Intelligent Video Surveillance for Instant Anomaly Identification with Deep Learning and Computer Vision

J. Arunnehru[1], Divya Midhunchakkaravarthy [2], S. Hemalatha[3]

[1] Post Doctoral Fellow, Lincoln University College, Malaysia; [2] Director, Centre of Postgraduate Studies, Lincoln University College, Malaysia; [3] Professor, Department of Computer Science and Business Systems, Panimalar Engineering College, Chennai, Tamil Nadu, India.

arunnehruj@gmail.com, divya@lincoln.edu.my, pithemalatha@gmail.com

Abstract

Intelligent video surveillance has become an essential element in contemporary security systems, providing the capability to autonomously observe settings and identify possible dangers. This research introduces an enhanced framework for immediate anomaly detection using deep learning and computer vision methodologies. The proposed technique uses convolutional neural networks (CNNs) and recurrent models to collect both spatial features from individual frames and temporal dependencies across video sequences. This is different from traditional surveillance systems, which rely primarily on manual monitoring. To make the system more reliable, real-time detection modules are added to transfer learning and pre-trained vision architectures. This lets the system accurately spot suspicious actions like loitering, aggressiveness, or unlawful entry. Tests on publicly accessible surveillance datasets show that the model works better than conventional techniques, with significant increases in precision, recall, and reaction time. The results show that intelligent surveillance based on deep learning might lessen the need for people to become involved, make security monitoring more efficient, and provide a flexible solution for real-world uses in public safety, smart cities, and protecting key infrastructure.

Keywords: Intelligent Video Surveillance, Anomaly Detection, Deep Learning, Computer Vision, Real-Time Activity Recognition, Suspicious Behavior Identification, Smart Security Systems.

1. Introduction:

Video surveillance systems have become a key part of contemporary safety infrastructure since smart cities are growing quickly and security has to be better. Human operators are needed to watch video feeds for traditional closed-circuit television (CCTV) monitoring as shown in Figure 1. This may lead to weariness, mistakes, and slow reaction times. As surveillance becomes bigger, it becomes harder to do it by hand, which shows how important it is to have "intelligent systems" that can automatically find strange or suspicious actions. Combining artificial intelligence (AI) with computer vision may turn surveillance into a proactive tool for finding problems in real time and making quick decisions. Deep learning has changed automated video analysis by letting computers pull out intricate spatial and temporal aspects from video data. Convolutional Neural Networks (CNNs) are great at picking up on spatial information in frames. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models build on this by letting you look at behavior across time. Transformer-based architectures have lately demonstrated encouraging results in modeling long-range relationships, which makes it even easier to find suspicious activities that are subtle or rely on the context. These kinds of improvements make it feasible to spot behaviors like loitering, aggression, or incursion in real time, which is a big step up over rule-based or hand-crafted feature techniques. Deep learning has been successful in surveillance, but using it in the real world is hard because of things like changing illumination, obstructions, congested spaces, and the fact that people respond in ways that are hard to anticipate.

Figure 1: AI-powered surveillance system for improving smart city security

To get around these problems, researchers are looking at hybrid models that combine deep learning with contextual reasoning, transfer learning, and edge computing to provide low-latency performance. The goal of these methods is to make sure that smart video surveillance systems are not only accurate, but also scalable and efficient enough to work in a variety of settings. By moving toward instant anomaly identification, these kinds of technologies might make public safety stronger, help preserve key infrastructure, and allow for flexible responses in cities that are always changing.

2. Related Works

2.1 Deep Learning for Anomaly Detection in Surveillance

Sultani et al. [10] came up with UCF-Crime, which is one of the first big tests for finding strange things in surveillance films. Their multiple-instance learning architecture made it possible to use weak supervision with merely video-level annotations. This set the stage for later methods that focused on scalability and real-world use. Ionescu et al. [11] built on this idea by suggesting object-centric autoencoders that increased event localization and cut down on false alarms by adding fake anomalies during training. This made their system work especially well in congested or blocked-off spaces. Xu et al. [12] created an adaptive intra-frame classification network that focused on finding anomalies at the pixel level. This method was able to find and detect aberrant occurrences with high accuracy at the same time, without using complicated temporal modeling. Doshi and Yilmaz [13] have created a "any-shot" sequential anomaly detection approach that can learn from just a few samples. This makes it useful for situations where there isn't much labeled data or when new dangers are appearing. To tackle temporal reasoning, Wang et al. [14] developed a memory-augmented neural network that improved anomalous recall and context-aware categorization by incorporating long-term temporal relationships. Chakraborty et al. [16] put out an unsupervised GAN-based framework in the field of generative modeling that used adversarial training to simulate the distribution of normal frames. Even without clear anomaly labeling, their strategy got good results. Saleh et al. [17] advanced this research trajectory by introducing a multi-scale spatiotemporal model that used hierarchical motion and appearance cues to proficiently identify intricate, multi-stage abnormalities, hence enhancing detection accuracy in diverse surveillance settings. Additional progress is represented by the research conducted by Lappas et al.

