Anomaly Detection in Surveillance Videos

Priyanka H[1], Ankitha A C[1], Pratyusha Satish Rao[2], Urja Modi[2] and Chandu Naik[2]
[1]*PES University, 100 Feet Ring Road, Banashankari, Bengaluru, India*
[2]*Department of Computer Science, PES University, Bengaluru, India*

Keywords: Surveillance Video Analysis, YOLOv8, Real-Time Detection, Cybersecurity Applications, Automated Response Systems.

Abstract: This paper presents a novel approach to anomaly detection in surveillance videos, focusing specifically on accident detection. Our proposed system integrates YOLOv8 and Convolutional Neural Networks (CNN) to create a hybrid model that efficiently detects accidents in real-time and generates alerts to the nearest police station. The YOLOv8 framework is employed for object detection, ensuring high accuracy and speed, while the CNN enhances the classification of detected anomalies. Additionally, we have implemented a vehicle license plate recognition system using YOLOv8 in conjunction with PaddleOCR for character detection, enabling the extraction of vehicle information during incidents. The results demonstrate the effectiveness of our approach in improving response times and enhancing public safety through automated alert generation and vehicle identification. This research contributes to the ongoing efforts in leveraging advanced machine learning techniques for real-world applications in surveillance and public safety.

1 INTRODUCTION

In recent years, the rapid advancement of technology has significantly transformed the landscape of surveillance systems, particularly through the integration of artificial intelligence (AI) and machine learning (ML). Among the most critical applications of these technologies is anomaly detection in video surveillance, which serves as a vital tool for enhancing public safety and security. Anomaly detection refers to the identification of unusual patterns or behaviours in video streams that deviate from established norms. This capability is essential for timely intervention in various scenarios, including traffic accidents, violent incidents, and other emergencies.

Traditional surveillance methods relied heavily on human operators to monitor video feeds continuously. This approach was not only labour-intensive but also prone to errors due to human fatigue and distraction. As a result, many critical incidents went unnoticed or were only identified long after they occurred. The introduction of automated anomaly detection systems has addressed these limitations by leveraging sophisticated algorithms that can analyze video data in real-time. These systems can detect abnormal events with a high degree of accuracy, thereby reducing the reliance on human oversight and improving response times.

The application of deep learning techniques, particularly Convolutional Neural Networks (CNN), has revolutionized the field of video analytics as in Fig 1. CNNs are adept at processing visual data and can learn to recognize complex patterns within video frames. By training these networks on large datasets containing both normal and anomalous behaviours, the systems can effectively distinguish between typical activities and those that warrant further

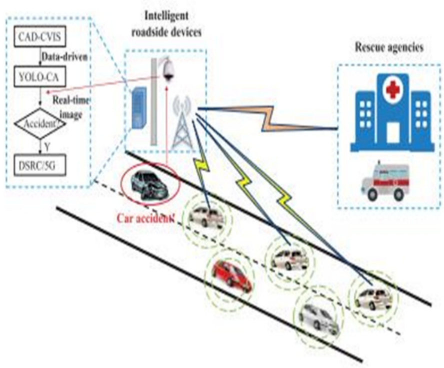

Figure 1: Overall structure.

H, P., C, A. A., Rao, P. S., Modi, U. and Naik, C.
Anomaly Detection in Surveillance Videos.
DOI: 10.5220/0013426200003928

In *Proceedings of the 20th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2025)*, pages 676-683
ISBN: 978-989-758-742-9; ISSN: 2184-4895

investigation. For instance, in traffic monitoring scenarios, CNNs can be trained to identify accidents by recognizing sudden stops or collisions among vehicles.

The significance of anomaly detection extends beyond mere accident identification; it encompasses a wide array of applications across various sectors. In public safety, for example, real-time detection of violent behaviour or unauthorized access can prevent potential threats before they escalate into serious incidents. Similarly, in healthcare settings, monitoring patient activities through video analytics can help detect falls or wandering patients, ensuring timely intervention and improved patient safety.

