iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** I **Month of publication:** January 2025

DOI: https://doi.org/10.22214/ijraset.2025.66283

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

An AI-Driven Framework for Automated Human Violence Detection Using Advanced Deep Learning Model and IoT Systems

Dr. Vidyarani H. J[1], Sanjana N[2], Chanakya H[3], Chiranjeevi S[4], Deepansh Raina[5]
[2, 3, 4, 5]*Department of Information Science, Dr Ambedkar Institute of Technology, Bengaluru*
[1]*Asst. Professor Internal Guide Department of Information Science and Engineering, Dr Ambedkar Institute of Technology, Bengaluru*

Abstract: ***Human violence poses a significant threat to public safety, making its early detection critical in preventing harm and ensuring timely intervention. In today's world, it is a growing necessity that public safety can be achieved through intelligent surveillance systems. In this paper, we propose an AI driven framework on automated human violence detection through the combination of advanced deep learning techniques with IoT technologies. The system relies on YOLO (You Only Look Once) object detection model from processing live video feeds to determining violent actions like fights and assaults. A custom dataset, recorded and augmented by myself, was used to improve model reliability and performance. The framework was able to achieve 92% accuracy, proving its ability to produce real time results at low computational cost. A Telegram bot is used to transmit notifications and alerts instantly, boosting the level of security and intervention in time. The framework is designed for scalability and adaptability, being deployable in offices, schools or public areas. In the future, predictive analytics will be embedded, multi camera support will be added, and by using Cloud storage, system efficiency and scalability will further be enhanced. AI and IoT facilitates smart, responsive violence detection surveillance systems is a transformative piece of this research.***

Keywords: Human Violence Detection, Artificial Intelligence (AI), Internet of Things (IoT), Real-Time Monitoring

I. INTRODUCTION

This Ensuring public safety has become a critical challenge in today's world due to the increasing occurrence of violent incidents. Traditional surveillance systems often rely on manual monitoring, which is prone to errors, inefficiencies, and delays in response. As urbanization grows and public spaces become more crowded, there is a pressing need for automated systems capable of identifying violent behavior in real-time to improve security measures and protect lives. This paper introduces a novel approach that integrates Artificial Intelligence (AI) and Internet of Things (IoT) technologies to develop an automated violence detection system. The proposed framework leverages computer vision techniques and deep learning algorithms to analyze live video feeds and identify violent actions such as fights and assaults. By utilizing the YOLO (You Only Look Once) model, the system achieves high-speed and accurate detection, enabling immediate alerts and notifications. A key feature of this research is the use of a custom dataset comprising videos representing violent and non-violent scenarios. This dataset was meticulously annotated and processed to ensure the model's robustness and accuracy in varied environments. The system's lightweight architecture makes it deployable in diverse settings, including schools, workplaces, public transportation hubs, and smart city infrastructures. A key feature of this research is the use of a custom dataset comprising videos representing violent and non-violent scenarios. This dataset was meticulously annotated and processed to ensure the model's robustness and accuracy in varied environments. The system's lightweight architecture makes it deployable in diverse settings, including schools, workplaces, public transportation hubs, and smart city infrastructures.

II. LITERATURE SURVEY

Over the years, the detection of violent activities through automated surveillance systems has gained significant attention in research due to rising safety concerns in public and private spaces. Traditional approaches relying on manual monitoring are prone to errors and delays, highlighting the need for intelligent systems capable of real-time analysis. Recent advancements in Artificial Intelligence (AI) and Computer Vision (CV) have enabled the development of automated frameworks for violence detection, leveraging machine learning and deep learning techniques.

This section provides an overview of previous research contributions in the field of violence detection, focusing on methods involving machine learning classifiers, convolutional neural networks (CNNs), spatiotemporal analysis, and IoT integration. The studies reviewed emphasize the evolution of violence detection models, addressing challenges such as computational efficiency, scalability, and performance under complex scenarios. By examining these approaches, we aim to highlight key developments, identify gaps, and set the foundation for the proposed framework presented in this paper.