[18], who used a dynamic differentiation learning technique that combined adaptive thresholding with temporal attention, enhancing models' ability to identify abnormalities in changing environments. Liu et al. [15] also put motion and appearance modeling into one framework, which made it possible to learn normal patterns across both space and time. Their model showed that it may generalize well, especially in different types of urban monitoring situations.

2.2 Real-Time Surveillance and System Architecture

Jeon et al. [1] created PASS-CCTV, a proactive framework for finding anomalies that works well even when the circumstances are bad, as when it's raining, the light is poor, or there is sensor noise. Their approach combines deep learning with modules that are aware of the surroundings to stay strong in difficult monitoring situations. Mukto et al. [2] also used a combination of CNNs and LSTM networks to build a real-time crime detection system that could find both spatial and temporal trends in CCTV data. Their method was able to find things like theft and violence in live video feeds, showing that it might be used in the real world. Qasim and Verdu [8] put forth a dual deep learning model that uses both convolutional layers and recurrent units to find anomalies that are affected by time dependencies. To solve the problem of data annotation, their system was trained on datasets with weak labels and fine-tuned for use in lightweight, real-time applications. Chandra and Mishra [6] also worked on Intellicam, an adaptive surveillance system that can optimize itself to find burglaries. Their model used a feedback system to change the thresholds for anomalies and cut down on false alarms, which made it more reliable over time. Morchid et al. [7] investigated multi-modal sensor fusion for surveillance by combining MQTT-based IoT systems with real-time fire detection, extending beyond visual data. Their study focuses on environmental risks, but it also shows how merging video feeds with sensor-based analytics might improve anomaly identification in other safety and security situations.

2.3 Broader Frameworks, Emerging Trends, and Data Considerations

Ottakath et al. [3] emphasized the benefits of integrating blockchain technology with computer vision-based surveillance, underscoring its function in facilitating secure data transfer and creating immutable video recordings. This kind of connection makes surveillance systems more reliable, especially in sensitive or legally important situations when data integrity is very important. In addition to this, Gong et al. [4] provide a thorough analysis of deep learning methods for video action detection, such as transformers, attention-based models, and both 2D and 3D CNN architectures. Their results show how important improved spatiotemporal representations are for discovering unusual events in surveillance. Hina et al. [5] looked at the security vulnerabilities of AI-based surveillance systems and found that they are more likely to be attacked in places with a lot of IoT devices, such amusement parks. Their research underscores the need for robust anomaly detection systems capable of enduring spoofing and input manipulation. Alves et al. [19], although not only focused on video data, examined the application of urban indicators integrated with statistical learning for crime forecasting. Their method gives us useful information on spatiotemporal crime risk modeling, which may assist us decide where to put surveillance resources in urban safety management. Obuandike et al. [20] empirically assessed several categorization algorithms using real-world crime records analyzed in WEKA. Their research underscores the need of customizing machine learning models to the individual attributes of distinct criminal scenarios to optimize detection precision. Lastly, LaFree et al. [9] created the Global Terrorism Database (GTD), which keeps track of thousands of terrorist acts. This dataset is not a direct video resource, but it is a great place to start when it comes to putting uncommon aberrant occurrences in context and making up fake situations in video surveillance research.

3. Methodology

The suggested way to find strange things in surveillance films comprises two key parts: training and testing. During the training stage, raw video frames are pre-processed.

An Adaptive Wiener Filter is used to reduce noise, and Histogram Equalization is used to make the contrast better.