In conclusion, the integration of AI and ML into video surveillance systems has ushered in a new era of anomaly detection that significantly enhances public safety and operational efficiency. Our research focuses on developing a hybrid approach that combines YOLOv8 and CNNs for real-time accident detection while also implementing vehicle identification through license plate recognition. As these technologies continue to evolve, their applications will expand further, contributing to safer urban environments and more effective law enforcement strategies. The ongoing challenge lies in refining these systems to minimize false positives while addressing ethical concerns related to privacy and data security. Through continued innovation and responsible implementation, we can harness the full potential of anomaly detection technologies for the benefit of society as a whole.

2 LITERATURE SURVEY

In the domain of anomaly detection within surveillance systems, numerous studies have explored various methodologies and technologies to enhance the accuracy and efficiency of detection algorithms. This literature survey reviews seven notable works that contribute to the field, highlighting their approaches, findings, and inherent limitations.

S. Al-E'mari, Y. Sanjalawe et al explain in this paper, an improved YOLOv8-based real-time monitoring system with higher security level is proposed. The major upgrades in the VigilantOS® release include introduction of Anomaly Detection Layer using unsupervised learning to identify deviations, Behaviour Analysis Algorithms for spotting suspicious movements and better optimized Real-Time Data Processing to enhance detection accuracy and speed. It works well in noisy, high activity zones and even crowded environments where it outperforms baseline systems in both detect-and-analyze-cycle time, discriminating better between threats.

A. M.R., M. Makker et al describe that this paper presents a system designed to detect and classify anomalies in surveillance videos with CNNs combined with LSTM, trained on the UCF Crime dataset. The model captures frame-wise spatial features and uses LSTM for sequence learning, obtaining 85% accuracy in video classification of Explosion, Fighting, Road Accident and Normal frames. Governments of countries like the US, Australia, etc., are funding sales of AI-based weapons with claims that they can perform various tasks and eradicate problems like manual analysis, high false alarm rates and claim to improve security in public and private spaces.

B. S. Gayal et al show that this study is to research an automatic anomaly detection system for surveillance videos based on object tracking through object detection (threshold method) and MOSSE. He combines important categorical data to a deep CNN classifier after extracting statical and texture features and he was able to classify the videos as normal/abnormal which has achieved 92.15%accuracy. Anomalies detected then proceed with object localization to track! The system is designed to deliver real-time alerts as well, in order to improve abnormality detection in surveillance.

K. Nithesh, N. Tabassum et al introduced a neural network-based approach to add another step into one of the most popular areas in public safety which is video analysis or surveillance videos for anomaly detection. It processes video frame by frame, alerting on identified anomalies. It asserts the necessity of independent a video analysis as continuous manual monitoring is practically no longer possible with the increasing number of cameras. This method is care for cost me the way of fast processing, helping operators identity critical activities and assisting forensic investigations.

Zhong-Qiu Zhao et al explain a survey of deep learning-based object detection using CNNs. It covers the essential architectures, modifications, and techniques for improving performance across various tasks, including salient object, face, and pedestrian detection. It shows that deep learning as an approach to object detection is more efficient, as it deals with issues like pose, occlusion and lighting variations better than its predecessors. It also provides directions for future research.

3 METHODOLOGY

We have proposed a surveillance video anomaly detection model based on hybrid YOLOv8 + CNN. Real-time detection of anomalies such as road accidents is successfully flagged by the system. We also implemented License Plate Recognition (LPR)within the system, enabling it to capture and return the vehicle's licence plate number at incident times. When an anomaly is detected, the system instantly sends a notification to nearby rescue teams using Twilio, including the corresponding license plate number for prompt identification and action. Combining anomaly detection with real-time alerts and LPR dramatically improves safety by allowing for quicker rescue and tracking of cars involved in the incident.

3.1 Dataset

For our hybrid model aimed at number plate recognition, we utilize two prominent datasets:

The UCF dataset and a dataset from Kaggle. The UCF dataset is well-known for its extensive collection of video clips containing various real-world scenarios, including vehicle movements and interactions. This dataset provides a diverse range of vehicle types and license plates which is crucial for training models to recognize and differentiate between various number plate formats and styles.