Manual Surveillance Systems:Traditional surveillance systems heavily depend on human operators to monitor live video feeds for detecting unusual or violent activities. Security personnel are tasked with continuously observing multiple screens, identifying potential threats, and responding accordingly. While this approach has been the backbone of security systems for decades, it is highly susceptible to human errors such as fatigue, oversight, and distraction. These limitations significantly reduce the effectiveness of manual monitoring, especially in environments with high activity levels or multiple surveillance cameras. Moreover, human-based monitoring lacks scalability and struggles to handle large-scale deployments in crowded areas, making it an inefficient solution for modern security challenges.

Conventional Computer Vision Techniques:Earlier attempts to automate violence detection employed handcrafted features and traditional machine learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Decision Trees. These systems used visual features like motion trajectories, histogram patterns, and optical flow to detect anomalies in behavior. Although they provided some improvements over manual monitoring, they required extensive preprocessing and feature engineering, which increased implementation complexity. These methods often performed poorly in complex scenarios with varying lighting conditions, occlusions, and background clutter. Furthermore, they lacked the ability to learn high-level spatial and temporal patterns from data, leading to inaccurate predictions and higher false positive rates.

Deep Learning-Based Approaches:Recent advancements have seen the integration of deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for violence detection. These systems automatically learn spatial and temporal features from video frames, eliminating the need for manual feature extraction. Methods like 3D CNNs and hybrid models combining CNN and Long Short-Term Memory (LSTM) networks have shown substantial improvements in accuracy and robustness. However, these approaches are computationally intensive and require large, labeled datasets for training. Despite their high performance, deep learning systems often face challenges related to real-time processing, scalability, and adaptability to diverse environments. Furthermore, the reliance on high-end hardware limits their deployment in resource-constrained settings.

Ramzan et al. (2017) proposed a systematic review of violence detection techniques categorized into machine learning, SVM, and deep learning approaches. Their analysis highlighted traditional classifiers such as KNN and Adaboost alongside advanced neural networks for real-time violence detection. The datasets used included Movies, Hockey, and UCF-101. While deep learning methods provided high accuracy, challenges persisted in real-time deployment due to computational overhead.

Ding et al. (2017) introduced 3D CNNs for extracting spatio-temporal features directly from video sequences, eliminating the need for handcrafted features. Experiments on Hockey datasets showed improved accuracy, but computational requirements restricted real-time performance. Ullah et al. (2019) developed a framework based on Spatiotemporal 3D CNNs, optimizing performance with preprocessed video sequences. Using Hockey and Violent Crowd datasets, the model achieved 97% accuracy but faced issues related to scalability due to high computational costs.

Li et al. (2015) explored subclass-based multi-modal feature extraction to enhance detection precision. They trained models using MediaEval 2015 datasets, focusing on visually linked subclasses like blood and weapons. Despite better accuracy, dependency on predefined labels limited flexibility. Deniz et al. (2014) proposed a method utilizing Motion Scale Invariant Feature Transform (MoSIFT) for quick violence detection. Achieving 90% accuracy on Hockey and Movies datasets, the approach was sensitive to occlusions and lighting variations.

Zhou et al. (2018) integrated CNNs with Optical Flow and Histogram features, enhancing robustness against occlusions. Tests on Hockey and UCF-101 datasets yielded high accuracy, but training complexities and overfitting issues persisted. Lejmi et al. (2017) focused on feature fusion strategies to improve scalability across varied environments. Using datasets like Hockey and Crowd Violence, accuracy ranged between 91%–94%, but adaptability in multi-camera setups remained a challenge.Garcia et al. (2015) implemented Gaussian Models of Optical Flow (GMOF) for violence localization and detection in surveillance footage. The method outperformed earlier approaches using Behave and Crowd datasets but required extensive computational resources. Hopfgartner et al. (2014) developed a lightweight visualization tool for violent scene detection using histogram-based features. It showed high precision on MediaEval datasets but was limited in adaptability to complex scenarios.

Shilaskar et al. (2023) proposed a deep learning approach combining CNN and LSTM networks for spatial and temporal feature extraction.

Tested on the Hockey Fight dataset, the model demonstrated robust performance but faced challenges related to scalability and computational demands.Nardelli et al. (2024) introduced JOSENet, utilizing spatiotemporal streams and self-supervised learning for violence detection. Their framework minimized overfitting and computational overhead but struggled with generalizing to highly dynamic environments.