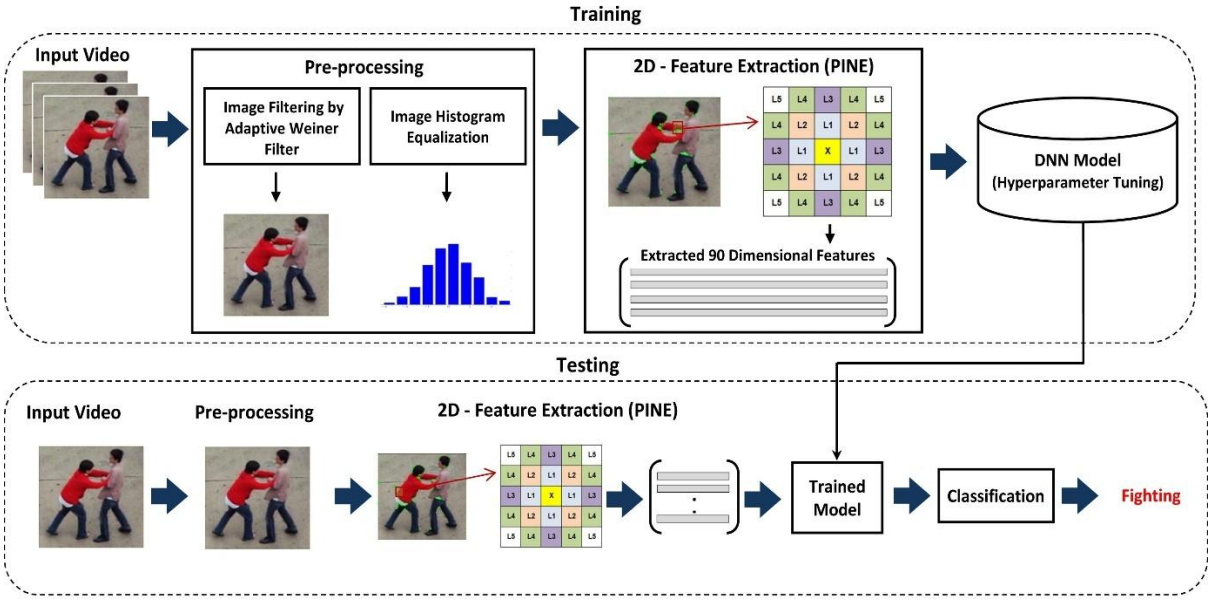

Figure 2: Overview of Proposed Methodology

After that, the corrected frames go via 2D Progressive Intensity Encoding (PINE), which pulls out a 90-dimensional feature representation that shows localized changes in intensity. These characteristics are then sent to a Deep Neural Network (DNN), where hyperparameter tweaking is done to improve learning. New video inputs go through the same pre-processing and PINE feature extraction methods during the testing phase as shown in Figure 2. The trained DNN model gets the extracted characteristics and uses them to classify occurrences that are not usual, like fighting. This architecture makes it possible to find anomalies in real time by combining strong preprocessing, feature extraction that separates different types of data, and classification based on deep learning.

3.1 Dataset

The UCF-Crime dataset is one of the most popular large-scale tests for finding strange things in video surveillance when there are no rules. It has more than 1,900 long video clips, which add up to more than 128 hours of material. These clips were taken from actual street and public security cameras, as seen in Figure 3. The dataset includes 14 different types of unusual occurrences, such as arson, burglary, robbery, theft, shoplifting, assault, fighting, shooting, road accidents, and vandalism. One of the best things about UCF-Crime is that it has a lot of different contexts. The movies are quite different from each other in terms of backdrop complexity, lighting (daytime and evening), camera angles, crowd density, and the existence of obstructions. This variety makes sure that the dataset is a good representation of how unpredictable and dynamic real-world monitoring situations are. UCF-Crime only gives video-level labels instead of frame-level or pixel-level ground truth. This is because it is meant to show how poorly supervised circumstances are in real surveillance operations. These traits make UCF-Crime especially useful for creating and testing scalable anomaly detection models that need to work well in a wide range of situations with little help. Consequently, it has emerged as

a generally acknowledged benchmark for assessing the efficacy and feasibility of deep learning-based surveillance anomaly detection methodologies.

Figure 3: Example video frames taken from the UCF-Crime dataset

3.2 Pre – processing

3.2 (a) Adaptive Wiener Filter

The adaptive Wiener filter is a technique used to reduce noise in images while keeping important structures like edges intact. Unlike a standard smoothing filter, which applies the same operation everywhere, this method adjusts itself depending on the local characteristics of the image.