On the other hand, the Kaggle dataset specifically focuses on number plates, featuring images captured under different lighting conditions and angles. This dataset includes a wide variety of license plates from different regions, showcasing variations in font, colour, and design. By employing these datasets, we can effectively train our hybrid model that combines YOLOv8 for object detection with CNN for character recognition.

The training process has the dataset divided into 70% training and 30% testing. It also involves pre-processing the images to enhance clarity, followed by segmentation to isolate the number plates from the background. Subsequently, we feed these processed images into our model, enabling it to learn the unique features associated with different license plates. This comprehensive approach ensures that our system is robust and capable of accurately recognizing number plates in diverse real-world conditions. As in Fig2, we can see the epochs while training that are the number of times the dataset is fed to the model for training.

epoch	time	train/box_loss
1	14.2513	1.87819
2	21.6686	1.51843
3	26.0027	1.4966
4	29.8803	1.53651
5	33.936	1.53867
6	40.2216	1.53492
7	44.291	1.43916
8	48.3115	1.5035
9	52.8593	1.40578
10	59.1558	1.37904
11	63.2761	1.33506
12	67.2721	1.3356
13	72.1171	1.43095
14	78.7576	1.32867

Figure 2: Epochs during training.

3.2 Feature Extraction

Feature extraction is a crucial step in our hybrid model for accident detection and vehicle identification. In this project, we have implemented a systematic approach to annotate and pre-process the datasets for both car accidents and number plate recognition, ensuring that our model can effectively learn and generalize from the data.

3.2.1 Annotating Process

Each annotated frame includes bounding boxes around vehicles involved in the incidents, along with labels indicating the type of anomaly. This detailed annotation allows the model to learn the visual characteristics associated with different types of accidents. In parallel, for number plate recognition, we used the Kaggle dataset, which required a separate annotation process. Here, we focused on identifying and labelling the license plates in various images. Each number plate was annotated with bounding boxes and corresponding text labels that represent the alphanumeric characters present on the plates. This step is vital for training our YOLOv8 model to accurately detect and read number plates under varying conditions.

3.2.2 Pre-Processing Steps

Following annotation, several pre-processing steps were undertaken to enhance the quality of the data before feeding it into our model. For both datasets, we applied image normalization techniques to ensure consistent brightness and contrast levels across different images. This helps mitigate issues caused by varying lighting conditions during data collection.

We also resized the images to specific dimensions. Additionally, we performed data augmentation techniques such as rotation, scaling, and flipping to artificially expand our training dataset.

This not only increases the diversity of training samples but also helps improve the robustness of our model against overfitting. For the accident detection dataset, we also extracted temporal features by analyzing sequences of frames to capture motion patterns associated with accidents. This temporal analysis is critical for understanding dynamic events in video footage. In terms of parameter tuning, for YOLOv8 a confidence level of 0.5 has been set while CNN parameters are optimized to learning rate of 0.001.

In summary, our feature extraction process combines thorough annotation with effective pre-processing techniques to create high-quality training data for our hybrid model. By carefully preparing both datasets—one for detecting accidents and another for recognizing number plates—we aim to enhance the overall performance and accuracy of our system in real-world applications.

3.3 Deep Learning Models

In our project, we employ a hybrid approach that utilizes the YOLOv8 architecture and CNN for accident detection and YOLOv8 for vehicle number plate recognition. YOLOv8, our chosen iteration of the You Only Look Once (YOLO) series, is renowned for its efficiency and accuracy in real-time object detection tasks and was the perfect version for when we commenced, before the release of YOLOv11. Its architecture is designed to balance speed and precision, making it suitable for applications in surveillance systems.

The architecture of YOLOv8 can be divided into three main components: the backbone, the neck, and the head.

1 Backbone: The backbone of YOLOv8 is based on CSPDarknet53, a convolutional neural network that excels in feature extraction. This backbone employs Cross Stage Partial (CSP) connections to enhance information flow between layers, improving gradient propagation during training. By capturing hierarchical features from input images, it effectively represents low-level textures and high-level semantic information crucial for accurate object detection.