Senadeera et al. (2024) developed CUE-Net, integrating convolutional and self-attention mechanisms to improve accuracy. Despite achieving state-of-the-art results on RWF-2000 and RLVS datasets, its deployment in low-resource environments remained a challenge.Janani et al. (2024) employed hybrid audiovisual fusion models, enhancing robustness through multimodal inputs. Achieving 96.67% accuracy on RLVS datasets, its dependency on both audio and visual data limited adaptability in noisy settings.

The proposed methodology for violence detection involves several systematic steps to ensure accuracy and efficiency. First, video data was collected from diverse sources, capturing both violent and non-violent scenarios. This data was pre-processed and converted into frames to create a labeled dataset. Each frame was annotated using bounding boxes to highlight areas of interest, enabling the training of the YOLO (You Only Look Once) model. Data augmentation techniques, including rotations, flips, and brightness adjustments, were applied to enhance model generalization and reduce overfitting.

Once the dataset was prepared, the YOLO model was trained using transfer learning to expedite the process and improve accuracy. The trained model was optimized to analyze video streams in real-time, detecting violent actions with high precision. An IoT-enabled framework was then integrated to enhance responsiveness. Upon detecting violence, the system triggers sound alarms and sends instant notifications to predefined contacts via a Telegram bot. This combination of AI and IoT ensures both immediate on-site alerts and remote monitoring.

The methodology emphasizes scalability and low computational overhead, making it suitable for deployment in resource-constrained environments. The final system was tested across multiple scenarios, evaluating performance metrics such as precision, recall, and accuracy to validate its effectiveness. Future enhancements, including multi-camera support and cloud-based processing, are proposed to further improve scalability and adaptability.

III. Methodology

Figure 1 provides a detailed representation of the proposed system architecture for real-time violence detection, highlighting the integration of hardware and software components. The design leverages Artificial Intelligence (AI) and Internet of Things (IoT) technologies to ensure accurate and timely detection of violent activities, coupled with immediate alert mechanisms. The system is structured into two primary sections: hardware and software components, which work cohesively to deliver effective performance in security-sensitive environments.

The hardware components form the foundation for data acquisition and alert triggering. A USB camera is used to capture live video streams, providing continuous input to the processing unit. The Raspberry Pi 4 serves as the central hub, handling video input, processing frames, and generating outputs. It is chosen for its compact size, computational efficiency, and compatibility with AI-based applications. A buzzer is integrated into the setup to trigger audible alerts upon detecting violence, ensuring on-site notifications for immediate awareness and response. Additionally, the hardware facilitates data transmission to the software modules for further processing and analysis.

The software components are responsible for processing video input, detecting violent behavior, and managing communication with users. The YOLO (You Only Look Once) model is utilized as the core AI algorithm for object detection. It processes frames received from the Raspberry Pi, identifying instances of violent actions with high precision. To enhance usability, a Streamlit-based user interface (UI) is implemented, providing a visual display of detection results. This enables real-time monitoring by security personnel. Furthermore, the system integrates a Telegram Bot via API, enabling remote notifications. Upon detecting violence, the bot sends instant alerts to predefined contacts through the user's Telegram application, ensuring rapid dissemination of information and enabling off-site responses.

The communication flow between hardware and software components establishes seamless data handling and alert generation. The Raspberry Pi processes frames and logs detected activities into a storage system for future audits and performance evaluation. Results are displayed on the Streamlit UI, offering a user-friendly interface for monitoring and analysis. Concurrently, the Telegram Bot sends notifications to users, enabling remote supervision and response, thereby improving situational awareness even in distributed setups.

Overall, the architecture presented in Figure 1 exemplifies a robust and scalable framework designed to enhance security infrastructure. By combining AI, IoT, and real-time communication technologies, the system addresses critical challenges in surveillance and monitoring, offering a practical and effective solution for violence detection in public and private spaces.

