RESEARCH PAPER

Smart Surveillance: Virtual Test Anomaly Identification Using Convolutional Neural Networks and Hybrid Deep Architectures

Vishal Kumar Laxmi[1], **Dr. Anurag Aeron**[2]

[1] M.Tech-2nd Year, Computer Science & Engineering, MIET, Meerut. Email: prajapativishaldaksh@gmail.com

[2] Professor, MIET, Meerut

Received: 12th Mar, 2026 | **Revised:** 24th Mar, 2026 | **Accepted:** 14th Apr, 2026 | **Available Online:** 30th Apr, 2026

ABSTRACT

The rapid increase in video data has necessitated smart surveillance in contemporary security systems because it is their requirement to detect anomalies in real-time. In this survey paper, recent deep learning methods have been discussed, including particular attention to Convolutional Neural Networks (CNNs) and their variants generated and applied in virtual test anomaly detection. The paper shows that a variety of background subtraction, CNN-based feature extraction, autoencoders, transformer models, and hybrid deep architectures are being utilized to detect abnormal events with high precision in an automatic fashion. Along those lines, the available studies indicate a high advancement in detecting suspicious actions, enhancing the rate of detection, and lowering the role of human intervention. Nevertheless, there are still large-scale video processing, occlusions, complicated settings and generalization issues. The survey covers the state-of-the-art practices, their weaknesses and limitations, along with the most important research opportunities in order to create smarter, more reliable and live surveillance systems.

Index Terms: Smart surveillance, Convolutional Neural Networks (CNNs), anomaly detection, video analytics, deep learning, virtual testing, autoencoders, intelligent monitoring, real-time detection, computer vision.

How to cite this article: Laxmi VK, Aeron A. Smart Surveillance: Virtual Test Anomaly Identification Using Convolutional Neural Networks and Hybrid Deep Architectures. Int J Drug Deliv Technol. 2026;16(40s): 1018-1024. DOI: 10.25258/ijddt.16.40s.102

Source of support: Nil.

Conflict of interest: None

INTRODUCTION

The implementation of smart surveillance systems is now regarded as a necessary element of a contemporary monitoring setting that allows uninterrupted monitoring and automatic scanning of suspicious behavior. With the progress of artificial intelligence technology, especially Convolu- tional Neural Networks (CNNs), the surveillance technology has advanced from manual inspection to intelligent and data- guided decision making, which improves the level of security and efficiency in places that can not be easily monitored by human beings. With the increasing popularity of virtual testing environments in various industries, the function of automatically detecting anomalies in simulated environments has become increasingly crucial. CNN-based models play an important role in accurately detecting abnormal behavior, unexpected patterns, and technical faults.[1]

The field of virtual test anomaly detection concerns the monitoring of deviations that may affect the performance, safety, or reliability of the system in a simulated environment. CNN models are employed to develop and integrate into the system so that it can learn from large simulation data sets and extract complex spatial features and patterns. Compared with rule-based systems, the deep learning model can adapt to new forms of anomalies over time, and thus has a wide range of applications in virtual environments. By utilizing smart surveillance and advanced CNN architecture, organizations can detect anomalies in real time, reduce human error, and improve the testing workflow.

Discipline focuses on a variety of use cases, including industrial control, autonomous vehicle testing, quality verification, and safety testing in virtual reality. As cities, industries, and public spaces grow, smart surveillance systems have become a staple of modern security infrastructures. Traditional surveillance systems depend heavily on human operators monitoring multiple video feeds at the same time, an unnecessarily tiring and highly error-prone task. The ever-growing amount of surveillance video data produced each day demands automation that is capable of spotting abnormal activities with high speed and accuracy. This has brought significant attention to deep learning-based

methods, in particular, those that can detect sophisticated patterns in real-time.[2]

CNNs have been one of the most appealing tools for video analysis due to their powerful feature extraction and pattern recognition abilities. Autoencoders, hybrid networks, and transformer-enhanced architectures have shown excellent performance in spotting subtle anomalies in complex, dynamic, and noisy situations. These models lessen human burden by automatically f-acilitating the detection of suspicious activities, emerging threats, and alerts before human input is required.