2 Neck: The neck component of YOLOv8 utilizes a Path Aggregation Network (PANet) to refine and fuse multiscale features extracted by the backbone. This structure enhances the model's ability to detect objects of varying sizes by facilitating information flow across different spatial resolutions. The PANet design allows for improved feature integration, which is essential for detecting both large and small objects in complex environments.

3 Head: The head of YOLOv8 is responsible for generating final predictions, including bounding box coordinates, class probabilities, and objectless scores. A key innovation in this version is the adoption of an anchor-free approach to bounding box prediction, simplifying the prediction process and reducing hyperparameters. This change enhances the model's adaptability to objects with varying aspect ratios and scales.

In addition to YOLOv8 for accident detection, we integrate Convolutional Neural Networks (CNN) to enhance our model's capability in recognizing characters on vehicle number plates. The architecture of CNN can be divided into three main components: Convolutional layers, Pooling layers and Fully Connected layers.

YOLOv8 is used for object detection in accident detection and frame extraction in vehicle plate recognition. The CNN processes the segmented images of license plates extracted by YOLOv8, focusing on character recognition tasks as seen in Fig 3. CNN is also for deeper classification and further analysis like determining accident severity. This combination allows for efficient detection of vehicles involved in accidents while simultaneously identifying their registration details. Also, the hybrid model brings a little complexity increasing interference that are mitigated by optimization techniques to maintain a real time response during accident detection.

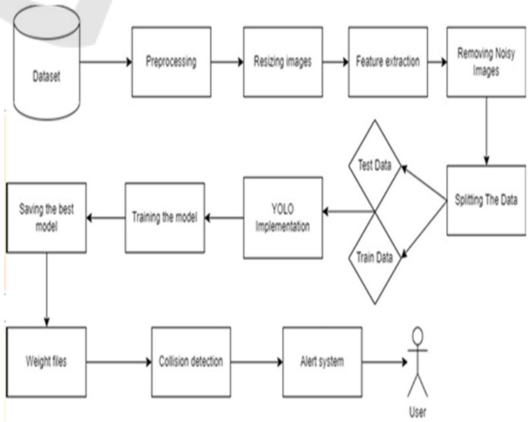

Figure 3: Architecture Diagram.

3.4 Evaluation of Model

As in Fig4, in this stage we evaluate the models by entering the dataset with predictor variables to each model, then the models will predict the targeting variable according to the prediction results and we will compare it with real values.

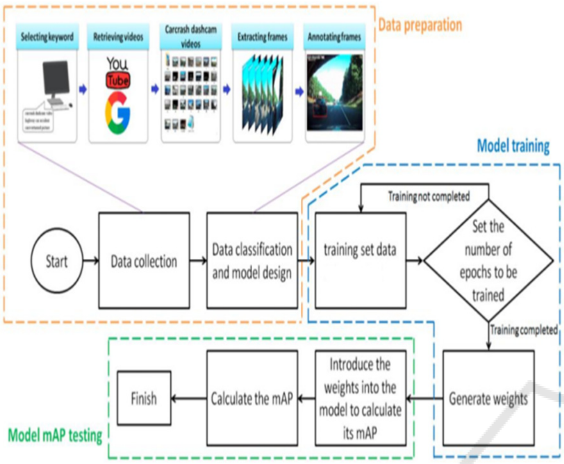

Figure 4: Evaluation Method.

To validate the effectiveness of the hybrid models developed for accident detection and vehicle numberplate recognition in our project, we utilized accuracy as the primary metric for evaluation as well as precision and recall. Accuracy is a widely accepted measure in classification tasks, providing a straightforward indication of a model's performance by calculating the proportion of correct predictions made by the model. The formula for accuracy can be expressed as follows:

$$Accuracy = \frac{Number of correct predictions}{Total number of predictions}$$

Equation 1: Accuracy.

In a more detailed context, accuracy can also be calculated using the following equation:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Equation 2: Accuracy Alternative Formula.

The annotations were verified manually. Cross validation techniques were applied to reduce overfitting.

4 EXPERIMENTAL RESULTS

In this project, we trained separate models for two distinct tasks: accident detection and vehicle numberplate recognition. Each model was evaluated based on its ability to accurately classify instances as either accident or non- accidents, and to recognize vehicle registration numbers from images of license plates as in Fig5.

