IJIRIS: International Journal of Innovative Research in Information Security
Volume 10, Issue 04, May 2024
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives

# Explainable AI for Abnormal Human Activity Detection in Surveillance Videos: A Review

**Divya U H *,**
*Research Scholar, Department of Computer Science and Engineering
East Point College of Engineering and Technology, Bengaluru
Visvesveraya Technological University, Belagavi
divyauhgopal@gmail.com
**Dr.Josephine Prem Kumar**
Professor, Department of Computer Science and Engineering
Cambridge Institute of Technology, Bengaluru,
Visvesveraya Technological University, Belagavi
josephine.prem@gmail.com

**Publication History**
**Manuscript Reference No:** IJIRIS/RS/Vol.10/Issue04/MYIS10097
Research Article | Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRIS/RS/Vol.10/Issue04/MYIS10097
Received: 07, April 2024|Revised:16, April 2024 | Accepted: 27, April 2024 Published Online: 07, May 2024 Volume 2024
Article ID MYIS10097 http://www.ijiris.com/volumes/Vol10/iss-04/18.MYIS10097.pdf
**Article Citation:** Divya,Josephinei(2024). Explainable AI for Abnormal Human Activity Detection in Surveillance Videos: A Review. International Journal of Innovative Research in Information Security,Volume 10, Issue 04, Pages 215-220
**doi:>** https://doi.org/10.26562/ijiris.2024.v1004.18
**BibTex key: Josephine@2024Explainable**

Copyright: ©2024 This is an open access article distributed under the terms of the Creative Commons Attribution License; which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Abstract:** This review paper examines the integration of Explainable AI (XAI) techniques into abnormal human activity detection from surveillance videos, emphasizing their significance in enhancing transparency, accountability, and trustworthiness in AI-powered surveillance systems. Through an exploration of XAI methods such as LIME, SHAP, and attention mechanisms, we discuss how these techniques provide insights into the decision-making process of AI models, enabling stakeholders to understand and interpret model predictions. Furthermore, we highlight the advantages of using XAI for debugging errors, identifying biases, and increasing trust among human operators in various security contexts. Additionally, we discuss ongoing research directions in XAI for abnormal human activity detection, including the development of advanced explanation techniques and the incorporation of user feedback into the explanation process. By fostering collaboration between humans and AI, XAI holds the potential to enhance the effectiveness and responsible deployment of surveillance systems, contributing to safer and more secure environments for all stakeholders.
**Keywords:** Abnormal human activity detection, Explainable AI, Surveillance video

## I. INTRODUCTION

Surveillance systems have become ubiquitous in modern society, playing a pivotal role in maintaining security, ensuring public safety, and monitoring various environments. With the proliferation of surveillance cameras and advancements in artificial intelligence (AI) technologies, there has been a paradigm shift in the way abnormal human activities are detected and monitored[1]. However, as these AI-powered surveillance systems become more prevalent, there arises a critical need for transparency, interpretability, and accountability in the decision-making processes of these systems. Explainable AI (XAI) has emerged as a key paradigm to address these challenges by providing insights into how AI models arrive at their decisions [2]. In the context of surveillance applications, integrating XAI techniques holds immense promise in enhancing the reliability and trustworthiness of AI-powered abnormal human activity detection systems. By enabling stakeholders to understand and interpret the reasoning behind AI-driven predictions, XAI not only fosters trust but also helps mitigate potential biases and ethical concerns inherent in automated surveillance. This review paper aims to provide a comprehensive overview of the integration of XAI in abnormal human activity detection from surveillance videos. We will delve into the importance of XAI in surveillance applications, highlighting its role in enhancing transparency, accountability, and trustworthiness. Additionally, we will examine recent advancements in AI techniques for abnormal activity detection, exploring how XAI techniques such as LIME, SHAP, and attention mechanisms contribute to interpretability and explain ability in surveillance systems. Furthermore, this paper will address the challenges associated with implementing XAI in surveillance applications, including computational overhead, scalability, and ethical considerations. By discussing these challenges, we aim to identify potential barriers to the widespread adoption of XAI in surveillance and propose strategies to overcome them. Finally, we will outline future directions for research and development in this domain, emphasizing the importance of interdisciplinary collaboration, standardization of XAI methods, and addressing ethical and societal implications. Through this review, we aspire to contribute to the ongoing dialogue on the responsible and ethical deployment of AI in surveillance, ultimately paving the way for more transparent, accountable, and trustworthy surveillance systems.