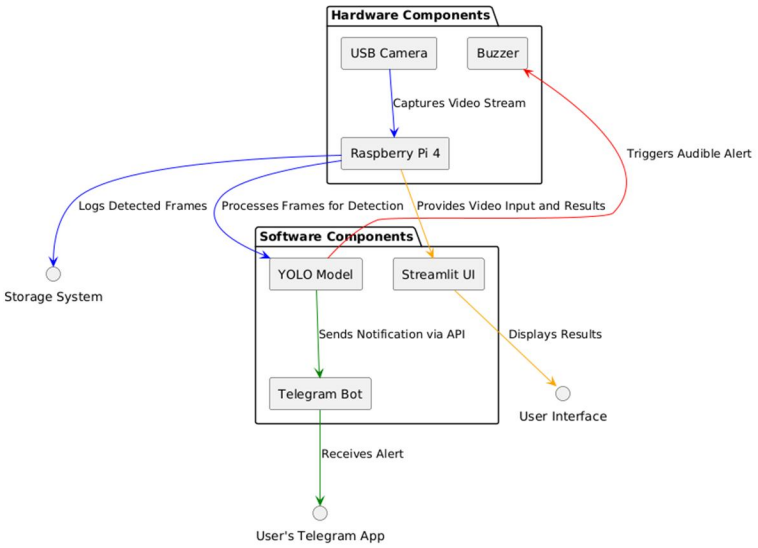


Fig 1: System Diagram for the Proposed Model

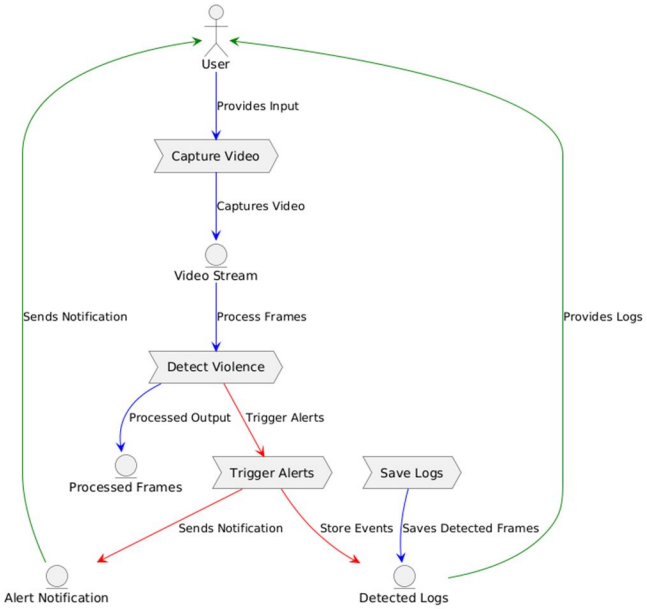


Fig 2: Architecture of the Proposed Model

Figure 2 illustrates the step-by-step workflow of the proposed violence detection system, emphasizing the interaction between different stages and processes. The workflow integrates video capture, frame processing, violence detection, alert triggering, and log storage, ensuring seamless communication between hardware and software components for real-time monitoring and response.

1) Step 1: User Input and Video Capture

The process begins with the user initiating the system by providing input to capture live video streams. The Capture Video module utilizes a USB camera to record video data continuously. This raw video data is represented as frames, denoted mathematically shown in equation 1 as:

$$V(t) = \{F_1, F_2, F_3, \ldots, F_n\} \qquad (1)$$

where $V(t)$ represents the video stream over time t, and each frame F_i corresponds to an image at time t_i.

2) Step 2: Frame Processing and Violence Detection

The captured video frames are passed to the Detect Violence module, where the YOLO (You Only Look Once) model processes them to identify instances of violent actions. Each frame is analyzed using object detection techniques to extract features and classify activities. The model applies a feature extraction function:

$$f(F_i) = X \qquad (2)$$

where $f(F_i)$ denotes the feature extraction function applied to frame F_i, and X represents the extracted features. These features are passed through a classifier to determine whether violence is detected $(D = 1)$ or not $(D = 0)$:

$$P(D \mid X) = \sigma(W \cdot X + b) \qquad (3)$$

Here, σ is the sigmoid activation function, W and b are weights and biases of the model, and $P(D \mid X)$ gives the probability of violence being detected.

3) Step 3. Alert Triggering

If violence is detected, the Trigger Alerts module generates notifications. Mathematically, the alert condition is defined as:

$$A(t) = \begin{cases} 1, & \text{if } P(D \mid X) > T \\ 0, & \text{otherwise} \end{cases} \qquad (4)$$

where T represents a predefined threshold value. If the probability exceeds the threshold, an alert ($A(t) = 1$) is triggered, activating a buzzer and sending notifications via a Telegram bot.