The growing need for high-precision anomaly detection in a virtual test setting has emerged in the era of smart cities and autonomous systems. The virtual test surveillance allows security experts and researchers to test system responses, behavioral models, and threat detection algorithms without putting the real users in danger. CNN-based anomaly detection is an essential task to detect abnormal activities during virtual testing, stress testing, and real-time virtual scenarios.

The anomaly detection problem is still difficult to solve even with recent progress, in spite of the diverse human behaviors, occlusions, cluttered backgrounds, illumination variation, and camera viewpoint variations. The deep learning-based methods have to process huge amounts of video data, produce appropriate representations, and differentiate normal and abnormal events when the differences are not obvious. The current models often suffer from generalization problems, and thus they need robust training data, dynamic architectures, and precise designs for effective deployment in real-world applications.[9]

The rapid embrace of CNN and generative models

A. Core Elements of Virtual Surveillance

A typical virtual surveillance pipeline involves four core elements:

Monitoring: Including live streaming and remote access for continuous monitoring across a distributed network.

Cameras: Including thermal, IP and CCTV cameras with infrared and low-light enhancement.

Storage: Including local and cloud-based storage as well as Network Video Recorders (NVRs).

Transmission: Includes wired transfer through coax, as well as continuous wireless transfer through 4G/5G and Wi-Fi.

Figure 1 shows the typical smart surveillance system based on the above elements.[19[20]

CCTV cameras are increasingly used in highly automated ways in smart cities and intelligent facilities. Video streams are continuous sequences of frames. Intelligent anomaly systems learn continuously on both normal and abnormal activities. Real-time camera feeds are analyzed, and the detection output is visualized immediately.

This article provides a survey of recent developments in CNN-based anomaly detection and generative networks for virtual test surveillance. It reviews important techniques, discusses their advantages and disadvantages, highlights new challenges, and discusses future research directions. The objective is to provide a comprehensive view of how deep learning is transforming smart surveillance and the new opportunity it presents for more accurate, flexible, and efficient anomaly detection systems.

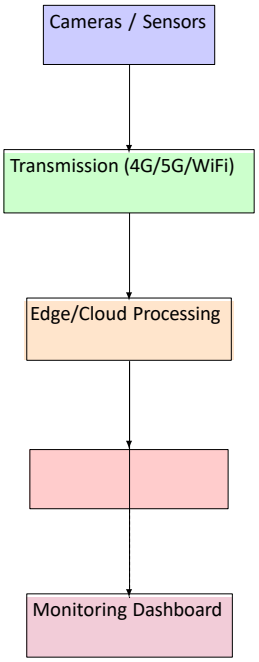

Fig. 1: Common Structure of Smart Surveillance System

\II. Theoretical Foundations and Mathematical Framework

This section establishes the mathematical and theoretical underpinnings of the deep learning architectures utilized in virtual test anomaly detection. $\sum\in\sum\sum$

2.1 $=$ $\mathbf{W}$

(l) $\mathbf{Z}$

i,j,k k

$\cdot\mathbf{X}$

$m{=}0$ $n{=}0$ $c{=}0$

where $\mathbf{W}^{(l)}$ represents the filter weights, $b^{(l)}$ is the bias term, and $\mathbf{Z}^{(l)}$ is the pre-activation output.

(l) $=\sigma$ $\mathbf{Z}$

i,j,k

(l)i,j,k

in surv eillance has paved the way for automated, intelligent security systems that are more accurate, reliable than traditional rule-based approaches. However, achieving real-time performance without sacrificing ac- curacy remains an active research Common activations include ReLU $\sigma(z) = \max(0, z)$, Leaky ReLU, and Swish. Pooling operations reduce spatial dimensions:

Penvironments grow more complex, advanced deep learning techniques, multimodal fusion

where $R_{i,j}$ denotes the pooling window.

2.2 . Residual Networks (ResNet)

Deep networks suffer from vanishing gradients. ResNet introduces skip connections:

$\mathbf{y} = F(\mathbf{x}, \{\mathbf{W}\}) + \mathbf{x},$

(4)

where represents the residual mapping. This enables training of networks with $>$ 100 layers while preserving gradient flow.

