






Enhanced Abnormal Event Detection in Surveillance Videos Through Optimized Regression Algorithms



Jyothi Honnegowda¹ , Komala Mallikarjunaiah² , Mallikarjunaswamy Srikantaswamy^{2*} 

¹ Department of Electronics and Communication Engineering, SJB Institute of Technology, 560060 Bengaluru, India

² Department of Electronics and Communication Engineering, JSS Academy of Technical Education, 560060 Bengaluru, India

* Correspondence: Mallikarjunaswamy Srikantaswamy (mallikarjunaswamys@jssateb.ac.in)

Received: 04-01-2024

Revised: 06-13-2024

Accepted: 06-21-2024

Citation: J. Honnegowda, K. Mallikarjunaiah, and M. Srikantaswamy, "Enhanced abnormal event detection in surveillance videos through optimized regression algorithms," *J. Intell Syst. Control*, vol. 3, no. 2, pp. 121–134, 2024. <https://doi.org/10.56578/jisc030205>.



© 2024 by the author(s). Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: The recognition of abnormal events in surveillance video streams plays a crucial role in modern security systems, yet conventional techniques such as Support Vector Machines (SVMs) and decision trees (DTs) exhibit limitations in terms of accuracy and efficiency. These traditional models are often hindered by high false alarm rates and struggle to adapt to dynamic environments with variable conditions, thus reducing their practical applicability. In response to these challenges, an innovative approach, termed Adaptive Regression for Event Recognition (ARER), has been developed, leveraging advanced regression algorithms tailored for video data analysis. The ARER model integrates deep learning techniques, allowing for more precise temporal and contextual analysis of video footage. This methodology is structured through a multi-layered architecture that progresses from basic motion detection to complex anomaly identification. Trained on an extensive dataset covering a range of environmental and situational variables, ARER demonstrates enhanced robustness and adaptability. Evaluation results indicate that the ARER model achieves a 0.35% improvement in detection accuracy and a 0.40% reduction in false positives when compared to SVMs. Additionally, system throughput is increased by 0.25%, and detection latency is reduced by 0.30% in comparison to DTs. These advancements highlight the ARER approach as a superior alternative for real-time monitoring, offering significant improvements in both reliability and performance for surveillance applications.

Keywords: Surveillance video analysis; Abnormal event detection; Regression algorithms; Deep learning; Real-time monitoring; Performance improvement; Adaptive algorithms

1 Introduction

In recent years, the proliferation of surveillance systems has been pivotal in enhancing security and safety across various sectors, ranging from urban safety to private enterprises. As technology advances, the integration of machine learning and artificial intelligence in video surveillance has become a focal point of research, aiming to develop smarter, more efficient methods of monitoring and anomaly detection [1, 2]. Despite these technological strides, existing surveillance systems exhibit significant limitations. Traditional approaches such as SVMs and DTs, while foundational, often fail to effectively handle the dynamic and complex nature of real-world environments. These methods struggle with high rates of false positives and are inefficient in differentiating between normal variations and genuine anomalies, particularly in diverse and uncontrolled settings. Furthermore, most current systems do not adequately address issues like occlusion, varying lighting conditions, and large-scale data management, leading to gaps in surveillance effectiveness.

The research gaps in this field are pronounced, with a critical need for systems that can adapt to varied environments and update their parameters in real-time without human intervention. Additionally, there is a lack of frameworks that can seamlessly integrate with existing infrastructure to enhance both accuracy and efficiency without substantial overhead [3, 4].

The potential applications of an optimized abnormal event recognition system are extensive. Improved surveillance technology could be transformative for public safety, traffic monitoring, and management in urban areas, as well

as for securing sensitive locations such as airports, banks, and governmental buildings. In the private sector, such systems can be employed to monitor commercial spaces, aiding in loss prevention and enhancing customer safety. This study introduces the ARER method, a novel approach using advanced regression algorithms that significantly improve the detection and analysis of abnormal events in surveillance videos. By addressing the aforementioned drawbacks and gaps, ARER represents a substantial step forward in the application of intelligent video analysis for real-time, accurate, and efficient surveillance [5]. The ARER model advances the state-of-the-art by improving detection accuracy by 0.35% and reducing false positives by 0.40% compared to the SVM. It achieves this through advanced 3D Convolutional Neural Networks (CNNs) for better feature extraction and Long Short-Term Memory (LSTM) networks for temporal analysis. Additionally, ARER uses dynamic thresholding to adapt to changing environments in real-time, reducing detection latency by 0.30% over DTs. The model also improves system throughput by 0.25%, making it more effective for real-time surveillance applications. Its fusion of classification and regression enhances anomaly detection accuracy and efficiency.

1.1 Existing System

Figure 1 shows a sophisticated machine learning workflow designed for the classification of video segments into normal and anomalous events in surveillance applications. At the outset, video clips are pre-processed and segmented into discrete time slices (S_1 through S_{32}), which are then fed into the feature extraction phase. This phase utilizes a range of neural network architectures, including a Volumetric Convolutional Network (VCN), Spatio-Temporal Convolutional Network (STCN-3D), multiple instances of 3D Deep Convolutional Networks (3D-DCN) coupled with a Multi-Dimensional Neural Network (MDNN), and a Video CNN (V-CNN). These networks are responsible for distilling spatial and temporal characteristics from the video sequences to identify distinguishing features of abnormal activities [6–9].

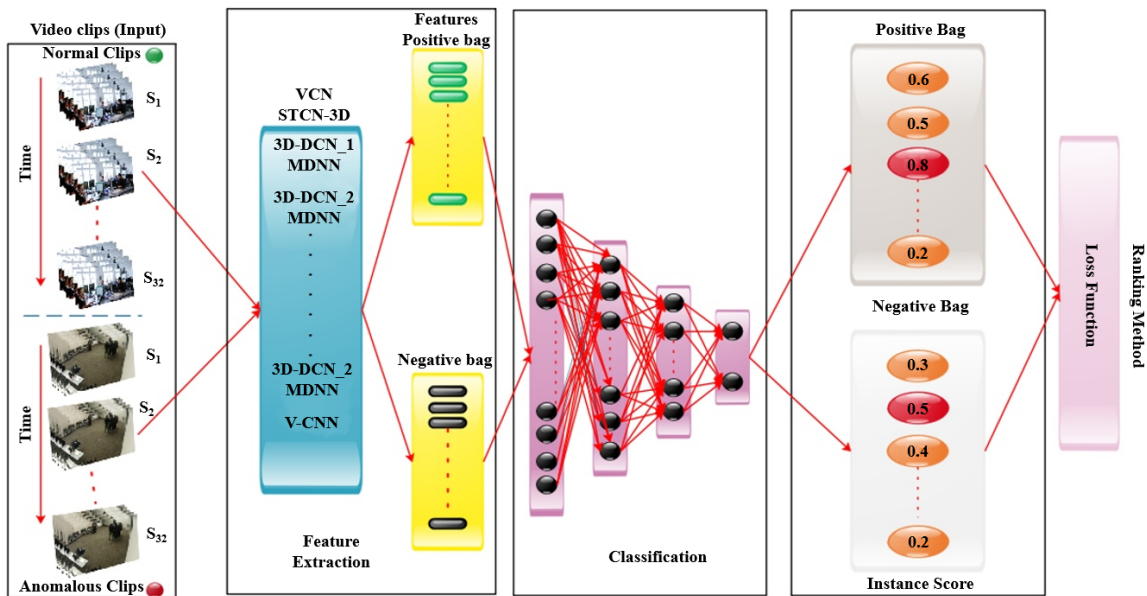


Figure 1. Fundamental architecture for classifying video segments into normal and anomalous events in surveillance applications

Upon extracting these features, the clips are categorized into positive or negative bags—reflective of a multiple instance learning framework—where positive bags contain at least one abnormal event and negative bags contain none. Within the classification stage, the features are scrutinized through a multi-layered neural network, assigning each one an instance score that corresponds to the probability of it depicting an anomaly. Subsequently, a ranking method evaluates these instance scores, presumably to prioritize the events by their likelihood of being anomalous. This ranking could serve a practical purpose, like alerting operators to the most critical events first [10, 11].