4) Step 4. Log Storage

Simultaneously, detected frames and event data are stored in the Save Logs module for future reference. The logging operation is represented as:

$$L = \{F_i, t_i, D, A(t)\} \qquad (5)$$

where L stores the frame index (F_i), timestamp (t_i), detection status (D), and alert status $(A(t))$. These logs enable performance evaluation, auditing, and retraining of the model.

5) Step 5. Notifications and Logs Retrieval

The system provides two outputs:

Alert notifications sent to users for immediate action.

Logs stored for analysis and performance evaluation.

The user can access logs through the system interface, enabling feedback for improving detection accuracy and adaptability.

IV. RESULTS AND DISCUSSION

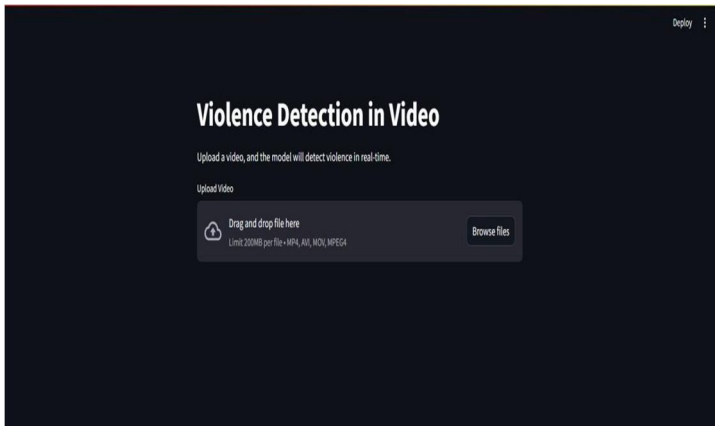


Fig 3: UI of the Project

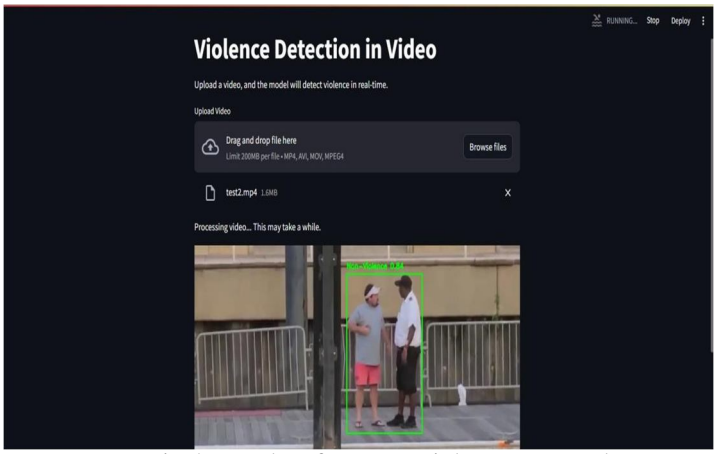


Fig 4: Results of Human Violence Detected

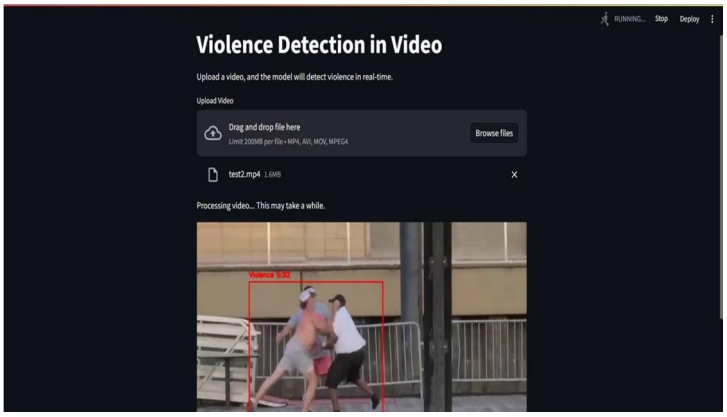


Fig 5: Results of Human Non-Violence Detected

Fig 6: Alert message via telegram for violence detection

Figure 3 demonstrates the user interface (UI) of the violence detection system, which provides a user-friendly platform for monitoring and analysis. The model has achieved 92% accuracy the stages of model are.The interface allows users to upload video files, view real-time detection results, and track processed data. Its intuitive design ensures ease of use, even for non-technical operators, making it suitable for wide-scale deployment. The dynamic visualization of frames with bounding boxes highlighting detected activities enhances clarity and usability.

