International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 5, Sep-Oct-2025, ISSN (Online): 2395-566X

A Comprehensive Overview of Deep Learning Methods for Violence Detection in Surveillance Systems

Sakshi Keshri[1], Nitin Namdev[2]
[1]M. Tech Student, 2Assistant Professor (HOD)
[1,2]Department of Computer Science & Engineering
Jawaharlal Institute of Technology , Borawan, Khargone , Madhya Pradesh, India

Abstract- This paper presents a comprehensive review of deep learning techniques designed to enhance violence detection in surveillance systems. With the rapid advancement of surveillance technologies, the accurate identification of violent activities has become crucial for ensuring public safety. Conventional approaches often fail to cope with the complexity of video data, which inherently involves both spatial and temporal dynamics. To overcome these limitations, modern deep learning models such as Convolutional Neural Networks (CNNs), InceptionV3, Long Short-Term Memory (LSTM) networks, and hybrid architectures have been widely adopted. These methods excel at capturing spatial representations while simultaneously modeling temporal dependencies, making them well-suited for real-time violence detection tasks. The review further discusses essential preprocessing strategies—including noise reduction, feature extraction, and data augmentation—that significantly improve model robustness. In addition, it outlines persistent challenges such as class imbalance, scalability issues, and high computational costs, which remain key barriers to practical deployment.

Keywords – Deep Learning, Violence Detection, Surveillance Systems, CNNs, InceptionV3, LSTM.

I. INTRODUCTION

In today's increasingly urbanized and interconnected world, ensuring the protection and security of the public has become more critical than ever. With rising concerns over violent incidents in public spaces, the demand for effective and efficient surveillance systems has grown significantly. Traditionally, surveillance video monitoring has been performed by human operators—a process that is both time-consuming and prone to error. However, rapid advancements in artificial intelligence (AI) and machine learning (ML) have created opportunities to automate real-time detection and recognition of violent behaviors, significantly enhancing the capabilities of modern surveillance systems.

Violence recognition using ML is an emerging and rapidly expanding research field. Its primary goal is to develop algorithms capable of accurately identifying and categorizing violent behaviors from visual data. Such recognition encompasses a wide range of harmful or aggressive activities, including physical assaults, altercations, and other violent acts that may occur in both public and private environments. Unlike object or face recognition, violence detection requires systems to interpret dynamic, subtle, and often context-dependent cues that distinguish normal from violent activities. This challenge is further complicated by varying lighting conditions, occlusions, diverse backgrounds, and the presence of multiple individuals—all of which can affect recognition accuracy.

Early efforts in violence recognition relied heavily on conventional ML techniques such as support vector machines (SVMs), decision trees, and k-nearest neighbors (KNN). These approaches required manual feature engineering, using motion patterns or spatiotemporal descriptors to train the models. While they achieved modest success, their dependence on handcrafted features limited their ability to generalize to complex real-world data, often resulting in poor robustness and scalability.

The introduction of deep learning fundamentally transformed the field. Models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) demonstrated superior performance in learning complex patterns directly from raw data, eliminating the need for manual feature extraction. These models more effectively capture the nuanced characteristics of violent behaviors, leading to more reliable and accurate recognition. Furthermore, advanced architectures such as long short-term memory (LSTM) networks and attention mechanisms have improved the ability to model temporal dependencies and contextual relationships, which are essential for distinguishing violent from non-violent activities. Despite these advancements, violence recognition remains a challenging problem. One of the most pressing issues is the

availability of large, high-quality annotated datasets. The collection and labeling of violent scenarios is not only resource-intensive but also fraught with ethical concerns. Moreover, ensuring generalizability across diverse environments further complicates model training. Overcoming these challenges requires both advanced algorithms and innovative data collection and augmentation strategies.

Evaluation is another critical component of violence recognition research. While common metrics such as accuracy, precision, recall, and F1-score are frequently used, they may not fully capture system performance due to the imbalanced nature of violence datasets, where violent events are relatively rare. More robust evaluation strategies often include the area under the ROC curve (AUC-ROC), the area under the precision-recall curve (AUC-PR), and the Matthews correlation coefficient (MCC). Additionally, cross-validation and testing across heterogeneous datasets are necessary to ensure the robustness and generalizability of proposed systems.