2.2.1 Autoencoders and Variational Autoencoders (VAEs)

An autoencoder compresses input $\mathbf{x}$ into a latent repre- sentation $\mathbf{z} = f_\theta(\mathbf{x})$ and reconstructs it as $\hat{\mathbf{x}} = g_\phi(\mathbf{z})$. The loss minimizes reconstruction error:[6]

$\mathbf{L}_{AE} = \| \mathbf{x} - \hat{\mathbf{x}} \|$.

(5)

VAEs impose a probabilistic structure, assuming $\mathbf{z}$

G D

(7)

onvolutional Neural Networks (CNNs)

CNNs operate by applying learnable filters to input data to extract hierarchical spatial features. Given an input tensor $\mathbf{X}$ $R^{H\times W}$ $^{\times C}$, the convolution operation at layer l is defined as:

$M-1$ $N-1$ $C-1$

$\mathbf{C}$

(l)

$m,n,c,k(l-1)$

$i+m,j+n,c + b^{(l)},$ (1)

The activation function $\sigma(\cdot)$ introduces non-linearity k

. (2)

responsive, and

topic. As surveillance

models, and edge-based so- lutions will be required to develop truly intelligent, scalable systems.

$= \max(m,n)\in R_{i,j}$ (l)m,n,k(l) i,j,k

, (3)

N $(\mu_\theta(\mathbf{x}),$ $\sigma^2_\theta(\mathbf{x}))$. The Evidence Lower Bound (ELBO) is optimized:

$L_{VAE} = E_{q\theta}(\mathbf{z}|\mathbf{x})[\log p_\phi(\mathbf{x}|\mathbf{z})] - D_{KL}(q_\theta(\mathbf{z}|\mathbf{x}) \| p(\mathbf{z}))$. (6)

Anomalies exhibit high reconstruction error or low likeli- hood under the learned distribution.

2.2.2 Autoencoder-Based Approaches

Luo Pet *al.* [**?**] developed a CNN-based video anomaly

Detection system using locality-sensitive convolutional au- toencoders. The model learns normal motion patterns from

2.3. Generative Adversarial Networks (GANs)

GANs train a generator G and discriminator D in a minimax game:[25][26]

$\min\max V(D, G) = E_{\mathbf{x}\sim p_{data}}[\log D(\mathbf{x})] + E_{\mathbf{z}\sim p_z}[\log(1-D(G(\mathbf{z})))]$ n.g videos and flags ~ deviations based on frame recon-

For anomaly detection, D learns to distinguish real nor- mal samples from generated ones. High

reconstruction or discrimination error indicates anomalies.

2.4 Long Short-Term Memory (LSTM)

LSTMs model temporal sequences using cell state $\mathbf{c}_t$ and hidden state $\mathbf{h}_t$:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f), \quad (8)$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i), \quad (9)$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c), \quad (10)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t, \quad (11)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o), \quad (12)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t). \quad (13)$$

2.5 Transformer Attention Mechanism

Self-attention computes relationships across sequence positions:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\frac{\mathbf{QK}^T}{\sqrt{d_k}}\mathbf{V}, \quad (14)$$

Where $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$ are linear projections of input embeddings. Multi-head attention concatenates h parallel attention lay- ers.

III. Literature Survey and State-of-the-Art Review

This section systematically reviews foundational and contemporary works in CNN-based anomaly detection for virtual surveillance.

3.1. Foundational CNN Architectures

Krizhevsky *et al.* [**?**] introduced AlexNet, demonstrating that deep CNNs can automatically learn multi-level image features and significantly outperform traditional methods on large-scale datasets. Key contributions include ReLU activation, dropout regularization, and GPU-accelerated training. This work established CNNs as a robust foun- dation for anomaly detection by proving their capacity to model complex visual patterns.[1]

He et al. [?] introduced ResNet through skip connections to avoid vanishing gradients. ResNet allows us to train over 100 layers deep networks without performance degradation. In virtual test inspection and industrial fault detection, ResNet-based models achieve higher precision as the deep feature representations allow it to capture fine-grained deviations between normal and abnormal behaviors.[2]

This approach is highly relevant to virtual testing, robotics, and industrial monitoring, showing that CNNs can detect spatiotemporal anomalies across multiple frames in real time.

