International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Smart Surveillance for Violence Detection

G. Sudeepthi[1], R. V. Anjana Reddy[2], T. Vaishanvi[3], Swapna. C[4]

[1,2,3,4]Stanley College of Engineering & Technology For Women, Hyderabad.

ABSTRACT

By automatically identifying violent actions in real-time, smart surveillance for violence detection is a state-of-the-art technology that aims to improve public safety. The technology analyzes live video feeds from security cameras using sophisticated machine learning algorithms and computer vision techniques. The device notifies authorities in real time when it detects anomalous behaviors, such physical attacks, fights, or other violent conduct, allowing for prompt response and intervention.

The project's main elements include motion analysis, object identification, and behavior recognition, all of which are incorporated into a strong surveillance framework. The system can reliably categorize and distinguish between normal and violent actions in a variety of settings, including public areas, businesses, and educational institutions, by utilizing convolutional neural networks (CNNs) and deep learning models. This smart surveillance technology guarantees faster threat identification, which lowers the potential harm caused by violent situations, and it also increases the efficacy of standard video monitoring by decreasing manual oversight. With its emphasis on maximizing accuracy, reducing false positives, and guaranteeing scalability for extensive monitoring networks, the project is a useful addition to contemporary security infrastructure.

Keywords: Smart surveillance, Violence detection, Machine learning, Computer vision, Real-time video analysis, Convolutional neural networks (CNNs).

INTRODUCTION:

Using automated video surveillance, the "Smart Surveillance for Violence Detection Using Domain Deep Learning" initiative is a creative way to increase public safety. Concerns about security are becoming more prevalent in the modern world, especially in public areas where human surveillance may be ineffective, expensive, or prone to mistakes. Due to their heavy reliance on manual observation, traditional surveillance systems may overlook crucial times during violent occurrences. In order to overcome these constraints, this research uses a deep learning-based system that can identify violence in real time, minimizing the need for human operators and facilitating quicker reaction times. Deep learning models that are domain-specific and trained to identify violent behavior in a variety of settings power the system's core.

These models are constructed utilizing state-of-the-art methods such as Recurrent Neural Networks (RNNs) for comprehending the temporal sequence of activities and Convolutional Neural Networks (CNNs) for extracting spatial features.

This system may be used in a variety of monitoring settings, such as cities, transit hubs, educational institutions, and other busy places. One of the project's main goals is to make sure the system can function in real-time, providing prompt and precise identification with few false positives. In order to guarantee reliable performance, it also focuses on solving the problem of imbalanced datasets, where violent episodes are comparatively uncommon in comparison to typical behavior. Overall, by providing a scalable, effective, and intelligent system for violence detection—which uses cutting-edge deep learning techniques to make public areas safer—this study advances the field of smart surveillance.

BACKGROUND:

Due to the quick development of deep learning and artificial intelligence (AI) technologies, the field of smart surveillance has seen tremendous growth in recent years. Conventional surveillance systems mostly depend on human operators to keep an eye on video feeds for any threats or illegal activity. Although human monitoring works well in controlled environments, it is naturally constrained by things like exhaustion, attention, and the inability to cover several regions at once. Large-scale surveillance networks, like those found in shopping centers, schools, public transportation systems, and metropolitan areas, where ongoing monitoring is essential to maintaining public safety, make these flaws much more noticeable.

Machines can now "learn" from massive volumes of data and provide intricate predictions thanks to deep learning, which has completely changed the field of computer vision and pattern recognition. Deep learning methods, especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have demonstrated impressive results in the surveillance domain in terms of video stream analysis, object identification, and the detection of anomalous behaviors. RNNs are excellent at temporal analysis, detecting patterns of activities across time, but CNNs are especially well-suited for spatial analysis, such as identifying gestures or actions. An intelligent, automated system that can identify violence in real time is more important than ever because of the growing threat of violence in public areas. An important tool for public safety, such a system can increase the effectiveness of current surveillance infrastructure by decreasing the need for human monitoring and increasing the speed and precision of incident detection.

PROBLEM STATEMENTS:

Even with the improvements in surveillance technology, there are still a lot of holes in the ability to identify violent activity in public spaces in real time. Human operators are frequently used in current systems, and while watching many channels, they may become distracted or tired and miss important moments. Additionally, the difficulty of precisely identifying violence persists, especially in busy and complicated surroundings, even if current AI-based systems have made progress in object detection and facial recognition.

Furthermore, a prevalent problem with a lot of automated systems is the high prevalence of false positives, which occurs when non-violent acts are mistakenly reported as violent. These erroneous warnings have the potential to cause needless actions and erode systemic confidence. Resolving this issue calls for precise models that can detect aggressive behavior in addition to the capacity to filter out noise and minimize false positives. Without a highly effective and dependable system for automated, real-time violence detection, public areas are susceptible to delayed reactions, which might cause violent occurrences to escalate before help arrives. A more advanced, domain-specific deep learning approach is therefore required in order to detect violent behaviors with accuracy while reducing false positives and operational inefficiencies.