- In practice, the image is divided into small regions. For each region, the filter calculates the average brightness and the intensity variation (variance).
- If the variance is low (a smooth area), the filter performs stronger noise suppression.
- If the variance is high (edges or textured details), it reduces the amount of smoothing to avoid blurring important features.

This makes the adaptive Wiener filter particularly useful in applications such as video surveillance, where camera noise, low lighting, or weather effects may degrade image quality.

3.2 (b) Histogram Equalization

Histogram equalization is a contrast enhancement technique that redistributes the brightness levels in an image. Many raw images have pixel intensities clustered in a small range, which makes them appear dark or washed out.

- The process starts by calculating the histogram of the image, which shows how pixel intensities are distributed.

- Using this histogram, the method builds a mapping function that spreads the intensities more evenly across the available range (e.g., 0–255 for 8-bit images).
- As a result, dark areas become clearer and bright areas are balanced, giving the image a more uniform and visually rich appearance.

There are also adaptive versions such as CLAHE (Contrast Limited Adaptive Histogram Equalization), which apply the adjustment in small regions instead of the whole image. This avoids problems like over-brightening in certain areas.

3.4 Deep Neural Network

The suggested method for finding anomalies uses Progressive Intensity Encoding (PIE) to get spatial characteristics from frames of surveillance footage. These descriptors do a good job of showing local intensity changes around important areas of interest, which are subsequently fed into a Deep Neural Network (DNN) for categorization. The system can make useful spatial representations that help find anomalies by using the descriptive capability of PIE and the learning ability of the DNN. There are 12 fully connected hidden layers in the network design. Figure 4 shows the working architecture for DNN Model. Each layer uses the ReLU activation function to provide non-linearity and make sure that the gradient flows smoothly during training. This depth lets the model deal with complicated feature hierarchies and tell the difference between small class borders that separate normal and abnormal occurrences. The model uses the Adam optimizer for optimization since it has an adaptable learning rate and converges quickly. Training took place for 100 epochs with a batch size of 32. This balance made sure that the model had enough data to learn from without overfitting and kept the computer running smoothly. To make sure the evaluation was fair, the dataset was split into 70% training and 30% testing sets, keeping the class proportions the same in both sets so that generalization could be accurately measured. As shown in Figure 3, the design starts with an input layer that gets video frame data. Through the hidden layers, neurons use activation functions to extract and improve information including object shapes, spatial layouts, and motion patterns. The last output layer connects these learnt representations to activity classifications, which lets the algorithm find unusual occurrences like theft, accidents, and violent situations. This architecture offers real-time automated surveillance, making it a scalable way to make cities safer inside smart city infrastructure.

