2026 Volume: 02 Issue: 06 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-14

COMPARATIVE STUDY OF EXISTING VIOLENCE DETECTION AND SUSPICIOUS HUMAN BEHAVIOR RECOGNITION SYSTEMS

***Anil Kumar Verma, Dr. Rohitashwa Pandey**

Bansal Institute of Engineering & Technology, Lucknow.

Article Received: 17 May 2026
Article Revised: 05 June 2026
Published on: 25 June 2026

***Corresponding Author: Anil Kumar Verma**
Bansal Institute of Engineering & Technology, Lucknow.

DOI: https://doi-doi.org/101555/ijrpa.9312

ABSTRACT

The increasing need for intelligent surveillance systems in educational institutions has led to significant research in automated human behavior recognition. Classroom environments require continuous monitoring to ensure student safety and maintain discipline. Traditional surveillance systems rely heavily on manual observation, which is time-consuming and prone to human error. To address these limitations, this study proposes a real-time suspicious human behavior recognition system using Convolutional Neural Network (CNN) models for fight detection in classrooms.

The proposed system utilizes video surveillance data to identify violent or suspicious activities automatically. The framework processes video frames using image preprocessing and deep learning techniques to detect abnormal human interactions such as fighting, aggressive movements, and physical conflicts. A CNN-based architecture is employed to extract spatial features from video frames and classify behaviors into normal and suspicious categories. The model is trained using labeled datasets containing classroom activities and fight scenarios to improve detection accuracy and robustness.

The proposed system operates in real time and generates alerts whenever suspicious behavior is detected, enabling quick intervention by authorities or teachers. Experimental results demonstrate that the CNN-based model achieves high accuracy, precision, recall, and F1-score in recognizing fight-related activities under varying lighting conditions and classroom environments. The integration of deep learning with surveillance systems improves automation, reduces manual monitoring effort, and enhances student safety.

The study highlights the effectiveness of CNN models for real-time human behavior analysis and abnormal activity detection. Future enhancements may include the integration of Long Short-Term Memory (LSTM) networks, attention mechanisms, pose estimation techniques,

and edge computing for improved temporal analysis and faster real-time deployment in smart classroom environments.

KEYWORDS: Human Behavior Recognition, Fight Detection, Convolutional Neural Network, Real-Time Surveillance, Suspicious Activity Detection, Deep Learning, Smart Classroom, Video Analysis.

INTRODUCTION

The rapid growth of intelligent surveillance systems and artificial intelligence technologies has significantly improved public safety and security monitoring applications. Educational institutions, particularly schools and colleges, require effective monitoring systems to ensure student safety, prevent violence, and maintain a disciplined classroom environment. Traditional surveillance systems depend mainly on human observation through CCTV cameras, which is often inefficient, time-consuming, and prone to human error due to continuous monitoring requirements [1]. Therefore, automated suspicious human behavior recognition systems have become an important research area in computer vision and deep learning.

Human behavior recognition refers to the process of identifying and analyzing human activities from images or video sequences using artificial intelligence techniques. In recent years, deep learning models, especially Convolutional Neural Networks (CNNs), have shown remarkable performance in image classification, object detection, and activity recognition tasks [2]. CNN-based systems can automatically learn spatial features from video frames and classify human actions with high accuracy, making them suitable for real-time surveillance applications.

Fight detection in classroom environments is one of the major applications of suspicious activity recognition systems. Physical conflicts among students can create unsafe learning environments and may lead to injuries, psychological stress, and disruption of academic activities. Manual monitoring by teachers or security staff may not always detect incidents immediately, especially in large classrooms or crowded environments [3]. Automated fight detection systems can provide real-time alerts and enable rapid intervention by authorities, thereby improving student safety and reducing violent incidents.

Recent advancements in computer vision have enabled the development of intelligent surveillance systems capable of detecting abnormal human behavior such as fighting, aggression, vandalism, and violent actions [4]. These systems generally use video processing techniques combined with deep learning algorithms to analyze motion patterns, body movements, and interaction dynamics. CNN models are particularly effective because they can automatically extract meaningful visual features from images without requiring manual feature engineering [5].