Recent research also explores multimodal approaches to improve violence detection. Although most systems rely solely on visual data, integrating complementary modalities such as audio, depth, or infrared sensing can provide richer contextual understanding. For example, combining audio and video can help differentiate between similar-looking behaviors, such as an argument versus a physical altercation, by analyzing associated sound cues. Depth sensors further enhance spatial awareness, improving the system's ability to interpret complex interactions. These multimodal strategies are particularly valuable in challenging scenarios where visual information alone may be insufficient.

Despite the progress achieved, several open challenges remain. Real-time processing is essential for practical deployment, requiring algorithms optimized for both computational efficiency and hardware performance. Ethical considerations are equally important, particularly regarding privacy and the risk of misuse. To gain public trust and prevent unintended consequences, violence recognition systems must be developed and deployed with transparency, fairness, and accountability.

II. REVIEW OF LITERATURE

Research on violent media, surveillance, and automated detection spans diverse domains, ranging from entertainment studies to real-time security applications. A series of studies have investigated how violence is portrayed and how it impacts audiences, while others have focused on technical advancements in machine learning and deep learning for automated violence detection.

Early studies emphasized the representation of violence in media platforms. For instance, a large-scale content analysis of 2,520 YouTube videos compared amateur and professional productions across categories of popularity, user ratings, and random sampling [1]. The findings indicated that violent content on YouTube, although diverse in context, was less normalized and more grounded in real-world implications than television violence, often appearing in more frightening circumstances. Related research highlighted the importance of monitoring content accessible to children. Given the growing use of YouTube among toddlers, studies examined the cognitive, emotional, and social development effects of video content [2].

While physical violence appeared less frequently, high levels of emotional distress were observed, suggesting that prosocial behaviors and emotional awareness are more beneficial for child development than exposure to aggression.
With rising concerns about public safety, surveillance technologies have become critical in detecting violent behaviors. Traditional surveillance relies on human operators, but human fatigue and error limit effectiveness. Automated methods using ML and DL enable real-time recognition of suspicious actions in both video and image data [3]. Techniques such as VD-Net (Violence Detection Network) integrate IoT-enabled lightweight architectures (ST-TCN blocks, bottleneck layers) to identify hostile actions, achieving 1–4% higher accuracy than existing systems [4].

The social dimensions of violence have also been explored. During the COVID-19 pandemic, the normalization of dating violence on platforms like TikTok gained attention. Studies examined trends such as the "pretend to punch your girlfriend" fad, assessing how young audiences interpret violent dating scenarios and how such exposure reshapes perceptions of relationship equality [5]. Similarly, computational analysis of dissident interviews revealed that responses to governmental repression often dictate violent or non-violent resistance [6]. These findings underscore the complex relationship between violence, political behavior, and social context.

Advances in technical approaches to aggression detection have greatly influenced surveillance systems. For example, real-time frameworks utilizing Spatial Motion Extractors (SME), Short Temporal Extractors (STE), and Global Temporal Extractors (GTE) demonstrated strong performance across datasets such as RWF-2000, Hockey, and the Peruvian VioPeru dataset [7]. Other novel techniques, including Angle-level Co-occurrence Matrices (ALCM) for spatiotemporal video modeling, provided superior results over state-of-the-art video violence detection methods [8].

Meanwhile, media analyses continue to highlight depictions of violence across different formats. A study of 765 primetime television episodes revealed that although violence in children's programming has decreased over two decades, it still surpasses adult programming and is often sanitized or trivialized [9]. Similarly, analysis of 540 violent news reports

linked the prevalence of violent coverage with societal violence trends, shaped in part by media framing and journalist practices [10]. The portrayal of violence against marginalized groups has also been examined: one study analyzed 316 news articles covering transgender fatalities in the U.S., highlighting how language choices can perpetuate harmful stereotypes or support inclusive narratives [11].