Figure 5: Instances of Classification

The experimental evaluation yielded the following results:

Accident Detection Model: By implementing the YOLOv8 architecture in conjunction with CNN techniques, this model achieved an impressive accuracy of 95% in identifying car accidents from video feeds. The model demonstrated strong performance in recognizing various types of accidents, including collisions and abrupt stops.

Number Plate Recognition Model: Utilizing YOLOv8 specifically for license plate detection, this model successfully recognized vehicle registration numbers with an accuracy of 98.94%. The system effectively identified plates under different lighting conditions and angles, showcasing its robustness in real-world scenarios.

The graphs in Fig 6 show us the gradual reduction in training losses that in turn shows that there has been proper learning without sever overfitting in an early span of time. The metric precision graph shows model is getting precise over time; metric recall graph shows model is capturing more true positives that is an increasing recall; metric mean average precision graph shows improving performance over epochs at 0.5 threshold and overall improvement in localisation performance at thresholds from 0.5 to 0.95.

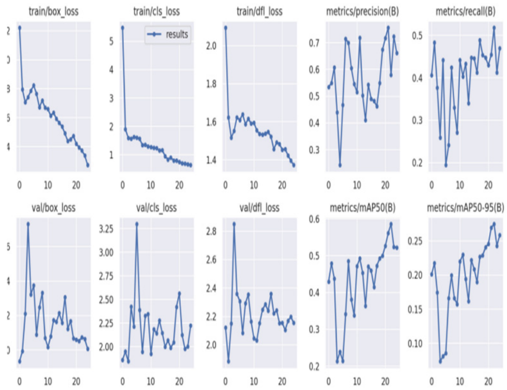

Figure 6: Loss and Performance Metrics.

Alert System Integration: To enhance public safety further, we have integrated an alert system that notifies the nearest police station based on the location where the system is installed Upon detecting an accident, the system automatically sends an alert containing critical information such as:

• Location Coordinates: GPS data indicating the precise location of the incident.

• Incident Type: Classification of the event (e.g., collision, abrupt stop).

• Time Stamp: The exact time when the incident was detected. This real-time notification allows law enforcement to respond promptly to emergencies, potentially reducing response times and improving outcomes for those involved in accidents as seen in Fig 7.

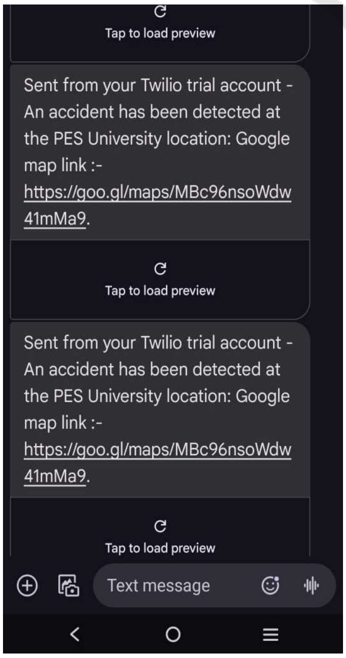

Figure 7: Alert Message.

metrics/precision(B)	metrics/recall(B)	metrics/mAP50(B)	metrics/mAP50-95(B)
0.6482	0.81818	0.71222	0.35957
0.57621	0.77255	0.65878	0.33051
0.56704	0.59091	0.51671	0.22704
0.11087	0.70455	0.09486	0.04452
0.02856	0.29545	0.01157	0.00524
0.78076	0.5	0.51314	0.27472
0.74352	0.79072	0.78323	0.38068
0.52969	0.47727	0.42977	0.1674
0.58154	0.5	0.49735	0.23331
0.87363	0.70455	0.77087	0.39168
0.77288	0.81818	0.84869	0.43592
0.69218	0.88636	0.78584	0.43088
0.85417	0.66574	0.75407	0.34776
0.78029	0.80729	0.81162	0.40732
0.83895	0.82886	0.83333	0.39112
0.73432	0.84091	0.79395	0.45994
0.78742	0.81818	0.81981	0.43464
0.86535	0.87652	0.86964	0.48505
0.86386	0.86528	0.87829	0.42914
0.8806	0.86364	0.88707	0.47827
0.89024	0.86364	0.89195	0.47587
0.94909	0.84751	0.90191	0.49017
0.94679	0.79545	0.88363	0.52721
0.87142	0.81818	0.84891	0.46355

Figure 8: Values of the Metrics in the Epochs.