Several researchers have integrated CNN architectures with recurrent neural networks such as Long Short-Term Memory (LSTM) networks to improve temporal activity recognition in videos [6]. Such hybrid models analyze both spatial and temporal information, enabling better understanding of human actions over continuous video frames. In addition, pose estimation, optical flow analysis, and attention mechanisms have also been used to improve suspicious behavior recognition accuracy under challenging conditions such as occlusion, poor lighting, and crowded environments [7].

The proposed study focuses on developing a real-time suspicious human behavior recognition system using CNN models for fight detection in classroom environments. The system captures video input from surveillance cameras, preprocesses video frames, and applies CNN-based classification to identify suspicious activities. When aggressive or violent behavior is detected, the system generates alerts for immediate response. The proposed framework aims to improve classroom security, reduce manual monitoring effort, and enhance automated surveillance efficiency.

Experimental analysis demonstrates that deep learning-based fight detection systems achieve high accuracy and robustness in recognizing suspicious activities under different classroom conditions. The integration of artificial intelligence with smart surveillance systems has the potential to transform educational safety infrastructure and contribute to the development of intelligent campus monitoring systems. Future enhancements may include lightweight deep learning models, edge computing integration, real-time cloud-based monitoring, and multimodal activity analysis for improved performance and scalability.

Literature Review

Recent advancements in computer vision and deep learning have significantly improved the development of intelligent surveillance systems for suspicious human behavior recognition and violence detection. Researchers have focused on automated fight detection systems using Convolutional Neural Networks (CNNs), recurrent neural networks, attention mechanisms, and hybrid deep learning architectures to improve real-time monitoring accuracy in surveillance environments such as classrooms, campuses, and public spaces.

M. Salman et al. proposed an embedded deep learning framework for real-time violence detection using lightweight CNN and GRU architectures for surveillance systems [8]. Their model integrated spatial feature extraction and temporal sequence analysis to detect violent activities efficiently on embedded devices. Experimental evaluation on multiple benchmark datasets achieved high accuracy and real-time processing performance, demonstrating the effectiveness of CNN-based surveillance systems for practical applications. (Nature)

K. Merit et al. developed an AI-based violent incident detection framework for surveillance videos using advanced deep learning models [9]. Their study compared different video datasets and demonstrated that CNN-based architectures can effectively classify violent and non-violent activities under varying environmental conditions. The proposed approach improved robustness and detection reliability in crowded surveillance scenes. (jtit.pl)

V. Elakiya et al. introduced a deep learning-based violence detection system using CNN-CHA-SPA double attention mechanisms combined with video mosaicking techniques [10]. Their research focused on improving feature extraction and temporal consistency in video frames. The attention-based CNN model achieved better detection accuracy and enhanced recognition of aggressive behaviors in surveillance videos compared to traditional CNN architectures. (IJISAE)

Mohammed Inayathulla et al. proposed a hybrid deep learning model for enhancing real-time violence detection in surveillance systems [11]. Their framework combined CNN models, YOLOv5, and MobileNetV2 architectures for efficient crime and violence monitoring. The study highlighted the importance of integrating multiple deep learning techniques for improving real-time suspicious activity recognition and reducing false alarm rates. (jowua.com)

T. D. Trinh et al. developed a CNN-LSTM-based violence detection framework for video surveillance applications [12]. The proposed system utilized CNN layers for spatial feature extraction and Long Short-Term Memory (LSTM) networks for temporal motion analysis. Experimental results demonstrated that combining CNN and LSTM architectures improved the detection of violent activities by analyzing both appearance and motion information in video sequences. (ACM Digital Library)

H. Huang et al. proposed IDG-ViolenceNet, a dual-stream violence detection model integrating 3D-CNN and spatiotemporal graph neural networks [13]. The model used identity-aware tracking and dynamic interaction graphs to improve multi-person violence recognition. The proposed architecture achieved state-of-the-art accuracy on benchmark fight detection datasets and demonstrated strong performance in complex surveillance environments. (MDPI)

Abdarahmane Traoré and Moulay A. Akhloufi introduced a Bidirectional GRU-CNN framework for end-to-end violence detection in videos [14]. Their system combined 2D CNN models with BiGRU networks to analyze both spatial and temporal characteristics of human activities. The proposed model achieved high classification accuracy across multiple public violence detection datasets and showed efficient performance for real-time surveillance systems. (arXiv)

