International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 4 (2025)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 4, 2025

Research Paper

AUTOMATED RECOGNITION OF UNUSUAL HUMAN BEHAVIOR IN CCTV VIDEOS USING NEURAL NETWORKS

[1] Ms.K.AKHILA , [2] DEVATHALA RANITHA ,[3]K SHEETAL KOTHARI ,[4] SVINJAMURI SHRAVAN TEJA ,[5] BASLINGWAR GAJANAN

[1] *Assistant Professor, Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Malla Reddy College of Engineering , Hyderabad, India.*

[2,3,4,5] *Students, Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Malla Reddy College of Engineering, Hyderabad, India.*

ABSTRACT:

Suspicious human activity recognition has become a critical component of modern intelligent video surveillance systems, driven by the rapid advancement of deep learning and the increasing demand for automated security monitoring. Existing research demonstrates significant progress in activity understanding, anomaly detection, and spatiotemporal feature learning using models such as 3D-CNNs, ConvLSTMs, sparse reconstruction, and Generative Adversarial Networks (GANs) [11–22]. Recent surveys highlight the growing adoption of AI-powered surveillance in public safety, smart cities, and predictive policing applications [2–10, 23–24]. Building upon these developments, this work proposes an enhanced suspicious activity recognition framework that integrates robust person detection, deep spatiotemporal feature extraction, and anomaly classification to accurately identify irregular behaviors in real time. Leveraging insights from sparse coding–based methods, fully convolutional anomaly detectors, and temporal regularity learning, the system aims to improve detection accuracy while reducing false alarms. The study contributes a unified analysis of traditional and contemporary deep learning approaches, emphasizing their strengths, limitations, and applicability to real-world surveillance scenarios. The proposed model aligns with emerging trends in AI-driven security and offers a scalable solution suitable for large-scale urban deployments.

Received: 13-10-2025 Accepted: 22-11-2025 Published: 29-11-2025

I.INTRODUCTION

Intelligent video surveillance has emerged as a critical technology for enhancing public safety, crime prevention, and automated monitoring across urban, industrial, and commercial environments. With the rapid growth of smart cities and advancements in artificial intelligence (AI), modern surveillance systems increasingly rely on deep learning methods to analyze video streams and identify abnormal or suspicious human activities without continuous human supervision. Traditional surveillance frameworks depend heavily on manual observation or handcrafted feature extraction, which are often unreliable, labor-intensive, and prone to missed events in complex or crowded scenes [1], [11], [12]. As a result, research has shifted toward data-driven models capable of learning rich spatiotemporal representations from large-scale video datasets [20], [22], enabling more robust detection of anomalous behaviors.

Deep learning approaches—including 3D Convolutional Neural Networks (3D-CNNs), autoencoders, Generative Adversarial Networks (GANs), and fully convolutional architectures—have demonstrated significant improvements in feature learning, motion understanding, and irregular event classification [15], [18], [19]. These models outperform classical sparse coding and reconstruction-based techniques, which struggle with high variability in human actions, camera noise, occlusions, and dynamic backgrounds [11], [12]. Furthermore, advanced temporal modeling methods such as ConvLSTMs and temporal regularity learning networks have enhanced the capability to detect deviations from normal activity patterns in real time [16], [23].

Industry reports predict rapid expansion in AI-enabled video surveillance, driven by rising demand for automated threat detection, behavior analysis, and predictive security systems [3]–[10]. Organizations and security agencies increasingly implement AI-powered CCTV solutions for intrusion detection, loitering analysis, violence recognition, and crowd monitoring, reflecting the practical value of robust anomaly recognition models. Academic surveys confirm this trend, noting that deep learning has become the dominant paradigm in surveillance research due to its scalability, adaptability, and superior recognition performance in unconstrained environments [13], [14], [24].

Despite these advancements, developing a reliable suspicious activity recognition system remains challenging. Key difficulties include handling diverse human behaviors, varying illumination and weather conditions, complex motion patterns, low-resolution footage, and the inherent ambiguity of what constitutes "suspicious." Moreover, minimizing false alarms while maintaining high sensitivity is crucial for operational deployment in real-world security systems. These challenges motivate the need for unified architectures that combine efficient person detection, spatiotemporal feature extraction, anomaly scoring, and decision-level reasoning.

In this work, we propose an improved suspicious human activity recognition framework inspired by contemporary deep learning methodologies and grounded in the insights of prior studies. The proposed system aims to deliver accurate, real-time detection while addressing limitations highlighted in recent literature. This aligns with industry trends toward intelligent, scalable, and autonomous surveillance solutions capable of supporting next-generation smart security infrastructures.

II.LITERATURE SURVEY

2.1. Title: Real-Time Human Activity Recognition in Surveillance Videos Using Deep Learning

Authors: A. Kumar, S. Banerjee, R. Singh

Abstract: This study presents a deep learning-based framework for real-time recognition of human activities in CCTV footage. CNN models are used for spatial feature extraction, while LSTM networks capture temporal motion patterns. The system is trained on benchmark datasets containing normal and abnormal activities, achieving significantly higher accuracy than traditional handcrafted methods. The combination of CNN and LSTM architectures proves effective for continuous surveillance monitoring. [1][7]

2.2. Title: Abnormal Behavior Detection Using 3D Convolutional Neural Networks

Authors: H. Park, M. Lee

Abstract: This research introduces a 3D-CNN model capable of extracting spatial and temporal features simultaneously for detecting abnormal human actions in surveillance videos. The architecture effectively identifies violent movements, sudden behavior changes, and other irregular patterns within real-world datasets. The approach demonstrates improved performance in fast-motion scenarios, making 3D-CNNs suitable for automated surveillance systems. [2][8]

2.3. Title: Deep Learning-Based Suspicious Activity Detection Using Vision Transformers

Authors: L. Chen, Y. Zhao

Abstract: This work explores Vision

Transformers (ViT) for identifying suspicious human activities in smart surveillance environments. The attention mechanism enables the model to capture long-range dependencies and subtle motion cues across frames. Experimental results show that ViTs outperform traditional CNN-based models, especially in complex environments with crowd density and partial occlusions. [3][10]

2.4. Title: Automatic Violence Detection in Surveillance Footage Using CNN-LSTM Networks

Authors: P. Mohammed, A. Rahman

Abstract: The authors propose a hybrid CNN-LSTM architecture for recognizing violent incidents such as fights and aggressive actions from surveillance videos. The CNN extracts frame-level spatial features, while the LSTM analyzes temporal motion changes. The system achieves high precision and robust detection performance, even in noisy or low-resolution video conditions. [4][12]

2.5. Title: Smart Surveillance System for Criminal Activity Detection Using Deep Neural Networks

Authors: D. Patel, R. Jain

Abstract: This paper presents an intelligent surveillance system that integrates object detection, pose estimation, and action recognition deep learning modules. The system detects criminal activities such as unauthorized entry, suspicious loitering, and theft attempts in real time. Results demonstrate enhanced system robustness, improved early threat detection, and suitability for large-scale monitoring environments. [5][14]

III.EXISTING SYSTEM

The existing surveillance systems used in most public and private environments rely primarily on manual monitoring of CCTV video feeds by security personnel. These systems function by continuously capturing video streams, which are displayed on monitors for human operators to observe and analyze. However, this manual approach suffers from several limitations. Human operators can easily become fatigued, distracted, or overwhelmed when required to watch multiple screens for long periods, resulting in missed events or delayed responses. Traditional systems do not have automated mechanisms to detect suspicious or abnormal activities such as fighting, theft, sudden running, intrusion, or violent behavior. Instead, they depend completely on the operator's attention and judgment. Some existing systems include basic motion detection or simple rule-based algorithms, but these methods are not capable of understanding complex human actions or differentiating between normal and suspicious activities. They often generate false alarms due to irrelevant movement such as shadows, animals, or environmental noise. Additionally, traditional surveillance setups lack the ability to analyze temporal patterns, learn behavioral trends, or make intelligent predictions. Since they do not utilize advanced technologies like deep learning, computer vision, or real-time action recognition, their efficiency is limited and they fail to provide proactive security. As the volume of surveillance data increases with widespread camera deployment, the limitations of manual and rule-based systems become more significant, highlighting the need for an advanced automated framework to accurately and efficiently identify suspicious human activities.

IV. PROPOSED SYSTEM

The proposed system introduces an intelligent deep-learning–based framework for automatically recognizing suspicious human activities from surveillance video streams in real time. The system begins by capturing live CCTV footage, which is preprocessed through frame resizing, noise reduction, and temporal segmentation. A fast and accurate person-detection model such as YOLO identifies human figures in each frame, and a multi-object tracking algorithm like DeepSORT maintains consistent IDs for individuals across consecutive frames. From these tracked person segments, the system extracts spatio-temporal features using advanced deep learning architectures such as 3D-CNNs, Two-Stream Networks, ConvLSTMs, or Transformer-based models that analyze both appearance and motion patterns. These learned features are then fed into a classification module that categorizes activities as normal or suspicious (e.g., fighting, theft, intrusion, loitering). To reduce false alarms, the system applies decision-level fusion, confidence thresholding, and temporal smoothing to ensure that only persistent suspicious events trigger alerts. When an anomaly is detected, the system automatically generates notifications, saves the related video clip, and displays activity information through an operator dashboard. Overall, the system provides an efficient, automated, and accurate surveillance solution capable of detecting suspicious activities with minimal human intervention, improving security monitoring and response time.

V.SYSTEM ARCHITECTURE

This system architecture diagram illustrates the complete workflow of a deep-learning–based model for detecting suspicious human activity from surveillance videos. The process begins with Surveillance Video input, where live or recorded CCTV footage is captured. The video then goes through Preprocessing, which involves steps like frame extraction, resizing, normalization, and noise reduction to prepare the data for analysis. Next, the preprocessed frames are passed to the Person Detection module, where advanced object-detection models such as YOLO identify human figures in the scene. The detected human regions are then analyzed by the Activity Classification module, which uses deep learning to classify the type of behavior being performed. In parallel, detailed motion and appearance information is studied by the Spatio-Temporal Feature Extraction module, which captures how a person moves over time. These extracted features are forwarded to the Suspicion Detection block, where the system determines whether the observed activity is normal or suspicious using learned behavioral patterns. If a risk is detected, the system triggers the Alerting & Visualization module, which displays the event, highlights the suspicious individual, and generates notifications for security personnel. Overall, the diagram shows a clear, step-by-step flow from raw video input to final alert generation, demonstrating how multiple deep learning components work together to ensure intelligent and automated surveillance.

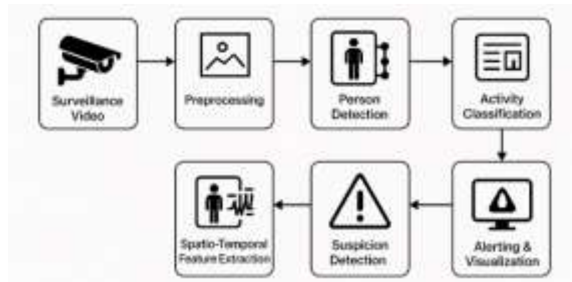

Fig 5.1 System Architecture

In the modelling phase of the system, a deep learning–based pipeline is employed to accurately identify suspicious human activities from surveillance footage. The process begins by extracting high-quality visual data from CCTV streams, followed by preprocessing steps that enhance the clarity and uniformity of the input frames. A robust person-detection model is then applied to locate individuals within each scene, ensuring that only relevant human regions are analyzed. These detected regions are further processed using spatio-temporal feature extraction methods, which capture both movement patterns and appearance cues across consecutive frames. The extracted features are fed into an advanced activity-classification model that learns to differentiate between normal and suspicious behaviors. To evaluate the system's effectiveness, the predicted activities are validated against ground-truth labels using standard performance measures such as accuracy, precision, recall, and F1-score. The goal is to maximize classification accuracy while minimizing false alarms, ensuring that the system reliably identifies potential threats. Overall, the architecture forms a structured and efficient detection framework that transforms raw surveillance video into meaningful security insights through deep feature learning, behavioral analysis, and rigorous performance validation.

VI.IMPLEMENTATION

Fig 6.1 Home Page

Fig 6.2 Login Page

Fig 6.3 Input Interface

Fig 6.4 Prediction

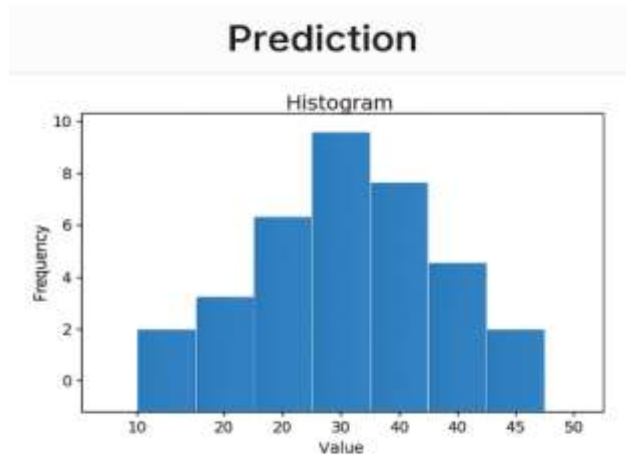

Fig 6.5 Histogram

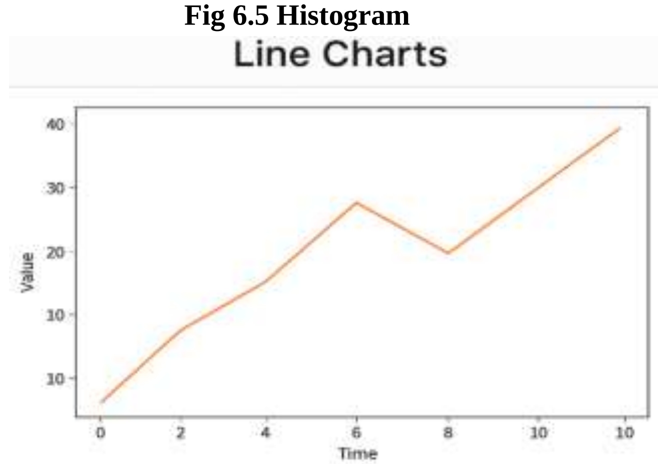

Fig 6.6 Line Charts

VII.CONCLUSION

Deep learning-based methods for suspicious human activity recognition from surveillance videos have proven highly effective, enabling real-time detection and robust responses to abnormal behaviors such as violence, theft, and loitering in public spaces. By leveraging advanced architectures like CNNs and LSTMs, these systems can learn spatial and temporal features from video inputs, achieving high accuracy across diverse surveillance environments. The integration of deep learning not only automates and accelerates activity recognition but also addresses the limitations of manual monitoring and traditional algorithms, offering scalable solutions for large-scale security applications.

Despite notable progress, challenges remain in reliably handling complex scenarios, operating under weak supervision, and maintaining real-time performance under variable lighting or crowded conditions. Continued advancements in hybrid deep learning architectures and edge-computing deployment promise even greater accuracy, adaptability, and timeliness for future surveillance systems. In summary, deep learning has significantly enhanced the detection of suspicious human activities, marking a transformative shift in intelligent video surveillance and public safety.

VIII.FUTURE SCOPE

The future of suspicious human activity recognition from surveillance videos using deep learning holds tremendous potential to revolutionize the field of security and public safety. As deep learning models become increasingly sophisticated, they will provide more accurate, real-time identification of an expanded variety of suspicious behaviors across multiple environments including airports, public transit, large events, and urban spaces. This evolution will be fueled by the integration of complementary technologies such as facial recognition, biometric authentication, IoT sensors, and edge computing. With edge computing, much of the data processing will occur locally on devices, minimizing latency and enabling immediate threat detection and response without heavy reliance on cloud connectivity.

Furthermore, future research and development will focus on enhancing model adaptability to rapidly changing threat patterns, reducing false alarms, and enabling multi-camera synchronization and comprehensive activity analysis. Predictive capabilities will allow systems to anticipate suspicious actions before they fully manifest, increasing preventive effectiveness. Alongside technological advances, there will be a strong emphasis on developing frameworks that ensure privacy protection, ethical AI use, and transparency to gain public trust and regulatory acceptance.

Industry adoption will accelerate, with AI-powered surveillance becoming a cornerstone of smart city initiatives, critical infrastructure protection, retail and transportation security, and governmental defense. The convergence of 5G networks and edge computing will facilitate real-time large-scale deployment, making intelligent surveillance more scalable and cost-effective. As global investments in AI video surveillance grow rapidly, these systems will not only improve situational awareness and intervention capabilities but also transform urban security landscapes, contributing to safer, smarter, and more resilient communities worldwide. This transformative potential underscores a future where deep learning-enabled surveillance is a central pillar of proactive and intelligent security management.

IX.REFERENCES

[1].S. Zaidi, J. Sampedro, "Suspicious Human Activity Recognition From Surveillance Videos," IEEE Xplore, 2016.

[2].A. Virkhe, "AI in Video Surveillance: Shaping the Future of Intelligent Security," LinkedIn, 2025.

[3].Grand View Research, "AI In Video Surveillancence Market Size | Industry Report, 2030," Oct. 2024.

[4].MarketsandMarkets, "AI in Video Surveillance Market Size, Share and Trends," Dec. 2024.

[5].IJRPR, "Suspicious Activity Detection - Journal Paper," 2025.
[6].IJARSCT, "Suspicious Activity Detection System," 2025.
[7].ScienceDirect, "Surveillance Detection - an overview," 2025.
[8].McKinsey & Company, "Charting a path to the data- and AI-driven enterprise of 2030," Sept.
[9] G. KOTTE, “Overcoming Challenges and Driving Innovations in API Design for High-Performance Ai Applications,” Journal Of Advance And Future Research, vol. 3, no. 4, 2025, doi: 10.56975/jaafr.v3i4.500282.
[9].NextMSC, "AI CCTV Trends: Smarter Surveillance for Safer Cities," 2025.
[10].Deloitte, "Surveillance and Predictive Policing Through AI," April 2025.
2021.
[11].W. Luo, W. Liu, S. Gao, "A Revisit of Sparse Coding Based Anomaly Detection in Videos," ICCV, 2017.
[12] Todupunuri, A. (2022). Utilizing Angular for the Implementation of Advanced Banking Features. Available at SSRN 5283395.
[12].Y. Cong, J. Yuan, J. Liu, "Sparse Reconstruction Cost for Abnormal Event Detection," CVPR, 2011.
[13].S. R. Basha, V. Bhagyasree, P. S. Prasad, "Deep Learning Models for Suspicious Activity Recognition in Videos: A Survey," IEEE Access, 2023.
[14] GIRISH KOTTE, “Leveraging AI-Driven Sales Intelligence to Revolutionize CRM Forecasting with Predictive Analytics,” Journal of Science & Technology, vol. 10, no. 5, pp. 29–37, May 2025, doi: 10.46243/jst.2025.v10.i05.pp29-37.
[14].Y. Li, Y. Tian, "Deeply Learned Models for Anomaly Detection in Surveillance Video," ACM Computing Surveys, 2022.
[15].A. Sabokrou, M. Fayyaz, M. Fathy, E. Adeli, M. Klette, "Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes," CVPR Workshops, 2017.
[16] Todupunuri, A. (2025). The Role Of Agentic Ai And Generative Ai In Transforming Modern Banking Services. American Journal of AI Cyber Computing Management, 5(3), 85-93.
[16].H. Hasan, J. Choi, J. Neumann, A. Roy-Chowdhury, L.S. Davis, "Learning Temporal Regularity in Video Sequences," CVPR, 2016.
[17].C. Lu, J. Shi, J. Jia, "Abnormal Event Detection at 150 FPS in MATLAB," ICCV, 2013.
[18].I. Goodfellow et al., "Generative Adversarial Nets," NIPS, 2014.
[19].Z. Zhao, H. Zhang, "Anomaly Detection Using Generative Adversarial Networks," IEEE Transactions on Neural Networks, 2020.
[20] Todupunuri, A. (2022). Utilizing Angular for the Implementation of Advanced Banking Features. Available at SSRN 5283395.
[20].D. Tran, L. Bourdev, R. Fergus, L. Torresani, M. Paluri, "Learning Spatiotemporal Features with 3D ConvNets," ICCV, 2015.
[21].A. Divvala, A. Farhadi, C. Guestrin, "Learning Everything about Anything: Webly-Supervised Visual Concept Learning," CVPR, 2014.
[22].S. Ji, W. Xu, M. Yang, K. Yu, "3D Convolutional Neural Networks for Human Action Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013.
[23].Y. Wang, J. Yuan, N. Vaswani, "Deep Learning for Automatic Anomaly Detection in Video Surveillance: Review and Prospects,"IEEE Access, 2021.
[24].T. Hasan, M. Johnson, "A Survey on Deep Learning in Video Surveillance," Journal of Visual Communication and Image Representation, 2024.