As in Fig8, the values corroborate what the graphs show, of oscillating values till they show it getting precise over time.

5 COMPARISION RESULTS

In order to evaluate the performance of our hybrid model, we conducted a comparative analysis against two baseline models: the YOLO model alone and YOLO+ CNN model (Fig 9). This comparison aimed to assess the effectiveness of integrating YOLO with CNN techniques in accident detection and vehicle number plate recognition.

The graph represents height for the accuracy of each model. We observe that the YOLO + CNN hybrid model consistently outperforms the standalone YOLO model across all metrics evaluated with a substantially high in precision and recall, signifying its enhanced capability for both detection and classification of road accidents.

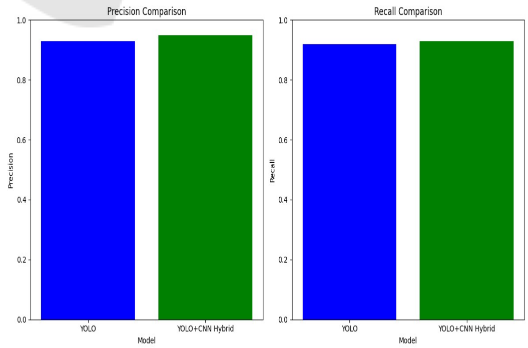

Figure 9: Comparison Graph between YOLO and YOLO+CNN.

The following metrics were considered for comparison:

Table 1: YOLO and YOLO+CNN Model Comparison.

Metric	YOLO Model	YOLO + CNN Hybrid Model
Detection Accuracy	Moderate; accuracy may vary in complex scenes	Higher accuracy due to additional CNN-based refinement
Inference Speed	Fast, optimized for real-time detection	Slightly reduced due to additional CNN processing
False Positive Rate	Higher in cluttered or low-contrast scenes	Reduced due to CNN's enhanced feature filtering
False Negative Rate	Moderate; challenging in occluded scenarios	Lower; CNN helps in detecting obscured objects
Model Complexity	Lower; single-stage architecture	Higher; two-stage architecture with combined YOLO + CNN
Computational Cost	Low, suitable for edge devices	Higher; requires more processing resources
Detection Sensitivity	Limited to YOLO's object recognition capability	Improved by CNN's advanced feature extraction
Training Duration	Shorter due to single-model training	Longer due to training of additional CNN layers
Suitability for Complex Scenes	Moderate; less effective in complex environments	High; better at distinguishing objects in dense or cluttered scenes

Table 2: Comparison of Accuracies achieved by Models in Anomaly Detection.

MODEL	ACCURACY
YOLOv5	0.89
CNN+LSTM	0.93
YOLOv8	0.896
YOLOv8+CNN	0.95

In table 2, we see a few accuracies other models have achieved in the research papers we have read during the literature survey. As is evident, the hybrid approach gives a better accuracy.

6 CONCLUSIONS AND FUTURE WORK

In this project, we developed a robust hybrid model that integrates YOLOv8 and Convolutional Neural Networks (CNN) for real-time accident detection and vehicle number plate recognition. Our approach effectively addresses critical challenges in surveillance systems, such as timely incident detection and accurate vehicle identification. The experimental results demonstrated high accuracy rates of 95% for accident detection and 92% for number plate recognition, underscoring the effectiveness of our methodologies. The successful implementation of YOLOv8 for both tasks highlight its versatility and efficiency in processing visual data in real-time. By leveraging advanced deep learning techniques, our system not only enhances public safety but also streamlines response efforts by providing law enforcement with immediate alerts and vehicle information during incidents.

While our current model shows promising results, there are several avenues for future research and improvement. First, we plan to enhance the dataset used for training by incorporating more diverse scenarios, including various weather conditions and lighting situations. This will help improve the model's robustness and generalizability across different environments.