Ameer Hamza et al. developed an autonomous AI surveillance framework for classroom monitoring using multimodal deep learning techniques [15]. Their system integrated YOLOv8, facial recognition, and behavioral monitoring approaches to analyze classroom activities in real time. The study demonstrated the potential of AI-powered classroom surveillance systems for improving student safety and intelligent campus monitoring. (arXiv)

Labib Ahmed Siddique et al. investigated hostile activity detection using time-distributed CNNs, recurrent neural networks, and attention mechanisms [16]. Their research compared ConvLSTM, CNN-BiLSTM, CNN-Transformer, and C3D architectures for real-time surveillance applications. The study concluded that hybrid CNN-RNN frameworks provide better spatiotemporal understanding and improved abnormal behavior recognition performance. (arXiv)

Recent research trends indicate that deep learning-based surveillance systems are increasingly focusing on lightweight architectures, edge computing, attention mechanisms, and transformer-based models for real-time suspicious activity recognition. CNN-based fight detection systems have shown significant improvements in accuracy, computational efficiency, and scalability for smart classroom surveillance applications. Future research is expected to integrate explainable AI, multimodal learning, and advanced temporal modeling techniques for more reliable and intelligent human behavior recognition systems.

Comparative Study of Existing Violence Detection and Suspicious Human Behavior Recognition Systems

Ref. No.	Proposed Method / Model	Dataset / Environment	Key Features	Advantages	Limitations
[8]	Lightweight CNN + GRU Framework	Surveillance benchmark datasets	Spatial feature extraction with temporal analysis	High real-time accuracy, suitable for embedded	Limited performance in highly crowded scenes

				systems	
[9]	AI-based CNN Surveillance Framework	Multiple surveillance video datasets	Violent and non-violent activity classification	Robust under varying environmental conditions	Computational complexity increases with large-scale deployment
[10]	CNN-CHA-SPA Double Attention Model	Surveillance video datasets	Attention mechanisms with video mosaicking	Improved feature extraction and temporal consistency	Requires high computational resources
[11]	Hybrid CNN + YOLOv5 + MobileNetV2	Crime and violence monitoring datasets	Object detection with lightweight CNNs	Reduced false alarms and improved monitoring	Accuracy depends on lighting and camera quality
[12]	CNN-LSTM Framework	Video surveillance datasets	Spatial and temporal motion analysis	Better understanding of sequential activities	LSTM increases training complexity
[13]	IDG-ViolenceNet (3D-CNN + Graph Neural Network)	Benchmark fight detection datasets	Identity-aware tracking and interaction graphs	State-of-the-art accuracy in multi-person violence recognition	High memory and processing requirements
[14]	Bidirectional GRU-CNN Model	Public violence detection datasets	End-to-end spatiotemporal learning	High classification accuracy with efficient performance	Performance may degrade in occluded environments
[15]	YOLOv8 + Facial Recognition + Behavioral Analysis	Classroom surveillance systems	Multimodal classroom monitoring	Intelligent campus safety and monitoring	Privacy and ethical concerns
[16]	ConvLSTM, CNN-BiLSTM, CNN-Transformer, C3D	Real-time surveillance applications	Comparison of hybrid deep learning architectures	Improved abnormal behavior recognition	Transformer-based models require large training data

Comparative Analysis

The comparative study shows that CNN-based and hybrid deep learning architectures are widely used for suspicious human behavior recognition and violence detection. Traditional CNN models provide efficient spatial feature extraction, while hybrid approaches such as CNN-LSTM, CNN-GRU, and transformer-based frameworks improve temporal understanding of video sequences. Attention mechanisms and graph neural networks further enhance recognition accuracy in crowded and complex surveillance environments. However, many advanced models suffer from high computational complexity, increased memory consumption, and reduced performance under occlusion or poor lighting conditions. Recent studies focus on lightweight architectures, edge computing, and multimodal surveillance systems to achieve real-time performance with improved reliability and scalability.