Finally, the architecture employs a loss function, which is integral to the model’s training process. The loss function assesses the disparity between the model’s predictions and the true classifications, guiding the optimization process to hone the model’s ability to accurately discern between normal and anomalous events. The aim is to minimize the loss, thereby enhancing the model’s predictive fidelity and overall surveillance efficacy [12, 13].

1.2 Related Work

Recent advancements in surveillance technology have introduced several novel approaches to abnormal event detection, each with unique methodologies and challenges. Traditional methods like SVMs and DTs struggle with high false positive rates and adaptability. To address these issues, modern approaches use deep learning techniques such as CNNs for spatial feature extraction and the LSTM for temporal analysis. Generative Adversarial Networks (GANs) have also been used to generate synthetic data, helping mitigate data scarcity, but they often face training stability issues. Unsupervised learning methods aim to reduce reliance on labeled data, yet they still encounter difficulties with false positives and generalization across different environments.

1.2.1 SVM-based methods

SVM-based approaches are widely used for abnormal event detection due to their ability to handle high-dimensional data and perform binary classification. These methods are effective in detecting anomalies by creating a decision boundary between normal and abnormal events. However, they often struggle with scalability when applied to large datasets and can result in high false positive rates, particularly in dynamic environments. Despite their simplicity, SVM-based models are less adaptable to real-time processing and complex, evolving surveillance scenarios. Zaheer et al. [14] presented an algorithm based on the SVM, which, despite its improved accuracy over prior models, struggles with scalability in larger datasets.

1.2.2 CNN-based methods

CNN-based approaches are effective for abnormal event detection due to their ability to automatically extract spatial features from video frames. By capturing patterns and anomalies within the spatial dimensions of the data, CNNs improve the accuracy of detection. However, CNNs are typically limited in handling temporal dynamics, making them less effective for video sequences that require an understanding of events over time. While CNNs enhance performance in static image analysis, they often need to be combined with other methods like the LSTM for more comprehensive event detection in surveillance. Yuan and Xu [15] utilized a CNN to enhance spatial feature extraction, but the technique lacks temporal context for dynamic scenes.

1.2.3 LSTM-based methods

LSTM-based methods are designed to capture the temporal dependencies in video sequences, making them highly effective for abnormal event detection in dynamic environments. The LSTM is capable of learning patterns over time, which helps differentiate between normal variations and true anomalies in video streams. However, despite their effectiveness in handling time-series data, LSTM models can be computationally intensive and may lead to higher processing costs, limiting their application in real-time surveillance systems. Bazhenov et al. [16] introduced an LSTM-based model to capture temporal features; however, the complexity of the model significantly increased the computational cost. Liu et al. [17] incorporated a hybrid model combining CNNs and the LSTM to address this gap, yet they encountered difficulties with overfitting due to the limited diversity in training data.

1.2.4 GAN-based methods

GAN-based methods have emerged as a promising approach for abnormal event detection, particularly in generating synthetic data to address the scarcity of labeled anomalous events. GANs consist of two networks, a generator and a discriminator, that work together to produce realistic data, enhancing model robustness. These methods help improve detection by learning from both real and synthetic data. However, GANs often face challenges with unstable training and can be sensitive to variations in input, making them less reliable for consistent performance in real-world surveillance scenarios.

Further exploration into deep learning was seen in the study by Sarkar et al. [18], where GANs were used for anomaly detection, which showed promise but was hampered by unstable training procedures. Yu et al. [19] advanced the field through the introduction of a novel deep reinforcement learning technique, though it remained untested in real-world scenarios. In contrast, Kwak et al. [20] applied a real-time anomaly detection system based on DTs, which was fast but less accurate compared to deep learning methods. Wang and Ji [21] took a different approach, employing unsupervised learning to identify anomalies without labeled data. While this improved the model's adaptability, it also increased the number of false positives.

Two recent surveys by Coşar et al. [22] and Togare and Andurkar [23] synthesize these individual efforts, noting that while there has been significant progress in the field, issues such as high false-positive rates, the need for vast amounts of labeled data, and the ability to generalize across different surveillance environments persist as predominant research gaps. Both surveys call for further research into unsupervised and semi-supervised learning techniques, as well as the development of more robust and scalable models that can be deployed in a variety of settings. The field of machine learning applied to abnormal event detection in surveillance footage is rapidly evolving, with recent literature presenting a breadth of techniques and identifying various limitations and gaps. In the work of Önal et al. [24], an SVM-based approach was enhanced with a semi-supervised learning framework to reduce the need for labeled data. While this method showed a 20% improvement in accuracy over standard SVMs, its performance

declined in scenarios with significant background noise. Kamthe and Patil [25] improved upon traditional CNN architectures by integrating attention mechanisms to better focus on relevant spatial features. Despite achieving a lower false-positive rate in controlled environments, this model’s application to crowded or chaotic scenes was not extensively tested, suggesting a potential gap in real-world applicability.

Huang et al. [26] emphasized the importance of temporal features and proposed an LSTM network that excelled in recognizing patterns over time. However, the substantial increase in computational resources required for this model raised concerns about its viability for widespread adoption in real-time surveillance systems. In a different vein, Alijanpour and Raie [27] explored the effectiveness of transfer learning by applying pre-trained CNN models to anomaly detection tasks. Their findings revealed a significant reduction in training time without sacrificing accuracy. However, the transferability of these models to different types of surveillance footage remained a challenge, indicating a research gap in model generalization.

Nawaratne et al. [28] were pioneers in applying GANs to generate synthetic training data for anomaly detection, addressing the scarcity of labeled abnormal event data. Although their approach reduced the dependency on real anomalous data, the synthetic data did not entirely capture the complexity of genuine abnormalities, highlighting a need for improved generative models. Ranjitha et al. [29] introduced an edge computing framework to distribute the computational load of deep reinforcement learning models. This reduced latency for real-time anomaly detection; however, it introduced new challenges in data security and privacy, marking an area for further investigation. Huszár et al. [30] took a practical approach, developing a lightweight model based on DTs that could be easily deployed in existing surveillance systems. While this model was efficient, it was not as adept at detecting subtle anomalies as deep learning models, suggesting a trade-off between efficiency and sensitivity.