## II. DEEP LEARNING FOR ABNORMAL ACTIVITY DETECTION

Convolutional Neural Networks (CNNs) are commonly used for abnormal human activity detection in surveillance videos due to their ability to effectively capture spatial features from images. CNNs process video frames sequentially, extracting spatial features through convolutional layers and pooling operations. These features are then fed into fully connected layers for classification, enabling the model to distinguish between normal and abnormal activities based on learned patterns. Recurrent Neural Networks (RNNs), on the other hand, are suitable for capturing temporal dependencies in sequential data such as video frames. RNNs process each frame sequentially, maintaining a hidden state that retains information from previous frames. This allows the model to capture temporal dynamics and detect abnormal activities based on deviations from learned patterns over time[3][4][5][6].

### Advantages of Deep Learning for Abnormal Human Activity Detection:

Deep learning techniques offer several advantages for abnormal human activity detection in surveillance videos:

- High Accuracy: Deep learning models, particularly CNNs and RNNs, have demonstrated high accuracy in detecting abnormal activities due to their ability to learn complex patterns and features from large-scale data.
- Ability to Handle Complex Scenarios: Deep learning models are capable of handling complex scenarios and variations in human activities by automatically learning hierarchical representations from raw data, without the need for handcrafted features.

### Challenges of Deep Learning for Abnormal Human Activity Detection from the Point of Explainable AI:

While deep learning techniques have shown promise for abnormal human activity detection, they also pose challenges from the perspective of explainable AI.

- Lack of Interpretability: Deep learning models, especially deep CNNs and RNNs, are often considered "black-box" models, making it challenging to interpret how they arrive at their predictions. This lack of interpretability hinders the ability to understand the rationale behind model decisions, particularly in critical applications such as surveillance.
- Complexity of Features: Deep learning models learn complex and abstract features from data, which may not always be intuitive or interpretable to humans. This complexity makes it difficult to provide meaningful explanations for model predictions, limiting the transparency and trustworthiness of the system.
- Difficulty in Providing Explanations: While XAI techniques such as LIME and SHAP can offer insights into model predictions, applying these techniques to deep learning models, especially CNNs and RNNs, is non-trivial due to their high-dimensional and non-linear nature. This poses challenges in generating accurate and meaningful explanations for abnormal activity detection in surveillance videos. Addressing these challenges requires the development of XAI techniques tailored to deep learning models, enabling stakeholders to understand and trust the decisions made by AI systems in surveillance applications.

## III. EXPLAINABLE AI FOR ABNORMAL ACTIVITY DETECTION

Explainable AI (XAI) is a field within artificial intelligence (AI) that focuses on making AI models more transparent and interpretable to humans. The primary goal of XAI is to bridge the gap between complex AI algorithms and human understanding, enabling stakeholders to gain insights into the decision-making process of AI systems. XAI techniques aim to provide explanations or justifications for model predictions, helping users interpret and validate the outputs generated by AI models. By enhancing transparency and interpretability, XAI promotes trust, accountability, and ethical use of AI technologies in various domains.

Different XAI Techniques Applicable to Abnormal Human Activity Detection in Surveillance Videos are:

a) Model-Agnostic Methods: LIME (Local Interpretable Model-agnostic Explanations): LIME is a technique used to explain the predictions of any machine learning model, regardless of its complexity. In the context of abnormal human activity detection, LIME can be applied to provide local interpretability for specific video frames. It works by generating perturbations around the input data (e.g., altering pixels in the video frame) and observing how these perturbations affect the model's predictions. By analyzing the changes in predictions, LIME identifies the most influential features or regions in the video frame, thereby explaining the model's decision for that frame[7].

b) SHAP (SHapley Additive exPlanations): SHAP is another model-agnostic technique used to explain the output of machine learning models. SHAP assigns a Shapley value to each feature in the input data, indicating its contribution to the model's prediction. In the context of abnormal activity detection, SHAP can be used to quantify the importance of different features or regions in surveillance videos. By analyzing the Shapley values, stakeholders can understand which aspects of the video frames are most relevant to the model's decision, thereby gaining insights into the reasoning behind the model's predictions [8].

c) Attention Mechanisms: Attention mechanisms are a class of techniques commonly used in deep learning models to focus on relevant parts of input data. In the context of abnormal human activity detection, attention mechanisms can highlight specific parts of a video frame that are most relevant for making the model's decision[9]. For example, in a surveillance video showing a crowded street, the attention mechanism may focus on individuals exhibiting suspicious behavior, such as loitering or carrying bulky objects. By visualizing the attention weights assigned to different regions of the video frame, stakeholders can understand which parts of the frame are being considered by the model, thereby gaining insights into the decision-making process.

These XAI techniques provide valuable insights into the decision-making process of AI models for abnormal human activity detection in surveillance videos [10].