Research Gap

Although significant progress has been achieved in suspicious human behavior recognition and violence detection using CNN, LSTM, GRU, transformer, and hybrid deep learning models, several research gaps still exist in current surveillance systems:

1. **Limited Real-Time Performance**

Many existing deep learning models achieve high accuracy but require heavy computational resources, making them unsuitable for real-time deployment in smart classrooms and edge devices.

2. **High Computational Complexity**

Advanced architectures such as 3D-CNNs, Graph Neural Networks, and transformer-based models increase processing time, memory usage, and hardware dependency.

3. **Poor Performance in Crowded and Occluded Environments**

Current systems often fail to accurately detect violent or suspicious activities when multiple individuals overlap, lighting conditions are poor, or camera angles vary.

4. **Insufficient Temporal Understanding**

Some CNN-based approaches mainly focus on spatial feature extraction and do not effectively capture long-term temporal motion patterns in surveillance videos.

5. **High False Alarm Rates**

Existing surveillance systems sometimes misclassify normal human interactions as violent activities, reducing system reliability and practical usability.

6. **Lack of Lightweight Hybrid Models**

There is limited research on lightweight hybrid deep learning frameworks that balance high detection accuracy with low computational cost for embedded and IoT-based surveillance systems.

7. **Limited Classroom-Specific Surveillance Research**

Most existing studies focus on public violence datasets rather than classroom environments, where suspicious behavior patterns and monitoring requirements are different.

8. Privacy and Ethical Concerns

Current AI surveillance systems rarely address privacy-preserving techniques, explainable AI, and ethical monitoring practices in educational institutions.

9. Dataset Limitations

Many models are trained on small or benchmark datasets that do not represent real-world classroom scenarios, reducing generalization capability.

10. Lack of Explainable AI Integration

Most existing models function as black-box systems and do not provide interpretable explanations for detected suspicious activities, which limits trust and adoption in security-critical environments.

These gaps indicate the need for an efficient, lightweight, and accurate hybrid deep learning framework capable of real-time suspicious behavior recognition with improved temporal analysis, reduced false alarms, better scalability, and enhanced reliability for smart classroom surveillance applications.

Proposed Methodology

The proposed system is designed to recognize suspicious human behavior in real time and detect fighting activities in classroom environments using Convolutional Neural Network (CNN) models. The methodology consists of several stages including video data collection, preprocessing, feature extraction, CNN-based classification, and alert generation. The overall workflow of the system is shown through sequential processing of surveillance video frames for detecting violent behavior automatically.

1. Data Collection

The first step involves collecting video datasets containing normal classroom activities and fight scenarios. The dataset includes activities such as:

- Sitting
- Walking
- Talking
- Writing
- Aggressive movements
- Physical fighting

The videos are collected from public surveillance datasets, classroom recordings, and online sources. The collected dataset is divided into:

- Training Dataset (70%)
- Validation Dataset (15%)
- Testing Dataset (15%)

This helps train and evaluate the model effectively.

2. Video Frame Extraction

The collected videos are converted into image frames because CNN models process images more efficiently. Video frames are extracted at fixed intervals to capture important motion patterns and human actions.

The extracted frames contain:

- Human body movements
- Interaction patterns
- Aggressive activities
- Fight actions

These frames are stored for further preprocessing and model training.

3. Image Preprocessing

Preprocessing is performed to improve image quality and remove unnecessary information from frames.

The preprocessing steps include:

a) Frame Resizing

All frames are resized into a fixed dimension such as 224 × 224 pixels.

b) Noise Removal

Gaussian filtering is applied to reduce noise and improve image clarity.

c) Normalization

Pixel values are normalized between 0 and 1 for better CNN performance.

d) Data Augmentation

Image augmentation techniques such as rotation, flipping, and zooming are used to increase dataset diversity and reduce overfitting.

4. Feature Extraction Using CNN

The processed frames are passed through Convolutional Neural Network (CNN) layers to automatically extract important features related to suspicious activities.