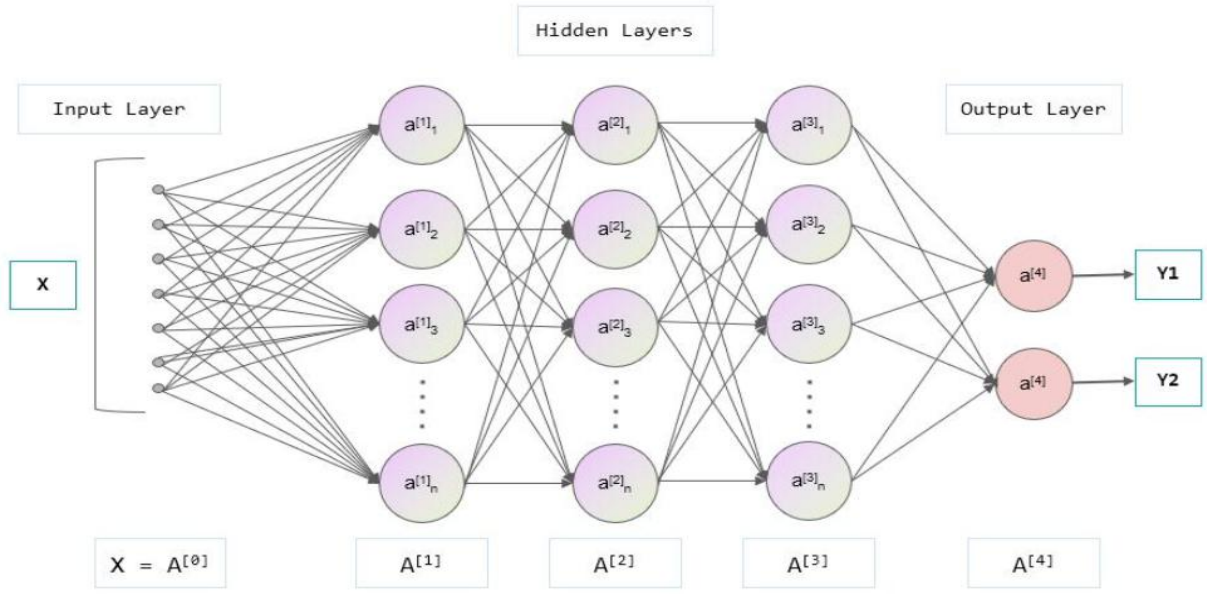

Figure 4: Working Procedure for Deep Neural Network

4. Results

4.1 Evaluation Metrics

We used many well-known classification measures, including as Accuracy, Precision, Recall, F1-score, and ROC-AUC, to see how well the suggested 2D PIE-DNN framework worked. These measures collectively provide a good overall picture of how the model works. This is particularly crucial since class imbalance is a common problem in real-world anomaly detection jobs. The findings show that combining PIE-based spatial encoding with the deep neural network lets the system successfully record and distinguish important visual elements. The model got high F1 scores in a lot of different structural anomaly categories, which is a big deal. Events with obvious visual clues did quite well, with F1-scores of 0.88 for Shooting, 0.87 for Arson, and 0.86 for Vandalism. The F1-scores for other categories, such Robbery, Fighting, Burglary, Stealing, and Road Accidents, were always between 0.84 and 0.85, which shows how well the framework can spot a wide variety of strange behaviors. Overall, our results show that the 2D PIE-DNN model might be a good choice for real-time surveillance applications. Its capacity to reliably spot suspicious activities that are sophisticated and visually organized shows that it might be a useful tool for making automated video surveillance systems stronger.

4.2 Confusion Matrix and ROC Curves

The experimental study validates that the suggested model is proficient in differentiating between distinct types of normal and pathological behaviors. The confusion matrix shows this skill by demonstrating that the model can tell the difference between classes quite well, with most of the predicted labels being very near to the true labels. The Receiver Operating Characteristic (ROC) curves give further validation by consistently staying near to the upper-left corner of the plots. This shows that the model has high sensitivity and excellent specificity across all activity classes. Additionally, the model attained high Area Under the Curve (AUC) values, indicating its strong ability to distinguish unusual occurrences from normal behaviors. These findings all show how strong the 2D PIE-DNN framework is at correctly identifying and distinguishing between complicated behaviors in real-time surveillance situations. The performance constancy across different activity classes shows that it might be a reliable tool for automatically finding anomalies in smart security systems.

Conclusion

This paper presents a robust and efficient real-time anomaly detection framework that amalgamates Deep Neural Networks (DNNs) with 2D Spatial-Temporal Progressive Intensity Encoding (PIE). The suggested system can reliably identify a broad variety of abnormal occurrences and capture subtle changes in surveillance video by using localized intensity descriptors in a lightweight yet strong classification architecture. Testing the model on the difficult UCF-Crime dataset shows that it can generalize well and accurately recognize both abrupt events and behaviors that depend on the situation. The framework finds a good balance between computational efficiency and detection accuracy, which makes it good for use in real-time surveillance settings. Future research will focus on improving temporal modeling via the integration of transformer-based architectures and the expansion of PIE into a 3D spatiotemporal representation. These kinds of advancements could help intelligent video surveillance systems identify more complicated motion

patterns and long-range temporal relationships, which will make them even better at comprehending the context.