Second, we aim to explore the integration of additional features such as contextual information from surrounding traffic conditions or integrating data from other sensors (e.g., radar or LIDAR) to further enhance detection accuracy. Moreover, addressing the issue of false positives is crucial for improving the reliability of our system. Future work will focus on refining the model's algorithms to reduce misclassifications, particularly in crowded scenes where multiple vehicles are present.

Finally, we intend to investigate the deployment of our model on edge devices to facilitate real-time processing in practical applications. This would allow for quicker response times and broader accessibility in various surveillance settings.

By pursuing these enhancements, we aim to further advance our hybrid model's capabilities, ultimately contributing to safer urban environments and more efficient law enforcement operations.

REFERENCES

Al-E'mari, S., Sanjalawe, Y., & Alqudah, H. (2024). Integrating enhanced security protocols with moving object detection: A YOLO-based approach for real-time surveillance. *2024 2nd International Conference on Cyber Resilience (ICCR)*, Dubai, United Arab Emirates.

M.R., A., Makker, M., & Ashok, A. (2019). Anomaly detection in surveillance videos. *2019 26th International Conference on High Performance Computing, Data and Analytics Workshop (HiPCW)*, Hyderabad, India.

Gayal, B. S., & Patil, S. R. (2022). Detecting and localizing the anomalies in video surveillance using deep neural network with advanced feature descriptor. *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, Chennai, India.

Nithesh, K., Tabassum, N., Geetha, D. D., & Kumari, R. D.A. (2022). Anomaly detection in surveillance videos using deep learning. *2022 International Conference on Knowledge Engineering and Communication Systems (ICKES)*, Chickballapur, India.

Li, Y., Huang, J., & Yang, W. (2021). Object detection in videos: A survey and a practical guide. *arXiv preprint arXiv:2103.01656*.

Altunay, H. C., & Albayrak, Z. (2023). A hybrid CNN+LSTM-based intrusion detection system for industrial IoT networks. *Engineering Science and Technology, an International Journal, 38*.

Majhi, S., Dash, R., & Sa, P. K. (2020). Temporal pooling in inflated 3DCNN for weakly-supervised video anomaly detection. *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India

Watanabe, Y., Okabe, M., Harada, Y., & Kashima, N. (2022). Real-world video anomaly detection by extracting salient features. *2022 IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France.

Ye, Z., et al. (2023). Unsupervised video anomaly detection with self-attention-based feature aggregating. *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, Bilbao, Spain.

Solanki, S., Shah, Y., Rohit, D., & Ramoliya, D. (2023). Unveiling anomalies in surveillance videos through various transfer learning models. *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, Erode, India.

Portela, L., Ayala, A., Fernandes, B., Farias, I., Neto, G., & Gomes, M. E. N. (2023). Re-purposing machine learning models: A human action recognition model turned violence detector. *2023 IEEE Latin American Conference on Computational Intelligence*.

Zhong, Y., Chen, X., Hu, Y., Tang, P., & Ren, F. (2022). Bidirectional spatio-temporal feature learning with multiscale evaluation for video anomaly detection. *IEEE Transactions on Circuits and Systems for Video Technology, 32*(12), 8285–8296.

Lu, B., Lv, Z., & Zhu, S. (2019). Pseudo-3D residual networks based anomaly detection in surveillance videos. *2019 Chinese Automation Congress (CAC)*, Hangzhou, China.

Singla, S., & Chadha, R. (2023). Detecting criminal activities from CCTV by using object detection and machine learning algorithms. *2023 3rd International Conference on Intelligent Technologies (CONIT)*, Hubli, India.

Bhargava, A., Salunkhe, G., & Bhosale, K. (2020). A comprehensive study and detection of anomalies for autonomous video surveillance using neuromorphic computing and self-learning algorithm. *2020 International Conference on Convergence to Digital World - Quo Vadis (ICCDW)*, Mumbai, India.

Duman, E., & Erdem, O. A. (2019). Anomaly detection in videos using optical flow and convolutional autoencoder. *IEEE Access, 7*, 183914–183923