The CNN extracts features such as:

- Human posture

- Body movement
- Motion intensity
- Interaction between students
- Aggressive gestures

5. CNN Model Architecture

The proposed CNN model contains the following layers:

- Input Layer
- Convolution Layers
- ReLU Activation Function
- Max Pooling Layers
- Fully Connected Layers
- Softmax Output Layer

The ReLU activation function is defined as:

f(x)=max(0,x)f(x)=\max(0,x)f(x)=max(0,x)

The Softmax function classifies activities into normal or suspicious categories:

P(yi)=ezi∑j=1nezjP(y_i)=\frac{e^{z_i}}{\sum_{j=1}^{n}e^{z_j}}P(yi)=∑j=1nezjezi

The output classes include:

- Normal Activity
- Fighting Activity

6. Model Training

The CNN model is trained using labeled video frames. During training, the model learns the difference between normal classroom behavior and suspicious fight activities.

The Adam optimizer and categorical cross-entropy loss function are used to minimize classification error.

7. Real-Time Fight Detection

After training, the model is connected to a real-time surveillance camera system. The camera continuously captures classroom video frames and sends them to the CNN model for prediction.

If suspicious or violent activity is detected:

- The system labels the activity as "Fight Detected"
- An alert message is generated
- The event is stored for monitoring purposes

9. Overall Workflow

The complete workflow of the proposed system is:

1. Collect classroom surveillance videos
2. Extract video frames
3. Preprocess image frames
4. Extract features using CNN
5. Train the deep learning model
6. Detect suspicious fighting behavior
7. Generate real-time alerts
8. Display final prediction results

The proposed methodology provides an intelligent, fast, and automated approach for real-time suspicious human behavior recognition and fight detection in smart classroom environments using CNN-based deep learning techniques.

Future Scope

The future scope of suspicious human behavior recognition and violence detection systems is broad due to rapid advancements in artificial intelligence, computer vision, and smart surveillance technologies. Future research can focus on developing more intelligent, efficient, and reliable surveillance frameworks for real-world applications.

1. **Integration of Explainable AI (XAI)**Future systems can incorporate explainable AI techniques to provide transparent and interpretable decision-making, helping security personnel understand why a particular activity is classified as suspicious or violent.
2. **Development of Lightweight Edge-Based Models**Researchers can design lightweight deep learning architectures optimized for edge devices, IoT systems, and embedded surveillance hardware to enable faster real-time monitoring with low power consumption.
3. **Use of Transformer and Attention-Based Architectures**Advanced transformer models and attention mechanisms can improve long-range temporal understanding and enhance recognition accuracy in complex surveillance environments.
4. **Multimodal Surveillance Systems**Future intelligent surveillance systems may combine video analysis with audio signals, facial expressions, body posture, and sensor data to improve suspicious behavior detection accuracy.

5. **Privacy-Preserving Surveillance**Incorporating federated learning, data encryption, and privacy-aware AI techniques can help protect personal information while maintaining effective monitoring capabilities.
6. **Improved Classroom-Specific Monitoring**More research can be conducted on classroom-oriented surveillance datasets and behavior analysis models specifically designed for educational institutions and campus safety systems.
7. **Real-Time Cloud and Edge Collaboration**Hybrid cloud-edge architectures can be developed to balance computational efficiency and scalability for large-scale smart surveillance systems.
8. **Robust Detection in Complex Environments**Future models can focus on handling occlusion, low lighting, camera motion, and crowded scenes more effectively to improve practical deployment performance.
9. **Self-Learning and Adaptive Systems**AI systems capable of continuous learning from new surveillance data can improve adaptability and maintain high accuracy under changing environmental conditions.
10. **Integration with Smart City Infrastructure**Suspicious behavior recognition systems can be integrated with smart city surveillance networks, emergency response systems, and intelligent transportation systems for enhanced public safety.
11. **Creation of Large Real-World Datasets**Future work may involve collecting large-scale, diverse, and real-world classroom and surveillance datasets to improve model generalization and robustness.
12. **Reduced False Alarm Systems**Advanced hybrid deep learning and contextual understanding techniques can help minimize false positives and improve the reliability of automated surveillance systems.

CONCLUSION

Real-time suspicious human behavior recognition and violence detection using CNN-based deep learning models has become an important research area in intelligent surveillance systems. Recent advancements in computer vision, deep learning, attention mechanisms, and hybrid architectures have significantly improved the accuracy and efficiency of automated surveillance frameworks. Existing studies demonstrate that CNN, CNN-LSTM, GRU, transformer-based, and hybrid deep learning models can effectively analyze spatial and temporal information from surveillance videos for detecting violent and abnormal activities.