Finally, technical research continues to push toward real-time violent event detection in surveillance. Using Deep Recurrent Neural Networks (DRNN) and spatiotemporal (ST) classification, systems trained on the UCF-Crime dataset achieved high performance, with reported accuracy of 98%, precision of 96%, recall of 80%, and F1-score of 78% [12]. These advances illustrate how deep learning models can effectively detect anomalies and violent behaviors in large-scale, real-world surveillance scenarios.

A key challenge in content moderation is balancing safety with the psychological well-being of human moderators. Manual screening often exposes moderators to graphic material, leading to severe emotional and psychological consequences. To address this, researchers proposed a machine learning algorithm that automatically classifies videos as violent or non-violent using both aural and visual cues [13]. The system begins by separating audio and video streams, applying an intelligent audio classifier to distinguish between high- and low-intensity sounds. These results inform a video classifier, which estimates the degree of violence based on the associated sound categories. Beyond content moderation, scholars have examined the online behaviors of violent extremists. Analysis of right-wing extremist engagement on Stormfront revealed five categories of user activity—super-posters, dedicated participants, engaged individuals, dabblers, and non-posters [14]. This classification uncovered distinct posting patterns with direct implications for intelligence and law enforcement agencies. A related study compared the online activities of violent extremists with those of ideologically similar but non-violent individuals [15]. The findings highlighted substantial differences in mobilization, grievances, and rhetorical framing, which could inform risk assessment frameworks for detecting serious online threats.

Research has also expanded toward peace and conflict studies, where video data offers unique opportunities. Video Data Analysis (VDA) was introduced as a valuable addition to the peace research methodological toolkit [16]. Unlike traditional reliance on written or symbolic data, VDA allows researchers to revisit recorded events, observe subtle interaction dynamics, and assess body language or facial expressions. This approach enables the exploration of complex processes such as violence escalation, mediation, and peacebuilding while also raising important ethical questions.

In parallel, surveillance research continues to focus on scalable AI methods for video violence detection. While most efforts have relied on supervised learning, recent advances explored semi-supervised reinforcement learning with a hard attention mechanism [17]. This approach prioritizes key video segments while filtering irrelevant frames, thereby improving precision and reducing annotation requirements. Leveraging a pre-trained I3D backbone, the proposed model achieved state-of-the-art accuracy—90.4% on the RWF dataset and 98.7% on the Hockey dataset.

Finally, deep learning architectures have been applied to automated violence recognition in CCTV footage. Researchers developed lightweight 3D convolutional networks capable of modeling spatiotemporal interactions between individuals and objects [18]. These models demonstrated efficiency with fewer parameters, robustness against compression artifacts in remote streaming scenarios, and a 2% accuracy gain compared to existing approaches. Evaluations across diverse public datasets confirmed their capacity to detect violent acts in complex, real-world environments.

III. CONCLUSION

The application of machine learning to violence detection represents a rapidly advancing field with substantial potential to enhance public safety and security. Through the integration of deep learning techniques, systems have achieved notable progress in accurately recognizing and classifying violent behaviors. Nevertheless, several challenges remain critical to the development of robust solutions. These include the limited availability of high-quality datasets, the need for standardized and reliable evaluation metrics, and the effective fusion of multimodal data sources. Addressing these gaps will be essential for building scalable and dependable violence detection frameworks. Continued research and methodological refinement in these areas can significantly strengthen the effectiveness of automated surveillance systems, thereby advancing their role in public safety, security management, and law enforcement applications.

REFERENCE

1. Weaver, Andrew J., Asta Zelenkauskaite, and Lelia Samson. "The (non) violent world of YouTube: Content trends in web video." Journal of Communication 62, no. 6 (2012): 1065-1083.
2. Choi, Yun Jung, and Changsook Kim. "A content analysis of cognitive, emotional, and social development in popular kid's YouTube." International Journal of Behavioral Development (2024): 01650254241239964.
3. Jain, Mahaveer, and Mukesh Kumar. "A Review of Violence Detection Techniques." In 2024 2nd International Conference on Computer, Communication and Control (IC4), pp. 1-6. IEEE, 2024.