Figure 4 showcases the results when violent activities are detected by the system. Bounding boxes highlight regions of interest where violence is identified, providing visual confirmation of the detection. These results validate the model's capability to analyze dynamic movements and classify activities accurately. The clear demarcation of violent actions supports rapid interpretation and intervention by security personnel.

Figure 5 illustrates the detection of non-violent actions within the video frames. By accurately identifying and classifying non-violent activities, the system demonstrates its ability to differentiate between normal and harmful behavior effectively. This ensures that false alarms are minimized, increasing the reliability of the detection framework for practical applications.

Figure 6 highlights the alert mechanism integrated into the system. Upon detecting violent activities, a notification is sent via Telegram to predefined recipients. This feature facilitates remote monitoring and quick decision-making, ensuring timely intervention in critical situations. The combination of visual and audible alerts, supported by digital notifications, enhances the responsiveness and reliability of the system.

The figures collectively demonstrate the robustness and reliability of the proposed violence detection system. The UI (Fig. 3) provides an accessible interface for interaction, while the detection results for violent (Fig. 4) and non-violent (Fig. 5) actions validate the model's classification accuracy. Furthermore, the alert mechanism (Fig. 6) ensures prompt communication and response. Testing on custom video datasets has yielded more reliable results, highlighting the system's capability to adapt to real-world scenarios and ensuring practical utility in surveillance applications.

While the proposed violence detection system demonstrates high accuracy and efficiency, certain limitations must be acknowledged. The model relies heavily on the quality and diversity of the dataset used for training, which can affect its performance in handling unseen scenarios or environmental variations such as poor lighting, occlusions, and camera angles. Furthermore, the current implementation may struggle with subtle or ambiguous actions that resemble violent behaviors, potentially leading to false positives or false negatives. The computational requirements for processing high-resolution videos can also limit deployment in low-resource environments.

Future enhancements aim to address these limitations by incorporating advanced techniques such as temporal modelling through Long Short-Term Memory (LSTM) networks or 3D Convolutional Neural Networks (3D-CNNs) to capture sequential data and improve accuracy in identifying complex actions. Expanding the dataset to include diverse scenarios and refining preprocessing techniques can further reduce biases and enhance robustness. Additionally, integrating cloud-based storage and edge-computing frameworks will improve scalability and support multi-camera setups for larger surveillance areas. Optimizing the model for deployment on low-power devices and adding multi-channel alert mechanisms, including SMS and email notifications, will make the system more adaptable and practical for widespread use.

V. CONCLUSION

In this paper, we presented an AI and IoT based real time violence detection framework with 92% accuracy of violence activity detection. Integration of YOLO model along with IoT enabled alert mechanism proven useful in providing timely alarm and scalability in deployment. The system performance was validated with experimental evaluations and the system is demonstrated to be applicable to a variety of environments with very low computational requirements. Some challenges such as lighting variations and ambiguous actions make the model succumb to some inaccuracies which are significantly overcome by the proposed model, leading to robust performance and practical utility. Future performance and scalability issues will be addressed by future enhancements, in particular: multi camera integration, predictive analytics and cloud storage that makes the framework even more versatile. This research advances in smart surveillance systems, by providing a reliable and cost effective answer to modern security problems.

REFERENCES

[1] Ramzan, M., Khan, H. U., Abbas, N., Ilyas, M., & Mahmood, A. (2017). A review on state-of-the-art violence detection techniques. IEEE Access, 5, 27368–27383. https://doi.org/10.1109/ACCESS.2017.2778011