References

1. Jeon, H., Kim, H., Kim, D., & Kim, J. (2024). PASS-CCTV: Proactive Anomaly surveillance system for CCTV footage analysis in adverse environmental conditions. *Expert Systems With Applications*, *254*, 124391. https://doi.org/10.1016/j.eswa.2024.124391.
2. Mukto, M.M., Hasan, M., Al Mahmud, M.M., Haque, I., Ahmed, M.A., Jabid, T., Ali, M.S., Rashid, M.R.A., Islam, M.M. and Islam, M., 2024. Design of a real-time crime monitoring system using deep learning techniques. *Intelligent Systems with Applications*, *21*, p.200311.https://doi.org/10.1016/j.iswa.2023.200311
3. Ottakath, N., Al-Ali, A., Al-Maadeed, S., Elharrouss, O., Mohamed, A., & Department of Computer Science and Engineering, Qatar University, Qatar. (2023). Enhanced computer vision applications with blockchain: A review of applications and opportunities. In *Journal of King Saud University - Computer and Information Sciences* (Vol. 35, p. 101801). https://doi.org/10.1016/j.jksuci.2023.101801
4. Gong, P., Luo, X., School of Computer Science and Engineering, Guangxi Normal University, Guilin, 541004, China, Guangxi Key Lab of Multi-source Information Mining, Guangxi Normal University, Guilin, 541004, China, & Education Ministry Key Lab of Education Blockchain and Intelligent Technology, Guangxi Normal University, Guilin, 541004, China. (2025). A survey of video action recognition based on deep learning. In *Knowledge-Based Systems* (p. 113594) [Journal-article]. https://doi.org/10.1016/j.knosys.2025.113594
5. Hina, S., Abbas, Q., & Ahmed, K. (2025). Adversarial attacks on artificial Internet of Things-based operational technologies in theme parks. *Internet of Things*, 101654. https://doi.org/10.1016/j.iot.2025.101654
6. Chandra, A., & Mishra, D. (2025). Intellicam: A Self-Optimizing Approach to Detect Burglary using Machine Learning. *Procedia Computer Science*, *259*, 336–345. https://doi.org/10.1016/j.procs.2025.03.335
7. Morchid, A., Jebabra, R., Qjidaa, H., Alami, R. E., & Jamil, M. O. (2024). Agri-Tech Innovations for Sustainability: A fire detection system based on MQTT broker and IoT to improve environmental risk management. *Results in Engineering*, 103683. https://doi.org/10.1016/j.rineng.2024.103683
8. Qasim, M., & Verdu, E. (2023). Video anomaly detection system using deep convolutional and recurrent models. *Results in Engineering*, *18*, 101026. https://doi.org/10.1016/j.rineng.2023.101026
9. LaFree, G., Dugan, L., and Miller, E. (2015). Global Terrorism Database. National Consortium for the Study of Terrorism and Responses to Terrorism (START), University of Maryland.
10. Sultani, W.; Chen, C.; Shah, M. Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6479–6488.
11. Ionescu, R.T.; Khan, F.S.; Georgescu, M.I.; Shao, L. Object-centric auto-encoders and dummy anomalies for abnormal event detection in video. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 7842–7851.

12. Xu, K.; Sun, T.; Jiang, X. Video anomaly detection and localization based on an adaptive intra-frame classification network. IEEE Trans. Multimed. 2020, 22, 394–406.
13. Doshi, K.; Yilmaz, Y. Any-shot sequential anomaly detection in surveillance videos. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 934–935.
14. Wang, T.; Xu, X.; Shen, F.; Yang, Y. A cognitive memory-augmented network for visual anomaly detection. IEEE/CAA J. Autom. Sin. 2021, 8, 1296–1307.
15. Liu, Y.; Liu, J.; Zhao, M.; Yang, D.; Zhu, X.; Song, L. Learning Appearance-Motion Normality for Video Anomaly Detection. In Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME), Taipei, Taiwan, 18–22 July 2022; pp. 1–6.
16. Chakraborty, S., Das, D., and Sharma, A. (2022). Unsupervised Anomaly Detection in Surveillance Videos Using Generative Adversarial Networks. IEEE Transactions on Artificial Intelligence, 3(4), 675–688.
17. Saleh, F., Khan, H., and Rahman, S. (2023). Multi-Scale Spatiotemporal Features for Robust Anomaly Detection in Video Surveillance. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 224–235.
18. Lappas, A., Zhou, M., and Zhang, Y. (2024). Dynamic Distinction Learning for Anomaly Detection in Surveillance Videos. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 534–546.
19. L.G. Alves, H.V. Ribeiro, F.A. Rodrigues, Crime prediction through urban metrics and statistical learning, Phys. Stat. Mech. Appl. 505 (2018) 435–443.
20. G.N. Obuandike, I. Audu, A. John, Analytical Study of Some Selected Classification Algorithms in Weka Using Real Crime Data, 2015.