4. Khan, Mustaqeem, Abdulmotaleb El Saddik, Wail Gueaieb, Giulia De Masi, and Fakhri Karray. "VD-Net: An Edge Vision-Based Surveillance System for Violence Detection." IEEE Access 12 (2024): 43796-43808.
5. Maddocks, Sophie, and Fallon Parfaite. "“Watch me pretend to punch my girlfriend”: exploring youth responses to viral dating violence." Feminist Media Studies 24, no. 1 (2024): 103-118.
6. Dornschneider-Elkink, Stephanie, and Nick Henderson. "Repression and dissent: How tit-for-tat leads to violent and nonviolent resistance." Journal of Conflict Resolution 68, no. 4 (2024): 756-785.
7. Huillcen Baca, Herwin Alayn, Flor de Luz Palomino Valdivia, and Juan Carlos Gutierrez Caceres. "Efficient human violence recognition for surveillance in real time." Sensors 24, no. 2 (2024): 668.
8. Hu, Xing, Zhe Fan, Linhua Jiang, Jiawei Xu, Guoqiang Li, Wenming Chen, Xinhua Zeng, Genke Yang, and Dawei Zhang. "TOP-ALCM: A novel video analysis method for violence detection in crowded scenes." Information Sciences 606 (2022): 313-327.
9. Martins, Nicole, and Karyn Riddle. "Reassessing the risks: An updated content analysis of violence on US children’s primetime television." Journal of Children and Media 16, no. 3 (2022): 368-386.
10. Ferguson, Christopher J., Anastasiia Gryshyna, Jung Soo Kim, Emma Knowles, Zainab Nadeem, Izabela Cardozo, Carolin Esser, Victoria Trebbi, and Emily Willis. "Video games, frustration, violence, and virtual reality: Two studies." British journal of social psychology 61, no. 1 (2022): 83-99.
11. Osborn, Max. "US news coverage of transgender victims of fatal violence: An exploratory content analysis." Violence against women 28, no. 9 (2022): 2033-2056.
12. Sahay, Kishan Bhushan, Bhuvaneswari Balachander, B. Jagadeesh, G. Anand Kumar, Ravi Kumar, and L. Rama Parvathy. "A real time crime scene intelligent video surveillance systems in violence detection framework using deep learning techniques." Computers and Electrical Engineering 103 (2022): 108319.
13. Rishab, K. S., P. Mayuravarsha, Yashwal S. Kanchan, M. R. Pranav, and Roopa Ravish. "Detection of Violent Content in Videos using Audio Visual Features." In 2023 International Conference on Advances in Electronics, Communication, Computing and Intelligent Information Systems (ICAECIS), pp. 600-605. IEEE, 2023.
14. Scrivens, Ryan, Garth Davies, Tiana Gaudette, and Richard Frank. "Comparing online posting typologies among violent and nonviolent right-wing extremists." Studies in Conflict & Terrorism (2022): 1-23.
15. Scrivens, Ryan. "Examining online indicators of extremism among violent and non-violent right-wing extremists." Terrorism and political violence 35, no. 6 (2023): 1389-1409.
16. Bramsen, Isabel, and Jonathan Luke Austin. "Affects, emotions and interaction: the methodological promise of video data analysis in peace research." Conflict, Security & Development 22, no. 5 (2022): 457-473.
17. Mohammadi, Hamid, and Ehsan Nazerfard. "Video violence recognition and localization using a semi-supervised hard attention model." Expert Systems with Applications 212 (2023): 118791.
18. Huszar, Viktor Denes, Vamsi Kiran Adhikarla, Imre Négyesi, and Csaba Krasznay. "Toward fast and accurate violence detection for automated video surveillance applications." IEEE Access 11 (2023): 18772-18793.
19. Cheng, Ming, Kunjing Cai, and Ming Li. "RWF-2000: an open large scale video database for violence detection." In 2020 25th International Conference on Pattern Recognition (ICPR), pp. 4183-4190. IEEE, 2021.