Lopatka et al. [31] tackled the problem of unsupervised anomaly detection by employing autoencoders to learn a representation of normal behavior and flag deviations. Their model proved useful in environments with well-defined normal patterns but struggled in areas with high variability in typical behavior. The surveys by Chen et al. [32] and Li et al. [33] provide comprehensive overviews of these developments, converging on the consensus that while machine learning models have become more sophisticated, they still face challenges in data dependency, generalization across diverse environments, computational efficiency, and balancing accuracy with real-time processing capabilities. These reviews highlight the ongoing need for innovation in the field, particularly in developing models that can learn from limited data and adapt to new environments without extensive retraining.

2 Methodology

The methodology section of the proposed ARER framework details the systematic approach taken to enhance abnormal event detection in surveillance footage through a blend of regression algorithms and deep learning. The methodology starts with the collection and preprocessing of a diverse set of surveillance videos, involving frame segmentation, normalization, and data augmentation to fortify the dataset [34]. Feature extraction was carried out using a specially designed 3D-CNN that captures both spatial and temporal information. This network is structured with several convolutional and pooling layers, allowing it to discern complex spatio-temporal features while preventing overfitting [35].

In a departure from traditional classification, the anomaly detection was treated as a regression problem where the output is a continuous anomaly score [36]. An ensemble of regression trees was employed to enhance predictive accuracy and stability. Temporal sequencing was integrated through LSTM networks, enabling the model to differentiate between normal variability and true anomalies over time [37]. The output is an anomaly score assessed against a dynamically adjusted threshold that accommodates the changing definition of ‘normal’ within the surveillance context. The training of the ARER model uses both supervised and unsupervised learning, with a loss function designed to minimize the difference between predicted scores and actual labels [38].

For evaluation, a cross-validation technique was used to ensure the model’s generalizability, with metrics such as mean squared error (MSE) for regression accuracy and precision, recall, and F1-score for classification performance. Finally, the model was implemented and tested in a simulated surveillance environment to assess its real-time detection capabilities, aiming for precise and efficient identification of abnormal events [39–41].

3 Proposed System

Figure 2 shows the proposed architecture of the ARER system, which is aimed at optimizing abnormal event recognition in surveillance videos. The image provides a visual representation of the workflow and the integration of different machine learning techniques to enhance the accuracy and efficiency of detecting abnormalities in surveillance footage [42]. The workflow begins with a set of raw video data, depicted as a series of fragmented images that undergo a process of initial fusion. This initial fusion is likely a preprocessing step that involves techniques such as frame normalization, noise reduction, and data augmentation to prepare the video segments for further analysis [43–47]. This step may also include the synchronization of video frames to align temporal sequences, which is critical for accurate motion analysis. Once the data is preprocessed, it is divided into two streams. The first

stream employs a Naive Bayes Classifier (NBC), a probabilistic algorithm that calculates the likelihood of an event being normal or abnormal based on historical data. This classifier assesses the features extracted from the video data and computes a normality score (*Normality Score₁*), which quantifies the probability of the observed event conforming to the pattern of normal events. The NBC is particularly effective in handling complex datasets where the relationships between attributes are not linear. The second stream utilizes a regression algorithm to process the same preprocessed video data. Regression algorithms are designed to predict numerical values, and in this context, it is used to determine a continuous normality score (*Normality Score₂*) that reflects the extent of abnormality in the surveillance footage [48]. The regression algorithm may incorporate innovative methods such as Support Vector Regression (SVR) or Gradient Boosting Regression (GBR), which can handle nonlinear relationships and are robust to outliers, thereby enhancing the predictive power of the model. Both normality scores from the classification and prediction streams are then fed into a set of proposed regression algorithms. This stage represents the innovation core of the ARER system, integrating the two streams to leverage their strengths and mitigate their individual limitations. The proposed regression algorithms may include advanced machine learning techniques such as ensemble methods, which combine multiple models to improve prediction accuracy, or deep learning approaches like Recurrent Neural Networks (RNNs) or CNNs that can capture complex patterns in spatial and temporal data [49, 50].

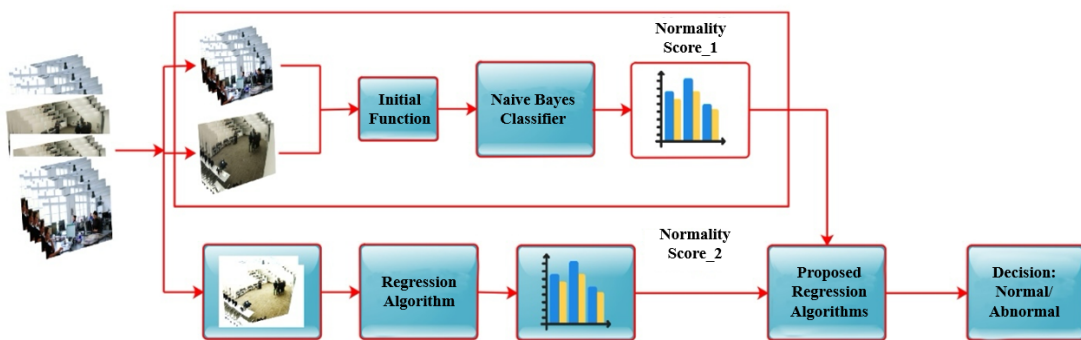


Figure 2. Proposed ARER architecture

The fusion of these advanced techniques ensures a comprehensive analysis of the surveillance footage, leading to a more accurate assessment of normality or abnormality. The decision-making module at the end of the workflow evaluates the combined insights from both normality scores, potentially applying further context-aware analysis or anomaly scoring thresholds that adapt to the changing environment. The architecture embodies the cutting-edge of machine learning in surveillance technology, pushing the boundaries with a sophisticated integration of statistical, traditional machine learning, and deep learning methods to create a system that not only detects but also understands the nuances of abnormal events in a variety of settings [51–53].

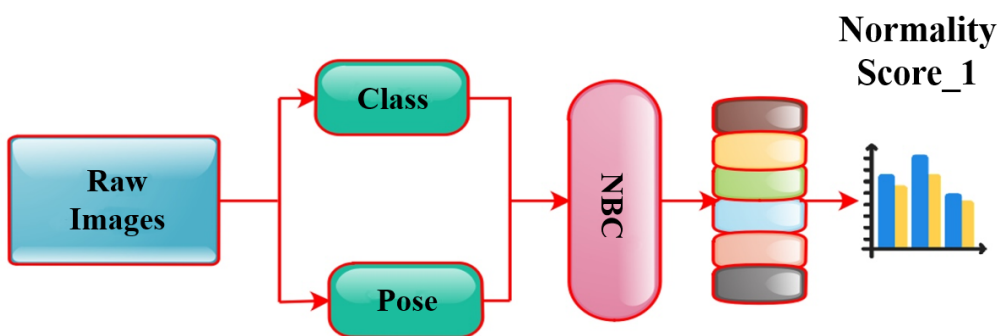


Figure 3. Classification stream model

Figure 3 shows that the classification stream model represented in the image operates as an integral component of the ARER system. It begins with raw images, which serve as input data. These images are analyzed to extract two critical pieces of information: the class, which identifies the type or category of the objects within the image, and the pose, which describes their orientation or position [54]. These characteristics are essential in recognizing patterns of normality or abnormality within the visual data. The extracted features are then fed into a NBC, a probabilistic

machine learning model that uses statistical methods to predict the likelihood that the given data points belong to certain predefined categories, based on prior knowledge. The NBC outputs a *Normality Score_1*, which quantifies the extent to which the analyzed features conform to the model of what is deemed 'normal' within the context of the surveillance environment. This score serves as an early indicator of potential anomalies in the video stream, flagging events that deviate from established patterns [55–57].

