong, Y. S., & Tay, Y. H. (2017, May). Abnormal event detection in videos using spatiotemporal autoencoder. In *International symposium on neural networks* (pp. 189-196). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-59081-3_23

Chu, W., Xue, H., Yao, C., & Cai, D. (2018). Sparse coding guided spatiotemporal feature learning for abnormal event detection in large videos. *IEEE Transactions on Multimedia*, *21*(1), 246-255. https://doi.org/10.1109/TMM.2018.2846411

Cong, Y., Yuan, J., & Liu, J. (2013).Abnormal event detection in crowded scenes using sparse representation. *Pattern Recognition*, *46*(7), 1851-1864. https://doi.org/10.1016/j.patcog.2012.11.021

Direkoglu, C. (2020). Abnormal crowd behavior detection using motion information images and convolutional neural networks. *IEEE Access*, *8*, 80408-80416. https://doi.org/10.1109/ACCESS.2020.2990355

Gnouma, M., Ejbali, R., & Zaied, M. (2018). Abnormal events' detection in crowded scenes. *Multimedia Tools and Applications*, *77*(19), 24843-24864. https://doi.org/10.1007/s11042-018-5701-6

Ilyas, A., & Bawany, N. (2025). Crowd dynamics analysis and behavior recognition in surveillance videos based on deep learning. *Multimedia Tools and Applications*, *84*(23), 26609-26643. https://doi.org/10.1007/s11042-024-20161-7

Jadhav, S., & Bartere, M. (2025).Generalizable and privacy-preserving multimedia forgery detection via semantically disentangled and temporally-aware deep learning architectures. *International Journal of Information Technology, 1*, 1–15. https://doi.org/10.1007/s41870-025-02978-z

Jency, A., & Ramar, K. (2025). A Review of abnormal behaviour detection in crowd for video surveillance: advances and trends, datasets, opportunities and prospects. *Expert Systems*, *42*(4), e70013. https://doi.org/10.1111/exsy.70013Digital Object Identifier

Kim, J. S., Yeom, D. H., & Joo, Y. H. (2011). Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems. *IEEE Transactions on Consumer Electronics*, *57*(3), 1165-1170. https://doi.org/10.1109/TCE.2011.6018870

Lalit, R., & Purwar, R. K. (2022). Crowd abnormality detection using optical flow and GLCM-based texture features. *Journal of Information Technology Research (JITR)*, *15*(1), 1-15. https://doi.org/10.4018/JITR.2022010110

Li, C., Wang, X., Wang, X., Chen, S., Ye, J., & Zhao, S. (2025, June). Pattern Learning for Trajectory Anomaly Detection: Recent Developments and Emerging Challenges. In *2025 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-10).IEEE.DOI: 10.1109/IJCNN64981.2025.11227964

Liang, R., Li, Y., Wu, Z., & Li, X. (2025). An interaction-scene collaborative representation framework for detecting traffic anomalies in driving videos. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2025.3553830

Lin, W., Zhang, Y., Lu, J., Zhou, B., Wang, J., & Zhou, Y. (2015). Summarizing surveillance videos with local-patch-learning-based abnormality detection, blob sequence optimization, and type-based synopsis. *Neurocomputing*, *155*, 84-98. https://doi.org/10.1016/j.neucom.2014.12.044

Lohithashva, B. H., Manjunath Aradhya, V. N., Basavaraju, H. T., & Harish, B. S. (2018, December). Unusual crowd event detection: an approach using probabilistic neural network. In *Information Systems Design and Intelligent Applications: Proceedings of Fifth International Conference INDIA 2018 Volume 1* (pp. 533-542). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-13-3329-3_50

Mancy, H., & Naith, Q. H. (2025). SwinIoT: A hierarchical transformer-based framework for behavioral anomaly detection in IoT-Driven smart cities. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3551207

Mou, H., Rong, H., & Teixeira, A. P. (2026). Detecting abnormal ship trajectory to avoid bridge collisions via a Transformer-BiLSTM model. *Ocean Engineering*, *343*, 123232. https://doi.org/10.1016/j.oceaneng.2025.123232

Paul, M. A., Kumar, K. S., Sagar, S., & Sreeji, S. (2023). LWDS: lightweight DeepSeagrass technique for classifying seagrass from underwater images. *Environmental Monitoring and Assessment*, *195*(5), 614. https://doi.org/10.1007/s10661-023-11183-z

Rajasekaran, G., & Sekar, J. R. (2023). Abnormal Crowd Behavior Detection Using Optimized Pyramidal Lucas-Kanade Technique. *Intelligent Automation & Soft Computing*, *35*(2). https://doi.org/10.32604/iasc.2023.029119

Song, H., Sun, C., Wu, X., Chen, M., & Jia, Y. (2019).Learning normal patterns via adversarial attention-based autoencoder for abnormal event detection in videos. *IEEE Transactions on Multimedia*, *22*(8), 2138-2148. https://doi.org/10.1109/TMM.2019.2950530

Sundaram, D. M., Mannam, S., & Jayachandran, S. (2025, August). Intelligent Video Surveillance for Human Behavior Analysis and Anomalous Activity Detection using Deep Learning. In *2025 5th International Conference on Soft Computing for Security Applications (ICSCSA)* (pp. 950-955). IEEE. https://doi.org/10.1109/ICSCSA66339.2025.11171248

Vallejo, D., Albusac, J., Jimenez, L., Gonzalez, C., & Moreno, J. (2009).A cognitive surveillance system for detecting incorrect traffic behaviors. *Expert Systems with Applications*, *36*(7), 10503-10511. https://doi.org/10.1016/j.eswa.2009.01.034

Vennila, T. J., & Balamurugan, V. (2020, February).A stochastic framework for keyframe extraction. In *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)* (pp. 1-5). IEEE. https://doi.org/ 10.1109/ic-ETITE47903.2020.294.

Vennila, T. J., & Balamurugan, V. (2023). A rough set framework for multihuman tracking in surveillance video. *IEEE Sensors Journal*, *23*(8), 8753-8760. https://doi.org/10.1109/JSEN.2023.3242007

Yang, Q., Wang, C., Liu, P., Jiang, Z., & Li, J. (2025). Video anomaly detection via self-supervised and spatio-temporal proxy tasks learning. *Pattern Recognition*, *158*, 111021. https://doi.org/10.1016/j.patcog.2024.111021