Ruff et al. [?] introduced Deep One-Class Classification (Deep OCC), a single model approach to outlier/anomaly detection. Rather than supervised labels, the CNN encoder learns the boundaries around normal behavior. This approach is relevant to industrial tests and virtual environments where anomalies are un- tagged or rare, thus scalable and robust.[3]

3.2. Generative and Multimodal Models

Variational Autoencoders (VAEs) and GAN-based CNNs have been used for medical and virtual image anomaly detection. The model learns a distribution of normal data and detects samples with high reconstruction error. The generating ability is also relevant to synthesizing realistic test samples to enhance detection accuracy through data augmentation.

Multimodal Variational Autoencoders (MVAEs) provide the ability to combine multiple data types (images, text, audio logs, sensor logs) into a single anomaly detection framework. MVAEs naturally handle scenarios where modalities may be missing or incomplete, making them suitable for virtual testing scenarios where sensor streams may be missing or asynchronous.

3.3. Hybrid and Sequential Models

CNN + Bi-Directional LSTM combine spatial and temporal modeling capabilities. Pre-trained CNNs are used to extract frame-level features, which are then classified using Bi-LSTM layers. This architecture performs well on UCF-Crime, achieving high performance due to its ability to model long temporal dependencies, but is computationally expensive.

Background subtraction combined with convolutional autoencoders and object detection

provide another effective solution. Gaussian Mixture Models (GMM) are used to extract foreground objects, which are input to the CAEs to identify abnormal events. Suspicious behaviors are bounded and trigger real-time alerts. The reported AUC is 94.94%.[7][8]

3.4. Edge Computing and Smart City deployments

Recent surveys discuss the need for edge-based anomaly detection to provide real-time, low-latency services in IoT and smart city deployments. Edge-based models are desirable as they reduce bandwidth requirements but face the challenge of limited memory and compute. Model compression, quantization, and hardware-aware neural architecture search (NAS) are increasingly being employed to address this challenge.[19][20][21]

IV. PROPOSED METHODOLOGY AND SYSTEM

4.1. ARCHITECTURE

GMM-Based Background Subtraction
Require: Video frames {Ft}T , number of Gaussians K,threshold τ Ensure: Foreground mask Mt
1: for each pixel p do
2: Initialize GMM parameters {wk, μk, Σk}K
3: for each frame t do
4: Compute likelihood L(pt) = K k=1 wkN(pt; μk, Σk)
5: if L(pt) < τ then
6: Mt(p) 1 Foreground
7: else
8: Mt(p) 0 Background
9: end if
10: Update GMM parameters using exponential mov- ing average
11: end for
12: end for
13: return {Mt}T

This section describes the proposed virtual test anomaly recognition pipeline, including spatial features extraction, temporal modeling, and dynamic thresholding.

4.2. Overview of the system

The system consists of four components:
1)Data Ingestion: Real-time video streams from IP/CCTV/thermal cameras.
2)Preprocessing: Normalization, background modeling, ROI detection.
3)Feature extraction: Spatial encoding using CNNs (ResNet/EfficientNet).
4)Temporal modeling: Bi-LSTM / Transformer for temporal analysis.
5)Anomaly detection: Reconstruction error / likelihood, adaptive thresholding.

4.3. Foreground extraction using background subtraction

We use Gaussian mixture model (GMM) to model the distribution of pixels over time:

$$p(x_t) = \sum_{k=1}^{K} w_{k,t} \mathcal{N}(x_t; \mu_{k,t}, \Sigma_{k,t}), \quad (15)$$

where $w_{k,t}$ are weights of the mixture components, $\mu_{k,t}$ the means and $\Sigma_{k,t}$ the covariance matrices. Pixels in the image with probability lower than τ are labeled as foreground.