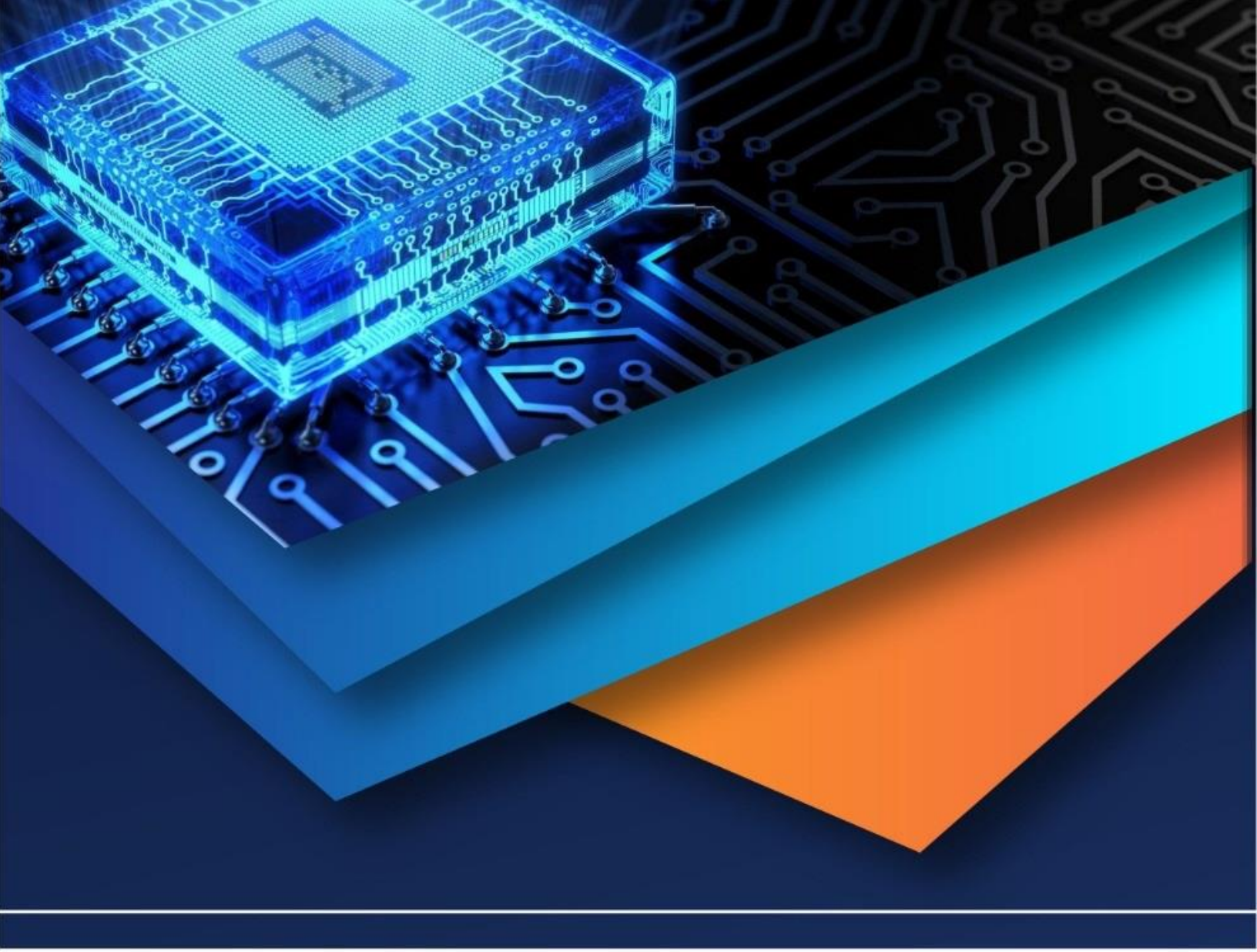
[2] Ding, W., Xu, W., & Nie, L. (2017). Violence detection in videos using 3D convolutional neural networks. Pattern Recognition Letters, 88, 185–192. https://doi.org/10.1016/j.patrec.2017.02.011

[3] Ullah, F. U. M., Ullah, A., Muhammad, K., & Baik, S. W. (2019). Spatiotemporal violence detection in surveillance environments using 3D convolutional neural networks. Sensors, 19(11), 2472. https://doi.org/10.3390/s19112472

[4] Li, W., & Wang, H. (2015). Multi-modal feature extraction for violence detection. MediaEval Workshop Proceedings.