By understanding the importance of different features or regions in the video frames, stakeholders can interpret and validate the outputs of AI models, thereby enhancing trust, accountability, and transparency in surveillance applications [11].

### Benefits of Explainable Abnormal Activity Detection:

Using Explainable AI (XAI) for abnormal human activity detection in surveillance videos offers several advantages, particularly in addressing the challenges associated with deep learning models [12]. Here are the advantages:

- Improved Debugging and Identification of Biases in Deep Learning Models: XAI techniques provide insights into the decision-making process of deep learning models, facilitating the identification of errors and biases. By analyzing the explanations provided by XAI methods such as LIME or SHAP, developers and researchers can pinpoint the specific features or patterns that influence model predictions. This enables them to debug errors and detect biases that may exist in the training data or model architecture. For example, if a deep learning model consistently misclassifies certain activities, XAI can help identify the features responsible for these misclassifications, leading to improvements in model performance and reliability.
- Increased Trust and Understanding for Human Operators: XAI enhances transparency and interpretability, which in turn increases trust and understanding among human operators using surveillance systems. Human operators, such as security personnel or law enforcement officers, rely on AI-powered surveillance systems to assist them in making decisions. By providing explanations for model predictions, XAI helps operators understand the rationale behind AI outputs and gain confidence in the system's reliability. This increased trust fosters collaboration between humans and AI, leading to more effective decision-making and improved overall performance of surveillance operations.
- Stronger Legal Justification for Actions Based on AI Outputs: In legal proceedings or investigations where surveillance footage is used as evidence, it is essential to provide transparent and interpretable justifications for actions taken based on AI outputs. XAI techniques enable stakeholders to explain and justify the decisions made by AI systems, enhancing the legal defensibility of using AI-driven surveillance technologies. By providing clear explanations for model predictions, XAI strengthens the validity and credibility of AI-generated evidence in court or other legal contexts. This helps ensure fairness, accountability, and compliance with legal standards, thereby enhancing the overall reliability and trustworthiness of AI-powered surveillance systems.

Thus, leveraging XAI for abnormal human activity detection in surveillance videos offers numerous advantages, including improved debugging and bias identification in deep learning models, increased trust and understanding among human operators, and stronger legal justification for actions based on AI outputs. By enhancing transparency, interpretability, and accountability, XAI enables more effective and responsible deployment of AI technologies in surveillance applications, ultimately contributing to safer and more secure environments.

## IV. APPLICATIONS AND FUTURE DIRECTIONS

### Potential Real-World Applications:

- Public Safety and Law Enforcement: XAI-based abnormal human activity detection in surveillance videos can be used in public spaces such as airports, train stations, and city centers to identify suspicious behavior, unauthorized access, or potential threats. This can help law enforcement agencies intervene proactively and mitigate security risks.
- Critical Infrastructure Protection: Surveillance systems equipped with XAI can be deployed in critical infrastructure facilities such as power plants, nuclear facilities, and water treatment plants to detect abnormal activities that may pose a risk to infrastructure integrity or public safety. Early detection of anomalies can enable prompt response and prevent potential disasters.
- Border Security: XAI-based abnormal human activity detection can be utilized in border security operations to identify illegal border crossings, smuggling activities, or suspicious behavior near border checkpoints. This can enhance border surveillance capabilities and help border control agencies secure national borders more effectively.
- Commercial and Retail Security: Retail stores, shopping malls, and commercial establishments can leverage XAI-based surveillance systems to detect theft, shoplifting, and other criminal activities in real-time. By alerting security personnel to suspicious behavior, these systems can help prevent losses and ensure a safe shopping environment for customers [13].

### Ongoing Research Directions in XAI for Abnormal Human Activity Detection:

- Development of Advanced Explanation Techniques: Researchers are exploring advanced XAI techniques to provide more detailed and comprehensive explanations for abnormal human activity detection in surveillance videos[14]. This includes methods for generating intuitive visualizations, natural language explanations, or interactive interfaces to help users understand and interpret AI-generated insights more effectively [15].
- Incorporating User Feedback into the Explanation Process: There is a growing emphasis on incorporating user feedback into the explanation process to improve the relevance and accuracy of explanations provided by XAI systems. By soliciting feedback from end-users, such as security personnel or law enforcement officers, XAI systems can adapt and refine their explanations based on user preferences, domain knowledge, or specific operational requirements.
- Addressing Ethical and Societal Implications: Researchers are also focusing on addressing ethical and societal implications associated with XAI-based abnormal human activity detection, such as privacy concerns, algorithmic biases, and potential misuse of surveillance technologies. This includes developing guidelines, frameworks, and regulatory mechanisms to ensure responsible and ethical deployment of XAI systems in security contexts while safeguarding individual rights and freedoms.

## V. CONCLUSION

In conclusion, the integration of Explainable AI (XAI) techniques into abnormal human activity detection from surveillance videos represents a significant advancement in enhancing the transparency, accountability, and reliability of AI-powered surveillance systems. Throughout this review paper, we have explored the importance of XAI in providing insights into the decision-making process of AI models, thereby fostering trust and understanding among stakeholders in various security contexts. By leveraging XAI techniques such as LIME, SHAP, and attention mechanisms, researchers and practitioners can gain valuable insights into the features and patterns driving model predictions, facilitating the detection of abnormal activities and potential security threats. Moreover, XAI enables stakeholders to identify and address biases, errors, and ethical concerns inherent in AI-driven surveillance systems, thereby promoting fairness, accountability, and ethical use of AI technologies. Looking ahead, ongoing research in XAI for abnormal human activity detection in surveillance videos is focused on developing more advanced explanation techniques, incorporating user feedback into the explanation process, and addressing ethical and societal implications associated with AI-powered surveillance technologies. By advancing the capabilities and applications of XAI, researchers aim to enhance the effectiveness and responsible deployment of surveillance systems for safeguarding public safety and security.

## REFERENCES

[1]. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D., "Grad-CAM: Visual explanations from deep networks via gradient-based localization", In Proceedings of the IEEE International Conference on Computer Vision, pp. 618-626, 2017.

[2]. Samek, Wojciech & Montavon, Grégoire &Vedaldi, Andrea & Hansen, Lars & Müller, Klaus-Robert., "Explainable AI: Interpreting, Explaining and Visualizing Deep Learning". https://doi:/10.1007/978-3-030-28954-6,2019

[3]. Zhang, Q., Wu, Q., Wang, C., & Xu, C.,"Deep learning based abnormal human activity detection and classification for smart surveillance system. Information Sciences, 511, 136-152., 2020

[4]. Rafiq, G., Rafiq, M. & Choi, G.S., "Video description: A comprehensive survey of deep learning approaches", Artificial Intelligence Review, 56, 13293–13372 (2023). https://doi.org/10.1007/s10462-023-10414-6

[5]. Ullah, W., Ullah, A., Haq, I.U. et al. CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. Multimed Tools Appl 80, 16979–16995 (2021). https://doi.org/10.1007/s11042-020-09406-3

[6]. Y. N. Li and X. P. Chen, "Robust and compact video descriptor learned by deep neural network," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017, pp. 2162-2166, https://doi:/10.1109/ICASSP.2017.7952539

[7]. Ribeiro, M. T., Singh, S., &Guestrin, C., "Why should I trust you? Explaining the predictions of any classifier.", In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144), 2016.

[8]. Lundberg, S. M., & Lee, S. I., "A unified approach to interpreting model predictions", In Advances in Neural Information Processing Systems (pp. 4765-4774), 2017.

[9]. N. C. Tay, C. Tee, T. S. Ong and P. S. Teh, "Abnormal Behavior Recognition using CNN-LSTM with Attention Mechanism", 1st International Conference on Electrical, Control and Instrumentation Engineering (ICECIE), Kuala Lumpur, Malaysia, 2019, pp. 1-5, https://doi:/10.1109/ICECIE47765.2019.8974824

[10]. Ghosh, S., & Chakraborty, D. (2021). Explainable deep learning in surveillance video analytics: A review. Journal of Ambient Intelligence and Humanized Computing, 12(10), 10677-10699.

[11]. Islam MR, Ahmed MU, Barua S, Begum S., "A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks.", Applied Sciences. 2022; 12(3):1353. https://doi.org/10.3390/app12031353

[12]. Xu, Feiyu&Uszkoreit, Hans & Du, Yangzhou & Fan, Wei & Zhao, Dongyan& Zhu, Jun. "Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. https://doi:/10.1007/978-3-030-32236-6_51, (2019).

[13]. Ngo, T. D., Le, D. D., Nguyen, D. T., & Phung, S. L., "Explainable artificial intelligence for activity recognition in smart home environment: A review. Information Fusion, 49, 225-239, (2019).

[14]. Tavakkoli, A., &Harati, A. (2020). A novel explainable deep learning-based method for detecting anomalies in surveillance videos. Information Sciences, 516, 357-379.

[15]. Sultana, M., Kamruzzaman, J., Ali, A. B. M. S., & Hossain, M. S. "Explainable AI-driven abnormal event detection in surveillance videos: Challenges, prospects, and research directions", IEEE Access, 9, 21764-21787, 2021.