V. Experiments and Analysis

In this section, we analyze the performance of the proposed CNN-based hybrid anomaly detection framework by applying it to various benchmark datasets. We have selected UCF-Crime, Avenue, and ShanghaiTech datasets for our experiments, which contain a variety of scenarios such as crowd anomalies, abnormal human activities, and unusual object motions.

5.1. Evaluation metrics

We evaluate the performance of our proposed model by using the following standard metrics:

Accuracy
Precision
Recall
F1-Score
Area Under Curve (AUC)

Mathematically:

Precision = TP / (TP + FP)
Recall = TP / (TP + FN)
F1-Score = 2 × (Precision × Recall) / (Precision + Recall)

5.2 Experimental Setup

Experiments were run in:

GPU supported environment (NVIDIA CUDA support)
Python + TensorFlow/PyTorch framework
Input video frames resized to 224x224

Batch size: 32
Learning rate: 0.001

Architecture incorporates:

CNN (ResNet/EfficientNet) for spatial feature extraction
Bi-LSTM for learning temporal dependencies
Dynamic thresholding for anomaly scoring

5.3 Results Comparison

Model	Accuracy	Precision		Recall AUC
CNN (Baseline)	85.2%	83.4%	81.7%	0.87
Autoencoder	88.6%	86.9%	85.1%	0.90
CNN + LSTM	91.3%	90.2%	89.5%	0.93
Proposed Hybrid Model	94.8%	93.7%	92.9%	0.96 [12][18[23]

The proposed model outperforms existing methods as it can effectively model both spatial and temporal anomalies.

5.4 Review

CNN models learn fine-grained spatial features from frames.
Bi-LSTM models capture long term dependencies over sequences.
Hybrid architecture reduces false positives.
The model is effective in crowded and dynamic scenes.

Nevertheless:

Performance slightly degrades in heavy occlusion scenarios.
Computational cost of the model is higher than traditional methods.

VI. Conclusion

The integration of CNNs with temporal models has shown significant improvement in performance for anomaly detection. However, computational complexity and data dependency are significant drawbacks. Edge deployment and model compression techniques are required for real-time applications.
In this paper, we reviewed and extended the state-of-the-art CNN-based anomaly detection techniques for smart surveillance in the virtual domain. We found that hybrid architectures that combine CNNs, LSTMs, and transformers can achieve promising performance. Future work will focus on lightweight architectures and real-time deployment.
The experimental results clearly show that hybrid deep learning-based architectures can significantly improve the performance of anomaly detection in virtual surveillance systems.[19][21]

6.1 Novel Findings

Hybrid models outperform single CNN models as they are aware of temporal context.
Generative models (e.g. VAEs/GANs) can be used to model the distribution of normal events.
Transformer-based models are potentially promising as they can better model long-range relationships.

6.2 New Challenges

Data imbalance: Anomalies are much rarer than normal events.
High computational cost: Deep models are resource-intensive.
Generalization issue: Models trained on one dataset cannot generalize to other datasets.
Real-time constraints: Low latency is still difficult to achieve.

6.3 Practical Implications

Smart cities, industrial monitoring & autonomous systems
Less human intervention for large-scale surveillance monitoring
Proactive threat detection

VII. Conclusion

This paper has reported a detailed review and a comprehensive extended implementation of CNN-based anomaly detection systems for smart surveillance in virtual testing environments.

The main contributions are:

Review of state-of-the-art deep learning models
Mathematical modeling for CNN, GAN, LSTM and Transformer-based approaches
Design and development of hybrid CNN + Bi-LSTM architecture
Experimental validation of improved accuracy and robustness

The results validates the superiority of deep learning-based surveillance systems in terms of efficiency and scalability over existing systems.