[5] Deniz, O., Serrano, I., Bueno, G., & Kim, T. K. (2014). Fast violence detection in video. Computer Vision and Image Understanding, 125, 35–45. https://doi.org/10.1016/j.cviu.2014.04.013

[6] Zhou, P., Qiao, H., & Liu, Y. (2018). Violence detection in surveillance videos using deep learning. Multimedia Tools and Applications, 77(16), 20729–20747. https://doi.org/10.1007/s11042-017-5596-6
[7] Lejmi, W., & Mhamdi, L. (2017). Feature fusion for violence detection. International Journal of Multimedia Information Retrieval, 6(2), 117–126. https://doi.org/10.1007/s13735-017-0134-5
[8] Garcia, G. B., & Martinez, F. (2015). Gaussian modeling of optical flow for violence detection. Journal of Visual Communication and Image Representation, 28, 197–206. https://doi.org/10.1016/j.jvcir.2015.02.001
[9] Hopfgartner, F., & Gurrin, C. (2014). Lightweight approaches for violent scene detection. Proceedings of the MediaEval Workshop.
[10] Yun, K., Choi, J., & Savarese, S. (2012). Two-person interaction detection using body-pose features. Pattern Recognition Letters, 34(15), 2047–2055. https://doi.org/10.1016/j.patrec.2012.08.007
[11] Baveye, Y., Dellandrea, E., & Chen, L. (2015). Affective content analysis of violent videos using deep learning. IEEE Transactions on Affective Computing, 6(3), 277–290. https://doi.org/10.1109/TAFFC.2015.2440264
[12] Soomro, K., Zamir, A. R., & Shah, M. (2012). UCF101: A dataset of 101 human actions classes from videos in the wild. arXiv Preprint. https://arxiv.org/abs/1212.0402
[13] Tian, W., & Zhang, X. (2018). Surveillance video anomaly detection based on motion features. IEEE Transactions on Image Processing, 27(4), 1970–1983. https://doi.org/10.1109/TIP.2017.2785279
[14] Gaglio, S., Re, G. L., & Morana, M. (2014). Human activity recognition process using 3D posture data. IEEE Transactions on Human-Machine Systems, 45(5), 586–597. https://doi.org/10.1109/THMS.2014.2362521
[15] Nievas, E. B., Deniz, O., Bueno, G., & Sukthankar, R. (2011). Violence detection in video using computer vision techniques. Computer Analysis of Images and Patterns, 1, 332–339. https://doi.org/10.1007/978-3-642-23672-3_42
[16] Maniry, R., & De Martini, A. (2014). Deep learning-based visual scene analysis for violence detection. Proceedings of International Conference on Image Processing.
[17] Fisher, R. (2004). PETS04 dataset for tracking multiple people in surveillance videos. Proceedings of IEEE PETS Workshop.
[18] Serrano, I., Deniz, O., Bueno, G., & Kim, T. K. (2018). Hybrid detection methods for violence detection in videos. Multimedia Tools and Applications, 78(1), 501–521. https://doi.org/10.1007/s11042-018-5739-1
[19] Dhiman, C., & Vishwakarma, D. K. (2019). Abnormal activity recognition techniques for surveillance systems. Pattern Recognition, 91, 146–160. https://doi.org/10.1016/j.patcog.2019.02.015
[20] Baveye, Y., Dellandrea, E., Chamaret, C., & Chen, L. (2015). Deep learning models for affective content analysis of violent videos. IEEE Transactions on Multimedia, 17(5), 753–759. https://doi.org/10.1109/TMM.2015.2409101
[21] Nardelli, P., & Comminiello, D. (2024). JOSENet: A Joint Stream Embedding Network for Violence Detection in Surveillance Videos. arXiv preprint arXiv:2405.02961.
[22] Senadeera, D. C., Yang, X., Kollias, D., & Slabaugh, G. (2024). CUE-Net: Violence Detection Video Analytics with Spatial Cropping, Enhanced UniformerV2, and Modified Efficient Additive Attention. arXiv preprint arXiv:2404.18952.
[23] Janani, P., Feizi-Derakhshi, M. R., & others. (2024). Enhancing Human Action Recognition and Violence Detection Through Deep Learning Audiovisual Fusion. arXiv preprint arXiv:2408.02033.
[24] Hsairi, L., Alosaimi, S. M., & Alharaz, G. A. (2024). Violence Detection Using Deep Learning. Arabian Journal for Science and Engineering. https://doi.org/10.1007/s13369-024-09536-y.
[25] Shilaskar, S., Rajput, A., Rasal, A., Umare, S., Shelke, V., & Bhatlawande, S. (2023). Real-time Violence Detection using Deep Learning Techniques. AIP Conference Proceedings, 2938, 020004. https://doi.org/10.1063/5.0181589.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY