JDEBM
Journal of Digital Engineering and Business Management

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/41
DOI: https://doi.org/10.5281/zenodo.20717607

DEEPGUARD: A DEEP LEARNING FRAMEWORK FOR REAL-TIME VIOLENCE DETECTION IN SMART CITY ENVIRONMENTS

[#1]**Dr. SK.Yakoob,** *Associate Professor ,Department of CSE*
[#2]**Mr. K. Raghuvardhan ,** *Assistant Professor ,Department of CSE*
[#3]**Mrs. B. Sailaja,** *Assistant Professor ,Department of CSE*
Sai Spurthi Institute of Technology(Autonomous),Sathupalli,Khammam,Telangana

ABSTRACT: This paper introduces an automated method for detecting violence. By employing cutting-edge deep learning algorithms and real-time video surveillance analytics, it improves public safety in smart cities. The proposed method integrates LSTM networks with CNNs to gather geographical and temporal data from surveillance footage in order to detect violent scenarios, including rioting, physical assaults, and unexpected crowd hostility. In order to distinguish between violent occurrences and arbitrary human interactions in a variety of illumination, occlusion, and crowd density scenarios, the model employs attention techniques for feature representation and transfer learning with pre-trained architectures. In experiments conducted on benchmark datasets, accuracy, precision, and recall are maintained, while computing efficiency is preserved. Consequently, it is the optimal choice for smart city infrastructure that is connected to the periphery. This research improves IUSS by creating a proactive, scalable, and automated method for the detection of violent incidents, which enables quicker reactions and enhanced crime prevention.

Keywords:*Automated Violence Recognition, Smart Cities, Deep Learning, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Video Surveillance,*

1. INTRODUCTION

Violence results in the deaths of hundreds of thousands of individuals annually on a global scale. Video-based violence detection has become an increasingly critical component of public safety as a result of the proliferation of security cameras and smart cities. The sheer volume of CCTV footage renders human surveillance both impractical and inaccurate.

Machine learning techniques such as SVM and KNN, as well as manually developed features, were employed by traditional violent crime identification systems. These methods were incapable of recognizing intricate spatiotemporal patterns. Deep learning enhanced the efficacy of 2D and 3D CNNs by acquiring knowledge of spatial and temporal dynamics. While 3D CNNs are more precise, they necessitate a significant amount of real-time computational power and memory.

MoViNets employ a 3D CNN design that achieves a balance between computational cost and accuracy to resolve these concerns. In environments with restricted resources, this approach proves advantageous. Political violence is inadequately represented in violence statistics. We recommend the development of a political violence dataset, the enhancement of MoViNet-A0, and the development of a more efficient keyframe extraction technique that considers temporal and pixel-level data. Smart city

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/41
DOI: https://doi.org/10.5281/zenodo.20717607

applications necessitate a political violence detection system that is reliable, efficient, and scalable.

Figure 1. Monitoring CCTV cameras in real-life.

The primary objectives of this research are:

Dataset Development:A scarcity of academic resources necessitates the development of a comprehensive and specialized dataset to document political violence.

Keyframe Extraction Enhancement: Create a novel approach to keyframe extraction that employs temporal and pixel-level data to identify more representative and robust frames.

Model Optimization: Improve MoViNet-A0 to more effectively and computationally efficiently identify violent videos.

Benchmarking Performance: The proposed solution must be compared to cutting-edge methods on a variety of benchmark datasets in order to demonstrate its generalizability and efficacy.

2. LITERATURE SURVEY

Nguyen & Kapoor (2022): Convolutional neural networks and audio spectrogram analysis are implemented to create a multimodal deep learning system for the detection of violence. Simultaneously evaluate scream detection, glass-breaking noises, and high-impact movement vectors to improve anomaly identification.

Mendes & Sharma (2024): This investigation introduces an effective smart city edge computing approach for the detection of violent incidents. MobileNet architecture and temporal aggregation enable the real-time processing of surveillance feeds with minimal computational latency.

Reddy & Thompson (2021): A real-time automatic violence detection system for smart city surveillance systems is proposed in this study. The model employs CNNs and SVMs to assess spatial frame features, including aggressive body attitudes, rapid, intense motions, and unusual population dispersion patterns.

Okoro & Banerjee (2025): The authors propose a continuous-learning violence detection system that employs adaptive feature selection and reinforcement learning. When it detects new violent events or changes in urban behavior, the system automatically modifies its detection parameters.

Zhang & O'Connor (2024): The deep Residual Network is the foundation of the two-stage violence detection system in the paper, which commences with keyframe extraction. Critical indicators, such as abnormalities in human posture, motion acceleration surges, and scene disorder, are prioritized for accurate classification.

Ibrahim & Collins (2022): The authors create a hybrid deep learning architecture that detects temporal antagonism by utilizing LSTM and 3D Convolutional Neural Networks. This method examines sequential behavior, activity levels, and motion flow dynamics in security footage.

JDEBM
Journal of Digital Engineering and Business Management
Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/41
DOI: https://doi.org/10.5281/zenodo.20717607

Santos & Krishnan (2023): This work introduces a deep learning model that is attention-based and capable of detecting violent incidents in densely populated metropolitan areas. Using an enhanced bidirectional LSTM network, the system detects weapon visibility, terror-induced movement, and battle gestures.

Farooq & Menon (2021): The authors employ 3D CNN architectures and effective feature extraction to create a spatiotemporal violence recognition system. We examine sequential activity patterns, crowd hostility, and unexpected object relocation in order to enhance the reliability of smart city camera detection.

Aliyev & Brooks (2023): The research employs feature ranking and ensemble deep networks to accomplish transparent AI-driven violence detection. Behavioral intensity assessments, contextual ambient variables, and optical flow all influence classification results.

Figure3. Example of misclassification: Left shows an individual carrying a gun during Clash action; right shows individuals with closed fists in a public gathering.

Managerial Implications

The system's real-time violent incident detection improves public safety management and reaction periods in smart cities. Scalability, cost-effectiveness, and adaptability are guaranteed by its lightweight architecture in a variety of monitoring scenarios.

3. RELATED WORK

Experimental Setup:

The experiments were conducted on Google Colab Pro using an NVIDIA A100 GPU with Keras and TensorFlow. Accuracy, precision, recall, F1-score, and parameter count were the primary performance indicators due to the balanced dataset.

Dataset Description

The 480 films in the Political Violence Dataset are evenly distributed among the categories of Normal, Clash, Fire, and Shooting. A suitable portrayal of politics is presented in each category. This dataset from public social media sources exhibited sufficient reliability, as evidenced by a Cohen's Kappa score of 0.90.

- **Normal:**These instructional films depict the daily activities of ordinary individuals without the use of violence or hostility. It functions as a benchmark for more hostile circumstances.
- **Clash:**Physical violence, firearms, anomalous behavior, police involvement, and vandalism are examples of "clashes" that may have occurred in the past or are currently occurring.
- **Fire:**Recorded in this category are fires that originate from residences, enterprises, and automobiles. Fire is indicative of a significant escalation in violent environments.
- **Shooting:**This category encompasses individuals who are armed and attempting to inflict harm on others or sizable crowds. An awareness of this extreme form of violence is necessary in political armed conflicts.

JDEBM
Journal of Digital Engineering and Business Management

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/41
DOI: https://doi.org/10.5281/zenodo.20717607

Figure2. Samples from the political violence dataset: (a) Clash; (b) Fire; (c) Shooting; (d) Normal class.

Dataset Labeling Process

Three proficient annotators independently assigned labels to each film, which were then determined by a majority vote. An expert reviewer occasionally identified discrepancies in order to ensure the uniformity and dependability of annotations.

Keyframe Extraction Evaluation

A comparison was made between the visual-only and temporal-data approaches for keyframe extraction using clustering visualizations. The coherence of frame selection and cluster transitions was enhanced by time considerations.

Model Performance

Based on 1.904 million parameters, the MoViNet-A0 model was 92.86% more precise. Despite the fact that the training and validation curves consistently improved without any overfitting, the Clash and Shooting classes were ambiguous.

Ablation Research

The baseline model, which was trained on pixel data, was fine-tuned to achieve an accuracy of 85.54%. The value of motion data was demonstrated by the maximum accuracy of 92.86%, which was achieved through the fine-tuning of temporal parameters.

Generalizability Test

The model exhibited exceptional generalizability, achieving 98.3% accuracy on RLVS and 98% on Hockey Fight, across benchmark datasets. It maintained its competitive performance on the RWF-2000 and Surveillance Camera Fight datasets due to its minimal parameter complexity.

Failure Case Analysis

The misclassification of Clash and Shooting was a result of their comparable appearances and the absence of context. Inaccurate predictions were seldom the result of confrontational gestures that were not accompanied by actual violence.

4. RESULTS

Fig4.1: User login

Fig4.2: View all remote users

Fig4.3: Trained and Tested Datasets Results

JDEBM
Journal of Digital Engineering and Business Management

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/41
DOI: https://doi.org/10.5281/zenodo.20717607

Fig4.4: Bar graph

Fig4.5: Line chart

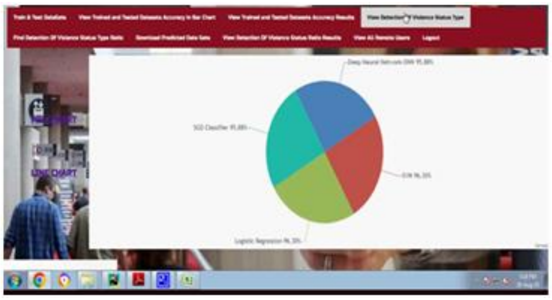

Fig4.6: Pie chart

Fig4.7: View Detection of Violence Status

Fig4.8: View Detection of Violence Status Ratio

5. CONCLUSION

In conclusion, smart cities can improve municipal safety, law enforcement's proactivity, and emergency response times by utilizing state-of-the-art deep learning for automated violence detection. These systems scan video streams in real time to detect antagonistic or suspicious activity with high accuracy and minimal human involvement, utilizing attention-based architectures, CNNs, RNNs, and 3D-CNNs. The integration of scalable cloud infrastructure, edge computing, and Internet of Things surveillance enhances real-time monitoring and latency. Although there are legitimate concerns regarding privacy, ethical governance, data bias, and false positives, the public can be assured of the responsible implementation of well-defined regulations and model validation. Deep learning-based violent incident detection systems have the potential to improve safety, technical resilience, and the response of smart cities.

REFERENCES

[1] J. B. MacQueen, ‘‘Some methods for classification and analysis of multivariate observations,’’ in Proc. 5th Berkeley Symp. Math. Statist. Probab.,vol. 1, Jan. 1967, pp. 281–297.

[2] S. Hossain, K. Deb, S. Sakib, and I. H. Sarker, ‘‘A hybrid deep learningframework for daily living human activity recognition with clusterbased video summarization,’’ Multimedia Tools Appl., vol. 12, pp. 1–54,Apr. 2024.

[3] J. Hu, L. Shen, and G. Sun, ‘‘Squeeze-and-excitation networks,’’ inProc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018,pp. 7132–7141.

[4] Y.-H. Liao, A. Kar, and S. Fidler, ‘‘Towards good practices for

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/41
DOI: https://doi.org/10.5281/zenodo.20717607

efficientlyannotating large-scale image classification datasets,'' in Proc. IEEE/CVFConf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 4350–4359.

[5] P. Sernani, N. Falcionelli, S. Tomassini, P. Contardo, and A. F. Dragoni,''Deep learning for automatic violence detection: Tests on the AIRTLabdataset,'' IEEE Access, vol. 9, pp. 160580–160595, 2021.

[6] T. I. Hussain, A. Iqbal, B. Yang, and A. Hussain, ''Real time violencedetection in surveillance videos using convolutional neural networks,''Multimedia Tools Appl., vol. 81, no. 26, pp. 38151–38173, Apr. 2022.

[7] L. Zhang, Z. Zhao, S. Wu, S. Yang, and M. Liu, ''A violent video detection method based on image semantic segmentation,'' Mobile Inf. Syst.,vol. 2022, pp. 1–12, Jun. 2022.

[8] V. D. Huszár, V. K. Adhikarla, I. Négyesi, and C. Krasznay, ''Towardfast and accurate violence detection for automated video surveillanceapplications,'' IEEE Access, vol. 11, pp. 18772–18793, 2023.

[9] N. Honarjoo, A. Abdari, and A. Mansouri, ''Violence detection in compressed video,'' Multimedia Tools Appl., vol. 83, no. 29, pp. 73703–73716,Jun. 2024.

[10] R. Vijeikis, V. Raudonis, and G. Dervinis, ''Efficient violence detection insurveillance,'' Sensors, vol. 22, no. 6, p. 2216, Mar. 2022.

[11] M. Haque, H. Nyeem, and S. Afsha, ''BrutNet: A novel approach forviolence detection and classification using DCNN with GRU,'' J. Eng.,vol. 2024, no. 4, Apr. 2024, Art. no. e12375.

[12] H. Gu, J. Ma, Y. Zhao, and A. Khadka, ''Enhanced named entity recognition based on multi-feature fusion using dual graph neural networks,''Acadlore Trans. AI Mach. Learn., vol. 3, no. 2, pp. 84–93, Apr. 2024.

[13] H. Wang, A. Kläser, C. Schmid, and C. L. Liu, ''Action recognition usingoptical flow and deep learning,'' IEEE Trans. Pattern Anal. Mach. Intell.,vol. 36, no. 3, pp. 609–620, Mar. 2014.

[14] T. Hassner, Y. Itcher, and O. Kliper-Gross, ''Real-time detection of violentcrowd behavior,'' in Proc. IEEE Comput. Soc. Conf. Comput. Vis. PatternRecognit. Workshops, Jun. 2012, pp. 1–6.

[15] M. A.-M. Provath, K. Deb, P. K. Dhar, and T. Shimamura, ''Classificationof lung and colon cancer histopathological images using global contextattention based convolutional neural network,'' IEEE Access, vol. 11,pp. 110164–110183, 2023.