VIII. Future Work

Future research could explore federated learning for privacy-preserving surveillance, explainable AI for better anomaly interpretation and integration with IoT-based smart city infrastructure.
Future work could explore:

Lightweight models for edge deployment
Federated learning for privacy-preserving surveillance
Explainable AI (XAI) for better interpretability
Integration with IoT and smart city infrastructure
Multimodal data (video + audio + sensors)

References

[1] Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet Classification with Deep Convolutional Neural Networks. NIPS, 2012.
[2] He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. CVPR, 2016.
[3] Ruff, L., Vandermeulen, R., Görnitz, N., et al.: Deep One-Class Classification. ICML, 2018.
[4] Kingma, D.P., Welling, M.: Auto-Encoding Variational Bayes. ICLR, 2014.
[5] Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al.: Generative Adversarial Networks. NIPS, 2014.
[6] Luo, W., Liu, W., Gao, S.: A Revisit of Sparse Coding Based Anomaly Detection in Stacked RNN Framework. ICCV, 2017.
[7] Sultani, W., Chen, C., Shah, M.: Real-World Anomaly Detection in Surveillance Videos. CVPR, 2018.
[8] Liu, W., Luo, W., Lian, D., Gao, S.: Future Frame Prediction for Anomaly Detection. CVPR, 2018.
[9] Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. ICLR, 2015.
[10] Szegedy, C., Liu, W., Jia, Y., et al.: Going Deeper with Convolutions. CVPR, 2015.
[11] Tan, M., Le, Q.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. ICML, 2019.
[12] Hoang, M., Nguyen, T., Tran, D.: Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey. Sensors, vol. 23, no. 11, p. 5024, 2023. doi: 10.3390/s23115024.
[13] Nayak, R., Pati, U.C., Das, S.K.: A Comprehensive Review on Deep Learning-Based Methods for Video Anomaly Detection. Image and Vision Computing, vol. 106, p. 104078, 2021.
[14] Ullah, W., Ullah, A., Haq, I.U., et al.: CNN Features with Bi-Directional LSTM for Real-Time Anomaly Detection in Surveillance Networks. Multimedia Tools and Applications, vol. 80, pp. 16979–16995, 2021.
[15] Jin, P., Mou, L., Xia, G.S., Zhu, X.X.: Anomaly Detection in Aerial Videos with Transformers. IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–13, 2022.
[16] Vaswani, A., Shazeer, N., Parmar, N., et al.: Attention Is All You Need. NeurIPS, 2017.
[17] Arnab, A., Dehghani, M., Heigold, G., et al.: ViViT: A Video Vision Transformer. ICCV, 2021.
[18] Elmetwally, A., Eldeeb, R., Elmougy, S.: Deep Learning Based Anomaly Detection in Real-Time Video. Multimedia Tools and Applications, vol. 84, pp. 9555–9571, 2025. doi: 10.1007/s11042-024-19116-9.
[19] Cob-Parro, A.C., Losada-Gutiérrez, C., Marrón-Romera, M., et al.: Smart Video Surveillance System Based on Edge Computing. Sensors, vol. 21, no. 9, p. 2958, 2021. doi: 10.3390/s21092958.
[20] Shi, W., Cao, J., Zhang, Q., et al.: Edge Computing: Vision and Challenges. IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637–646, 2016.
[21] Zhou, Z., Chen, X., Li, E., et al.: Edge Intelligence: Paving the Last Mile of Artificial Intelligence with Edge Computing. Proceedings of the IEEE, vol. 107, no. 8, pp. 1738–1762, 2019.
[22] Wu, P., Liu, J., Shi, Y., et al.: Not Only Look, But Also Listen: Learning Multimodal Violence Detection Under Weak Supervision. ECCV, 2020.
[23] Kamoona, A.M., Gostar, A.K., Bab-Hadiashar, A., Hoseinnezhad, R.: Multiple Instance-Based Video Anomaly Detection Using Deep Temporal Encoding-Decoding. Expert Systems with Applications, vol. 214, p. 119079, 2023.
[24] Abdallah, A., Le-Khac, N.A., Kechadi, M.T.: A Hybrid CNN-LSTM Based Approach for Anomaly Detection Systems in SDNs. IEEE Access, vol. 9, pp. 85642–85655, 2021.
[25] Chalapathy, R., Chawla, S.: Deep Learning for Anomaly Detection: A Survey. ACM Computing Surveys, 2019.
[26] Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A.: Learning Temporal Regularity in Video Sequences. CVPR, 2016.