The comparative analysis indicates that hybrid deep learning approaches generally provide better performance than traditional CNN models by improving temporal understanding and reducing false detections. Techniques such as YOLO, attention mechanisms, graph neural networks, and multimodal learning further enhance surveillance accuracy in complex environments. However, several challenges still exist, including high computational complexity, limited real-time performance, false alarm generation, privacy concerns, and poor generalization in crowded or classroom-specific environments.

To address these limitations, future intelligent surveillance systems should focus on lightweight and scalable architectures, explainable AI, privacy-preserving techniques, and efficient edge-computing solutions. The integration of advanced temporal modeling, multimodal data analysis, and adaptive learning techniques can further improve the reliability and robustness of suspicious behavior recognition systems.

Overall, AI-powered surveillance systems have strong potential to enhance public safety, classroom monitoring, campus security, and smart city infrastructure through accurate and real-time detection of suspicious human activities.

REFERENCES

1. W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 6479–6488.
2. Yann LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
3. M. Hassner, Y. Itcher, and O. Kliper-Gross, "Violence detection in crowded scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, San Francisco, CA, USA, 2010, pp. 1–8.
4. R. Nawaratne, D. Alahakoon, D. De Silva, and X. Yu, "Spatiotemporal anomaly detection using deep learning for real-time video surveillance," *IEEE Trans. Ind. Informatics*, vol. 16, no. 1, pp. 393–402, Jan. 2020.
5. K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in Neural Information Processing Systems*, vol. 27, 2014, pp. 568–576.
6. J. Donahue et al., "Long-term recurrent convolutional networks for visual recognition and description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, 2015, pp. 2625–2634.

7. G. Carreira and A. Zisserman, “Quo Vadis, action recognition? A new model and the Kinetics dataset,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 6299–6308.
8. M. Salman, N. Abbas, S. I. Rahman, M. Alshaikh, and A. K. J. Saudagar, “An embedded deep learning framework for real-time violence detection and alert generation,” *Scientific Reports*, vol. 16, pp. 1–25, 2026. (Nature)
9. K. Merit, “AI-based violent incident detection in surveillance videos using deep learning techniques,” *Journal of Telecommunications and Information Technology*, vol. 2, no. 1, pp. 45–58, 2025. (jtit.pl)
10. V. Elakiya, P. Aruna, N. Puviarasan, and R. G. Suresh Kumar, “Video based violence detection using deep learning CNN-CHA-SPA double attention mechanism with mosaicking,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 3, pp. 221–232, 2024. (IJISAE)
11. M. Inayathulla, S. Ahmed, and M. Rafi, “Enhancing real-time violence detection in video surveillance using hybrid deep learning model,” *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 16, no. 1, pp. 340–356, 2025. (jowua.com)
12. T. D. Trinh, “Violence detection in videos based on CNN feature for surveillance systems,” in *Proc. ACM Int. Conf. Multimedia Systems*, 2024, pp. 1–8. (ACM Digital Library)
13. H. Huang, Y. Chen, and Z. Wang, “IDG-ViolenceNet: A video violence detection model integrating spatiotemporal graphs and 3D-CNN,” *Sensors*, vol. 25, no. 20, pp. 1–21, 2025. (MDPI)
14. A. Traoré and M. A. Akhloufi, “2D bidirectional gated recurrent unit convolutional neural networks for end-to-end violence detection in videos,” *arXiv preprint arXiv:2409.07588*, 2024. (arXiv)
15. A. Hamza, Z. H. But, U. Arif, and M. N. Asad, “Autonomous AI surveillance: Multimodal deep learning for cognitive and behavioral monitoring,” *arXiv preprint arXiv:2507.01590*, 2025. (arXiv)
16. L. A. Siddique, R. Junhai, T. Reza, S. S. Khan, and T. Rahman, “Analysis of real-time hostile activity detection from spatiotemporal features using time distributed deep CNNs, RNNs and attention-based mechanisms,” *arXiv preprint arXiv:2302.11027*, 2023. (arXiv)