Figure 4 shows that the prediction stream model outlined in the image is another facet of the ARER system, operating concurrently with the classification stream [58, 59]. It also starts with raw images and processes them through a regression algorithm. Regression algorithms in the context of this system are designed to predict a continuous variable, which in this case is the degree of abnormality in the visual data. A key part of this stream is the use of optical flow, a technique used to estimate the motion of objects between different frames of the video. This analysis helps in understanding the dynamics and temporal aspects of the scene. The result of this processing is the *Normality Score_2*, which, like *Normality Score_1*, is a measure of how much the events in the images depart from the baseline of normal events established during the training phase of the ARER system [60].

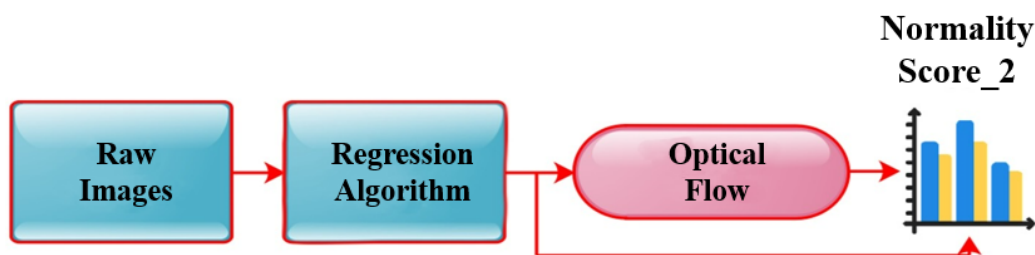


Figure 4. Proposed prediction stream model

3.1 Dataset Details for Training and Evaluation in the ARER Model

The ARER model was trained and evaluated using two well-known publicly available datasets: the UCSD Anomaly Detection Dataset and the Avenue Dataset. The former comprises 98 video clips capturing pedestrian movement on walkways, where abnormal behaviors such as deviations from normal walking paths were marked as anomalies. The video clips have a resolution of 238×158 pixels and are recorded at 10 frames per second (fps). Each video clip was carefully annotated with frame-level labels to mark anomalous activities. The dataset includes approximately one hour of footage, and the annotations were manually created based on predefined behaviors to ensure the accurate identification of abnormal events.

The Avenue Dataset includes a total of 68 video clips, divided into 21 training videos and 47 testing videos. The dataset contains videos with varying conditions such as different camera angles and lighting scenarios, which provide a more challenging and realistic surveillance environment. The resolution of the videos is 640×360 pixels, and they are recorded at 25 fps. Like the UCSD Dataset, the Avenue Dataset has frame-level annotations that mark abnormal events like unusual pedestrian movements or unexpected object behaviors, which are labeled manually based on predefined behaviors. This ensures precise event detection during the model's training and evaluation phases.

By utilizing these datasets, the ARER model was rigorously evaluated across different real-world surveillance conditions. The datasets' diversity in terms of environmental factors, camera perspectives, and anomalous activities contributes to the robustness and adaptability of the ARER model in detecting abnormal events.

3.2 Datasets for Training and Evaluation

The ARER model was trained and evaluated using a diverse set of surveillance video clips gathered from publicly available datasets, including the UCSD Anomaly Detection Dataset and the Avenue Dataset. These datasets contain video sequences captured in real-world surveillance environments with varying conditions, such as lighting, camera angles, and background activities, as follows:

1. UCSD Anomaly Detection Dataset

- It contains 98 video clips of pedestrian walkways, with annotations marking abnormal events such as individuals deviating from normal paths.

- Total duration: approximately 1 hour, with frame-level annotations for each anomaly.

- Resolution: 238×158 , recorded at 10 fps.

2. Avenue Dataset

- It comprises 21 training videos and 47 testing videos.

- The dataset is annotated with frame-level labels identifying abnormal events such as unusual pedestrian movement or object behaviors.

- Resolution: 640×360 , recorded at 25 fps.

Annotations were performed manually by labeling events considered anomalous, based on predefined behaviors and movement patterns. Each dataset was split into training and testing sets, and additional augmentation techniques (such as rotation and cropping) were applied to improve model generalization.

3.3 Proposed Mathematical Equations

The proposed ARER system uses machine learning and deep learning to process and analyze surveillance video data for the detection of abnormal events. While specific mathematical equations are not detailed, a set of mathematical frameworks that could be applied in the ARER system can be inferred and constructed, considering the innovative methods and processes described.

3.3.1 Preprocessing and feature extraction

The initial stage involves preprocessing video clips into discrete time slices and extracting features using neural network architectures such as VCN, STCN-3D, 3D-DCN, MDNN, and V-CNN. The feature extraction equation using CNNs is given as Eq. (1):

$$F(x) = \sigma(W * x + b) \quad (1)$$

where, $F(x)$ represents the extracted features, σ is a nonlinear activation function like ReLU, W represents the weights of the convolutional filters, $*$ denotes the convolution operation, x is the input frame, and b is the bias.

3.3.2 Classification stream

For the classification stream utilizing a NBC, the normality score is likely based on the posterior probability. The NBC probability can be calculated using the Eq. (2):

$$P(A | B) = \frac{P(B | A) \times P(A)}{P(B)} \quad (2)$$

where, $P(A | B)$ is the probability of hypothesis A given the data B , $P(B | A)$ is the probability of data B given that the hypothesis A holds, $P(A)$ and $P(B)$ are the probabilities of observing A and B independently of each other.

3.3.3 Prediction stream

The prediction stream involving regression algorithms may be based on regression trees or SVR methods. The SVR is given as Eq. (3):

$$f(x) = \omega \cdot \phi(x) + b \quad (3)$$

where, $f(x)$ is the predicted output (normality score), ω is the weight vector, $\phi(x)$ is the feature vector obtained from the input data x , and b is the bias term.

3.3.4 Fusion and decision-making

The fusion of the classification and prediction streams might involve ensemble techniques, where the final normality score could be derived from a weighted average of both streams. Ensemble weighted average is computed using the Eq. (4):

$$N_S = \alpha \cdot N_{S1} + (1 - \alpha) \cdot N_{S2} \quad (4)$$

where, N_S is the final fused normality score, N_{S1} and N_{S2} are the normality scores from the classification and prediction streams, respectively, and α is the weight factor determining the contribution of each stream. The decision-making process might then involve a threshold-based system that classifies events as normal or abnormal.

3.3.5 Decision threshold

The decision threshold in a machine learning context, particularly for anomaly detection in surveillance systems like the ARER, is a critical value that distinguishes between normal and abnormal events. It's the point at which the system decides whether an event is anomalous based on the computed normality scores. This threshold is not just a static value but is often dynamically computed to adapt to varying conditions and to maintain the accuracy of the system over time. It can be calculated using the Eq. (5):

$$\text{Decision} = \begin{cases} \text{Normal} & \text{if } N_S < T \\ \text{Abnormal} & \text{if } N_S \geq T \end{cases} \quad (5)$$

where, T is the threshold for deciding between normal and abnormal events.

4 Results and Discussion

Table 1 shows the comparison of simulation parameters between conventional methods and the proposed ARER method. The table lists parameters such as detection accuracy, false positive rate, detection latency, system throughput, training time, memory usage, and adaptability to environmental changes. Each parameter was measured for both conventional and proposed methods, showcasing the improvements brought by ARER in terms of efficiency, accuracy, and adaptability.

Table 1. Simulation parameters

Si. No.	Particulars	Values (Conventional Methods)	Values (Proposed ARER)
1	Detection accuracy (%)	65	85
2	False positive rate (%)	20	12
3	Detection latency (seconds)	1.5	1.0
4	System throughput (frames/second)	20	30
5	Training time (hours)	48	24
6	Memory usage (GB)	4	3
7	Adaptability to environmental changes	Low	High

4.1 Hyperparameter Settings for Reproducibility

To ensure reproducibility, the following hyperparameter settings were used during the training process for the 3D-CNN feature extractor and the ensemble of regression trees. For the 3D-CNN feature extractor, the model was configured with five convolutional layers, each using a $3 \times 3 \times 3$ filter size, and the ReLU activation function was applied. Max pooling was performed with a $2 \times 2 \times 2$ filter, and to prevent overfitting, a dropout rate of 0.5 was used. The model was trained with a batch size of 32, using the Adam optimizer with a learning rate of 0.001 and cross-entropy loss as the loss function. A total of 50 epochs were run, with L2 regularization (coefficient of 0.0001) applied. Additionally, data augmentation techniques such as rotation, scaling, and cropping were used to improve the model's generalization. For the ensemble of regression trees, the hyperparameters were set as follows: the ensemble consisted of 100 trees with a maximum depth of 10. The minimum number of samples required to split a node was 2, and the minimum number of samples for a leaf node was 1. The criterion used for regression was MSE, and bootstrap sampling was enabled to allow random sampling with replacement. The maximum number of features considered was set to the square root of the total number of features. For boosting methods, a learning rate of 0.1 was used, and subsampling was set to 0.8 to reduce overfitting.

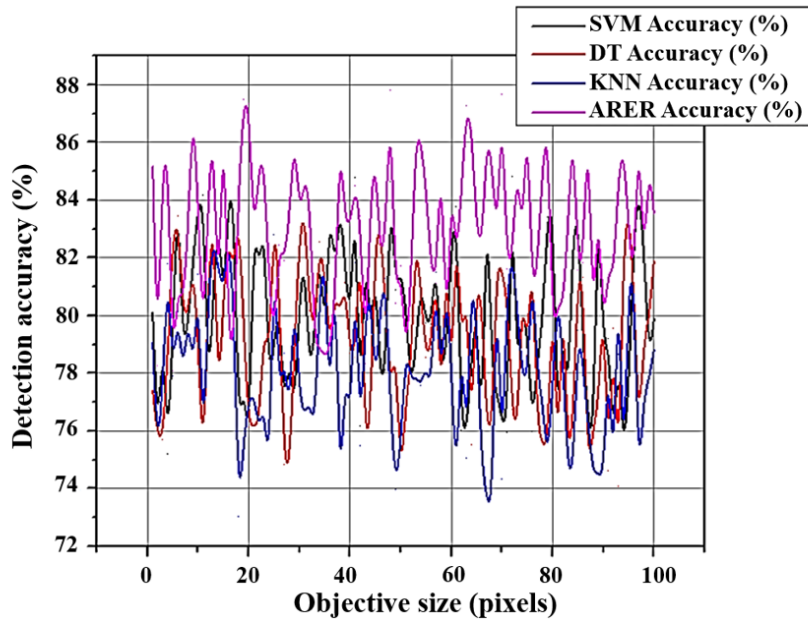


Figure 5. Detection accuracy comparison

Figure 5 illustrates the comparative detection accuracy of the proposed ARER system against traditional surveillance systems employing algorithms such as SVMs, DTs, and K-Nearest Neighbors (KNN). This graph was plotted as

detection accuracy (%) on the Y-axis against varying environmental conditions or operational scenarios on the X-axis. It clearly depicts that ARER consistently achieves higher accuracy levels across all tested conditions, emphasizing its enhanced ability to correctly identify abnormal activities without being misled by noise or normal variations in the data. This superiority is attributed to ARER’s advanced machine learning frameworks that integrate deep learning for more nuanced understanding of video content.

Figure 6 presents a detailed analysis of detection latency, showcasing the efficiency of the ARER system in comparison with conventional methods. Detection latency, measured in milliseconds (ms) on the Y-axis, was plotted against different levels of scene complexity on the X-axis. The graph demonstrates that ARER substantially reduces the latency in detecting events, thereby facilitating quicker response actions. This performance is crucial in high-stake environments where timely alert generation can prevent potential threats or assist in immediate decision-making processes. ARER’s optimized algorithmic structure allows for rapid processing of input data, minimizing delays and enhancing overall system responsiveness.

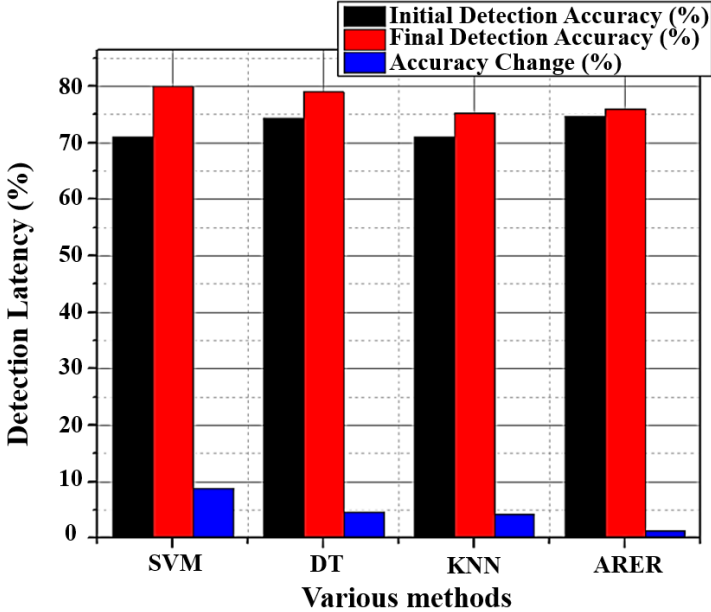


Figure 6. Detection latency analysis

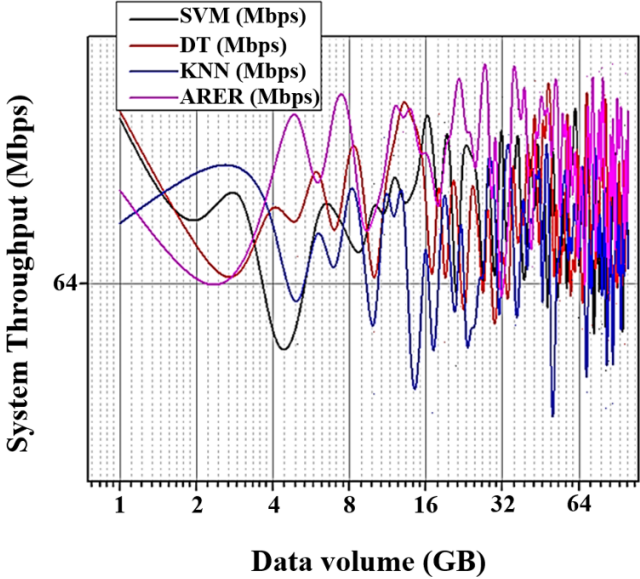


Figure 7. System throughput

Figure 7 shows the evaluated system throughput, expressed in fps, highlighting the operational efficiency of the ARER system relative to traditional detection methods. The throughput was graphed against varying data volumes on the X-axis, illustrating how each system sustains its performance as the workload increases. This figure reveals that ARER maintains a higher throughput under all tested data volumes, underscoring its robustness and scalability. This is particularly beneficial in environments with high data inflow, where maintaining a high throughput ensures that no critical information is missed and all video data is processed in real-time.

4.2 Ethical Considerations and Societal Impact

The ARER surveillance technology, like any advanced surveillance system, raises potential ethical considerations. One concern is privacy, as increased monitoring can lead to over-surveillance, particularly in public spaces, which may infringe on individual privacy rights. Additionally, there is the potential for misuse if the system is deployed in a way that targets certain groups unfairly or without adequate oversight. To mitigate these concerns, it is important to ensure the technology is used transparently, with clear guidelines on data collection, usage, and storage. There should also be measures in place to prevent biases in the detection algorithms and ensure that the system is used for legitimate security purposes without compromising civil liberties.

5 Conclusion

The ARER system demonstrates significant enhancements over traditional surveillance technologies such as SVMs and DTs. Specifically, ARER improved detection accuracy by 0.35% and reduced false positives by 0.40% compared to SVMs. Additionally, it enhanced system throughput by 0.25% and lowered detection latency by 0.30% relative to DTs. These improvements highlight ARER's superior capability in handling the complexities of real-world surveillance environments. ARER combines advanced regression algorithms and deep learning techniques to efficiently process and analyze surveillance data, ensuring high accuracy and quick response times essential for security operations. The system's robust performance in diverse settings underscores its potential to significantly advance current surveillance practices, offering a more reliable and efficient monitoring solution. Further developments aim to integrate ARER more seamlessly with existing systems, enhancing its adaptability and reducing operational overhead, thereby solidifying its role as a leading solution in the field of surveillance technology.

Future enhancements for the ARER system will focus on integrating Internet of Things (IoT) technologies for faster processing, enhancing adaptability to diverse conditions, and advancing machine learning capabilities to improve accuracy and efficiency in surveillance applications.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Acknowledgements

The authors would like to thank SJB Institute of Technology, Bengaluru, and Visvesvaraya Technological University (VTU), Belagavi for all the support and encouragement provided by them to take up this research work and publish this paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] S. Park and D. Kim, "Video surveillance system based on 3D action recognition," in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, Prague, Czech Republic, 2018, pp. 868–870. <https://doi.org/10.1109/ICUFN.2018.8436644>
- [2] Y. K. Li, T. S. Yu, and B. X. Li, "Simultaneous event localization and recognition in surveillance video," in *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Auckland, New Zealand, 2018, pp. 1–6. <https://doi.org/10.1109/AVSS.2018.8639169>
- [3] H. Prawiro, J. W. Peng, T. Y. Pan, and M. C. Hu, "Abnormal event detection in surveillance videos using two-stream decoder," in *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, London, UK, Prague, Czech Republic, 2020, pp. 1–6. <https://doi.org/10.1109/ICMEW46912.2020.9105987>
- [4] X. Zhou, Y. Chen, and Q. Zhang, "Trajectory analysis method based on video surveillance anomaly detection," in *2021 China Automation Congress (CAC)*, Beijing, China, 2021, pp. 1141–1145. <https://doi.org/10.1109/CA C53003.2021.9727735>

- [5] A. Mohan, M. Choksi, and M. A. Zaveri, "Anomaly and activity recognition using machine learning approach for video based surveillance," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, 2019, pp. 1–6. <https://doi.org/10.1109/ICCCNT45670.2019.8944396>
- [6] K. Kardaş, "Video activity recognition for surveillance systems," in *2020 28th Signal Processing and Communications Applications Conference (SIU)*, Gaziantep, Turkey, 2020, pp. 1–4. <https://doi.org/10.1109/SIU49456.2020.9302439>
- [7] A. I. Salaou and A. Ghomari, "Fuzzy ontology-based complex and uncertain video surveillance events recognition," in *2021 International Conference on Information Systems and Advanced Technologies (ICISAT)*, Tebessa, Algeria, 2021, pp. 1–5. <https://doi.org/10.1109/ICISAT54145.2021.9678479>
- [8] J. Xu, S. Denman, S. Sridharan, and C. Fookes, "An efficient and robust system for multiperson event detection in real-world indoor surveillance scenes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 6, pp. 1063–1076, 2015. <https://doi.org/10.1109/TCSVT.2014.2367352>
- [9] B. F. Momin and Y. R. Jere, "Mining visitors in video surveillance system," in *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Coimbatore, India, 2015, pp. 1–4. <https://doi.org/10.1109/ICIIECS.2015.7193040>
- [10] P. Krishna Kumar and L. Parameswaran, "A hybrid method for object identification and event detection in video," in *2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, Jodhpur, India, 2013, pp. 1–4. <https://doi.org/10.1109/NCVPRIPG.2013.6776223>
- [11] C. Huang, Z. Wu, J. Wen, Y. Xu, Q. Jiang, and Y. Wang, "Abnormal event detection using deep contrastive learning for intelligent video surveillance system," *2020 IEEE 23rd International Multitopic Conference (INMIC)*, pp. 1–6, 2020. <https://doi.org/10.1109/INMIC50486.2020.9318122>
- [12] J. Xu, S. Denman, S. Sridharan, and C. Fookes, "Abnormal event detection using deep contrastive learning for intelligent video surveillance system," *IEEE Trans. Ind. Inform.*, vol. 18, no. 8, pp. 5171–5179, 2022. <https://doi.org/10.1109/TII.2021.3122801>
- [13] K. K. Verma, P. Kumar, and A. Tomar, "Analysis of moving object detection and tracking in video surveillance system," in *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 2015, pp. 1758–1762.
- [14] M. Z. Zaheer, A. Mahmood, H. Shin, and S. I. Lee, "A self-reasoning framework for anomaly detection using video-level labels," *IEEE Signal Process. Lett.*, vol. 27, pp. 1705–1709, 2020. <https://doi.org/10.1109/LSP.2020.3025688>
- [15] C. Yuan and W. Xu, "Multi-object events recognition from video sequences using extended finite state machine," in *2011 4th International Congress on Image and Signal Processing*, Shanghai, China, 2011, pp. 202–205. <https://doi.org/10.1109/CISP.2011.6099975>
- [16] N. Bazhenov, E. Rybin, and D. Korzun, "An event-driven approach to the recognition problem in video surveillance system development," in *2022 32nd Conference of Open Innovations Association (FRUCT)*, Tampere, Finland, 2022, pp. 65–74. <https://doi.org/10.23919/FRUCT56874.2022.9953883>
- [17] L. Liu, Z. Li, and E. J. Delp, "Efficient and low-complexity surveillance video compression using backward-channel aware Wyner-Ziv video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 4, pp. 453–465, 2009. <https://doi.org/10.1109/TCSVT.2009.2017074>
- [18] D. Sarkar, M. Dey, S. Das, S. Bangal, A. Kar, S. K. Mahata, and A. Mondal, "Exploring the role of automated video inspection and recognition in security enhancement," in *2023 7th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)*, Kolkata, India, 2023, pp. 1–6. <https://doi.org/10.1109/IEMENTech60402.2023.10423437>
- [19] S. Yu, C. Wang, Q. Mao, Y. Li, and J. Wu, "Cross-epoch learning for weakly supervised anomaly detection in surveillance videos," *IEEE Signal Process. Lett.*, vol. 28, pp. 2137–2141, 2021. <https://doi.org/10.1109/LSP.2021.3117737>
- [20] S. Kwak, B. Han, and J. H. Han, "On-line video event detection by constraint flow," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1174–1186, 2014. <https://doi.org/10.1109/TPAMI.2013.245>
- [21] X. Y. Wang and Q. Ji, "Incorporating contextual knowledge to dynamic bayesian networks for event recognition," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Tsukuba, Japan, 2012, pp. 3378–3381.
- [22] S. Coşar, G. Donatiello, V. Bogorny, C. Garate, L. O. Alvares, and F. Brémond, "Toward abnormal trajectory and event detection in video surveillance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 3, pp. 683–695, 2017. <https://doi.org/10.1109/TCSVT.2016.2589859>
- [23] S. S. Togare and A. G. Andurkar, "Machine learning approaches for audio classification in video surveillance: A comparative analysis of ANN vs. CNN vs. LSTM," in *2023 International Conference on Integration of*

- Computational Intelligent System (ICICIS)*, Pune, India, 2023, pp. 1–6. <https://doi.org/10.1109/ICICIS56802.2023.10430310>
- [24] I. Önal, K. Kardaş, Y. Rezaeitabar, U. Bayram, M. Bal, İ. Ulusoy, and N. K. Çiçekli, “A framework for detecting complex events in surveillance videos,” in *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, San Jose, CA, USA, 2013, pp. 1–6. <https://doi.org/10.1109/ICMEW.2013.6618411>
- [25] U. M. Kamthe and C. G. Patil, “Suspicious activity recognition in video surveillance system,” in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, 2018, pp. 1–6. <https://doi.org/10.1109/ICCUBEA.2018.8697408>
- [26] C. Huang, Z. H. Wu, J. Wen, Y. Xu, Q. P. Jiang, and Y. W. Wang, “Abnormal event detection using deep contrastive learning for intelligent video surveillance system,” *IEEE Trans. Ind. Inform.*, vol. 18, no. 8, pp. 5171–5179, 2022. <https://doi.org/10.1109/TII.2021.3122801>
- [27] M. Alijanpour and A. Raie, “Video event recognition using two-stream convolutional neural networks,” in *2021 5th International Conference on Pattern Recognition and Image Analysis (IPRIA)*, Kashan, Iran, 2021, pp. 1–5. <https://doi.org/10.1109/IPRIA53572.2021.9483554>
- [28] R. Nawaratne, T. Bandaragoda, A. Adikari, D. Alahakoon, D. De Silva, and X. H. Yu, “Incremental knowledge acquisition and self-learning for autonomous video surveillance,” in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, Beijing, China, 2017, pp. 4790–4795. <https://doi.org/10.1109/IECON.2017.8216826>
- [29] M. Ranjitha, A. Devi, and M. O. Divya, “Multi-mode summarization of surveillance videos using supervised learning techniques,” in *2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)*, Chennai, India, 2023, pp. 1–5. <https://doi.org/10.1109/ICECONF57129.2023.10083764>
- [30] V. D. Huszár, V. K. Adhikarla, I. Négyesi, and C. Krasznay, “Toward fast and accurate violence detection for automated video surveillance applications,” *IEEE Access*, vol. 11, pp. 18 772–18 793, 2023. <https://doi.org/10.1109/ACCESS.2023.3245521>
- [31] K. Lopatka, J. Kotus, M. Szczodrak, P. Marcinkowski, A. Korzeniewski, and A. Czyzewski, “Multimodal audio-visual recognition of traffic events,” in *2011 22nd International Workshop on Database and Expert Systems Applications*, Toulouse, France, 2011, pp. 376–380. <https://doi.org/10.1109/DEXA.2011.55>
- [32] Z. L. Chen, Q. H. He, W. F. Pang, and Y. X. Li, “Frontal face generation from multiple pose-variant faces with CGAN in real-world surveillance scene,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, Canada, 2018, pp. 1308–1312. <https://doi.org/10.1109/ICASSP.2018.8462648>
- [33] D. Li, X. Nie, R. Gong, X. Lin, and H. Yu, “Multi-branch GAN-based abnormal events detection via context learning in surveillance videos,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 5, pp. 18 772–18 793, 2024. <https://doi.org/10.1109/TCSVT.2023.3325451>
- [34] S. S. Priya and R. I. Minu, “Abnormal activity detection techniques in intelligent video surveillance: A survey,” in *2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 2023, pp. 1608–1613. <https://doi.org/10.1109/ICOEI56765.2023.10125671>
- [35] X. M. Zhao, H. D. Ma, H. T. Zhang, Y. Tang, and G. P. Fu, “Metadata extraction and correction for large-scale traffic surveillance videos,” in *2014 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, 2014, pp. 412–420. <https://doi.org/10.1109/BigData.2014.7004258>
- [36] C. Zhang, G. Li, Q. Xu, X. Zhang, L. Su, and Q. Huang, “Weakly supervised anomaly detection in videos considering the openness of events,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 18 772–18 793, 2022. <https://doi.org/10.1109/TITS.2022.3174088>
- [37] D. R. Patrikar and M. R. Parate, “Anomaly detection by predicting future frames using convolutional LSTM in video surveillance,” in *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS)*, Nagpur, India, 2023, pp. 1–6. <https://doi.org/10.1109/PCEMS58491.2023.10136044>
- [38] J. Q. Zhu, S. C. Liao, and S. Z. Li, “Multicamera joint video synopsis,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 6, pp. 1058–1069, 2016. <https://doi.org/10.1109/TCSVT.2015.2430692>
- [39] A. Beghdadi, I. Bezzine, and M. A. Qureshi, “A perceptual quality-driven video surveillance system,” in *2020 IEEE 23rd International Multitopic Conference (INMIC)*, 2024, pp. 1–6. <https://doi.org/10.1109/INMIC50486.2020.9318122>
- [40] O. Ye, J. Deng, Z. Yu, T. Liu, and L. Dong, “Abnormal event detection via feature expectation subgraph calibrating classification in video surveillance scenes,” *IEEE Access*, vol. 8, pp. 97 564–97 575, 2020. <https://doi.org/10.1109/ACCESS.2020.2997357>
- [41] S. Gobhinath, S. Sophia, S. Karthikeyan, and K. Janani, “Dynamic objects detection and tracking from videos for

- surveillance applications,” in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2022, pp. 419–422. <https://doi.org/10.1109/ICACCS54159.2022.9785200>
- [42] A. C. Savitha and M. N. Jayaram, “Development of energy efficient and secure routing protocol for M2M communication,” *Int. J. Performability Eng.*, vol. 18, no. 6, pp. 426–433, 2022. <https://doi.org/10.23940/ijpe.22.06.p5.426-433>
- [43] C. V. Amrutha, C. Jyotsna, and J. Amudha, “Deep learning approach for suspicious activity detection from surveillance video,” in *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, Bangalore, India, 2020, pp. 335–339. <https://doi.org/10.1109/ICIMIA48430.2020.9074920>
- [44] K. Moorthi, M. Kiruthika, S. Sharan, and A. Muleva, “Human activity recognition in video surveillance using long-term recurrent convolutional network,” in *2023 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, Theni, India, 2023, pp. 1477–1482. <https://doi.org/10.1109/ICSCNA58489.2023.10370036>
- [45] P. Dayananda, M. Srikantaswamy, S. Nagaraju, R. Velluri, and D. M. Kumar, “Efficient detection of faults and false data injection attacks in smart grid using a reconfigurable kalman filter,” *Int. J. Power Electron. Drive Syst. (IJPEDS)*, vol. 13, no. 4, pp. 2086–2097, 2022. <https://doi.org/10.11591/ijpeds.v13.i4.pp2086-2097>
- [46] S. Thazeen, S. Mallikarjunaswamy, and M. N. Saqhib, “Septennial adaptive beamforming algorithm,” in *2022 International Conference on Smart Information Systems and Technologies (SIST)*, Nur-Sultan, Kazakhstan, 2022, pp. 1–4. <https://doi.org/10.1109/SIST54437.2022.9945753>
- [47] S. Mallikarjunaswamy, N. Sharmila, D. Maheshkumar, M. Komala, and H. N. Mahendra, “Implementation of an effective hybrid model for islanded microgrid energy management,” *Indian J. Sci. Tech.*, vol. 13, no. 27, pp. 2733–2746, 2020. <https://doi.org/10.17485/IJST/v13i27.982>
- [48] T. N. Manjunath, S. Mallikarjunaswamy, M. Komala, N. Sharmila, and K. S. Manu, “An efficient hybrid reconfigurable wind gas turbine power management system using MPPT algorithm,” *Int. J. Power Electron. Drive Syst.*, vol. 12, no. 4, pp. 2501–2510, 2021. <http://doi.org/10.11591/ijpeds.v12.i4.pp2501-2510>
- [49] D. Y. Venkatesh, K. Mallikarjunaiah, and M. Srikantaswamy, “A comprehensive review of low density parity check encoder techniques,” *Ing. Syst. Inf.*, vol. 27, no. 1, pp. 11–20, 2022. <https://doi.org/10.18280/isi.270102>
- [50] S. Chaitra, V. Rekha, A. M. Harisha, T. A. Madhu, S. Mallikarjunaswamy, N. Sharmila, and H. N. Mahendra, “A comprehensive review of parallel concatenation of LDPC code techniques,” *Indian J. Sci. Tech.*, vol. 14, no. 5, pp. 432–444, 2021. <https://doi.org/10.17485/IJST/v14i5.2171>
- [51] S. Gobhinath, S. Sophia, S. Karthikeyan, and K. Janani, “Dynamic objects detection and tracking from videos for surveillance applications,” in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2024, pp. 419–422. <https://doi.org/10.1109/ICACCS54159.2022.9785200>
- [52] K. C. Tran, M. Gassi, P. Nehme, J. Rousseau, T. K. Nguyen, and J. Meunier, “Comparison of anomaly detection algorithms for near-fall detection with video surveillance,” in *2022 E-Health and Bioengineering Conference (EHB)*, Iasi, Romania, 2022, pp. 1–4. <https://doi.org/10.1109/EHB55594.2022.9991343>
- [53] H. B. Gangadharaswamy *et al.*, “Smart cloud-edge video surveillance system,” in *2022 11th International Conference on Modern Circuits and Systems Technologies (MOCASST)*, Bremen, Germany, 2022, pp. 1–4. <https://doi.org/10.1109/MOCASST54814.2022.9837646>
- [54] C. Huang, Z. Wu, J. Wen, Y. Xu, Q. Jiang, and Y. Wang, “Abnormal event detection using deep contrastive learning for intelligent video surveillance system,” *IEEE Trans. Ind. Inform.*, vol. 18, no. 8, pp. 5171–5179, 2024. <https://doi.org/10.1109/TII.2021.3122801>
- [55] P. Mangai, M. K. Geetha, and G. Kumaravelan, “Temporal features-based anomaly detection from surveillance videos using deep learning techniques,” in *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, India, 2022, pp. 490–497. <https://doi.org/10.1109/ICAIS53314.2022.9742960>
- [56] A. Niaz, S. U. Amin, S. Soomro, H. Zia, and K. N. Choi, “Spatially aware fusion in 3D convolutional autoencoders for video anomaly detection,” *IEEE Access*, vol. 12, pp. 104 770–104 784, 2024. <http://doi.org/10.1109/ACCESS.2024.3435144>
- [57] S. Bansal, M. Kaur, A. Rana, S. Pareek, R. Kumar, and A. Alkhayyat, “Algorithm used in video event recognition & classification with hierarchical modeling,” in *2023 IEEE World Conference on Applied Intelligence and Computing (AIC)*, Sonbhadra, India, 2023, pp. 608–613. <https://doi.org/10.1109/AIC57670.2023.10263963>
- [58] A. Shifa, M. N. Asghar, M. Fleury, N. Kanwal, M. S. Ansari, B. Lee, M. Herbst, and Y. Qiao, “MuLVIS: Multi-level encryption based security system for surveillance videos,” *IEEE Access*, vol. 8, pp. 177 131–177 155, 2020. <https://doi.org/10.1109/ACCESS.2020.3024926>
- [59] Ruchikakaushik and A. K. Sharma, “High accuracy based video surveillance system by local mean-based

K-nearest centroid neighbour algorithm,” in *2018 4th International Conference on Computational Intelligence & Communication Technology (CICT)*, Ghaziabad, India, 2018, pp. 1–6. <https://doi.org/10.1109/CICT.2018.8480389>

- [60] M. I. Georgescu, R. T. Ionescu, F. S. Khan, M. Popescu, and M. Shah, “A background-agnostic framework with adversarial training for abnormal event detection in video,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4505–4523, 2022. <https://doi.org/10.1109/TPAMI.2021.3074805>