OBJECTIVES:

This project's major goal is to use domain-specific deep learning models to create and deploy an intelligent surveillance system for real-time violence detection. By analyzing surveillance camera data, this technology will identify and categorize violent incidents, sending out notifications that will enable security guards to react quickly. The following is an outline of the main goals:

- Real-time Detection: The system will be able to analyze live video streams and decide in real-time whether violent behavior is present, guaranteeing prompt actions in urgent circumstances.
- CNN-based feature extraction: Convolutional Neural Networks (CNNs) will be used by the system to identify postures and behaviors that could be signs of violent conduct, such as abrupt, aggressive gestures or physical altercations, by extracting spatial characteristics from video frames.
- Using Recurrent Neural Networks (RNNs) for Temporal Pattern Analysis: RNNs will be used to study action sequences across time, assisting in the differentiation of solitary gestures from continuous violent episodes.
- Decrease in False Positives: The system will use domain-specific datasets and sophisticated training techniques to minimize false positives, making sure that warnings are only triggered when actual violent behavior is found.
- Scalability and Adaptability: The system will be made to be flexible enough to work in a variety of situations, including private security settings and public areas, guaranteeing scalability for a range of use cases.

METHODOLOGY:

Using deep learning techniques, the Smart Surveillance for Violence identification Using Domain Deep Learning project's methodology is made to guarantee precise real-time identification of violent conduct. From gathering data to evaluating and deploying the model, there are a number of crucial milestones in the process. Using video footage as input, the system will be constructed and processed via a number of steps, such as preprocessing, feature extraction, classification, and the creation of real-time alerts.

Information Gathering:

The methodology's initial stage is to gather and prepare a suitable dataset:

- **Video Datasets:** Videos showing both violent and nonviolent acts will be collected as part of curated datasets. Training and testing will be conducted using publicly accessible datasets such as the Movies Action Dataset, Crowd Violence Dataset, and Hockey Fight Dataset. To add more variables to the dataset, artificial or simulated video material may also be produced.
- **Footage Labeling:** Annotations identifying violent and non-violent activities will be applied to the gathered video footage. For the deep learning model to be trained under supervision, these labels will be essential.

Preprocessing Data:

To guarantee compatibility with the deep learning models and to maximize performance, the raw video data will undergo preprocessing:

- **Frame Extraction:** To enable frame-by-frame analysis, videos will be divided into individual frames. Continuous video footage is converted into discrete picture data in this stage.
- **Data Normalization:** To increase training efficiency, video frames will be uniformly fed to the model by being normalized (rescaled to a consistent size, quality, and format).
- **Data Augmentation:** Methods like rotation, flipping, and zooming will be used to produce variants of the video data and artificially augment the dataset in order to increase the resilience of the model. This improves the model's ability to generalize and learn from a variety of inputs.

Extraction of Features

The system must extract temporal and geographical information from the video stream in order to recognize violence:

- **Convolutional Neural Networks (CNNs)**: The process of extracting spatial characteristics from individual video frames will be carried out using Convolutional Neural Networks (CNNs). CNNs are perfect at identifying patterns in a frame, including gestures, body position, and movement. Key spatial signs of aggression, including raised fists, aggressive body gestures, or abrupt physical contact, will be recognized by the CNN when it has been taught.
- **Pretrained Models:** Using pre-existing networks that have been trained on huge datasets, such as ImageNet, pretrained models such as VGG-16, ResNet, or InceptionNet can be utilized to extract features at this stage.

Analysis of Time

Temporal analysis is crucial because violent acts frequently follow a series of events throughout time:

- **Recurrent neural networks (RNNs):** These networks, also known as Long Short-Term Memory (LSTM) networks, will be used to examine the temporal patterns among video frames. By capturing the time dependencies and motion patterns that distinguish violent conduct from innocuous activities, these models will be able to identify violent behavior.
- **Spatial-temporal Feature Fusion:** To create a complete feature vector that captures the video's appearance and motion, the output from CNNs (spatial features) and RNNs (temporal features) will be fused. By combining visual patterns and action sequences, this fusion enables the model to make better judgments.

Categorization

Classifying the behavior as violent or non-violent comes next once the characteristics have been retrieved and fused:

- **Fully Connected Layers:** The fused feature vectors will be processed by a fully connected (dense) neural network layer. Using the learnt characteristics, this layer will operate as the classifier, predicting whether the observed action is violent or not.
- **Softmax Activation:** To generate probability ratings for each class (violent or non-violent), the final output layer will employ a softmax function. The model will forecast the class with the highest likelihood.

Training Models

A supervised learning strategy will be used to train the deep learning model:

- **Loss Function:** The error between each sample's actual label and projected class will be determined using a loss function, such as binary cross-entropy.
- **Optimization:** To minimize the loss function and iteratively update the model's weights, an optimizer such as Adam or SGD (Stochastic Gradient Descent) will be employed.
- **Training Procedure:** A subset of the dataset (the training set) will be used to train the model, while a different subset (the validation set) will be used to track the model's performance and avoid overfitting. When performance on the validation set no longer improves, training will be stopped using early stopping approaches.

Dealing with Unbalanced Information

The dataset is unbalanced since violent acts are usually few in comparison to non-violent activities. To deal with this:

- **Data Resampling:** To balance the dataset, methods such undersampling the majority class (non-violent activities) or oversampling the minority class (violent actions) will be used.
- **Cost-sensitive Learning:** This is an alternate strategy that encourages the model to concentrate on accurately recognizing violence by penalizing violent incident misclassification more severely than non-violent incident misclassification.

Model Assessment

The model's performance will be assessed following training:

- **Metrics:** F1-score, ROC-AUC curve, recall, accuracy, and precision are important performance indicators. These measures will shed light on how well the algorithm can identify violent crimes while reducing false positives.
- **Cross-validation:** To make sure the model is resilient and generalizable to unknown data, K-fold cross-validation will be employed.
- **Confusion Matrix:** To see how well the model performs in categorizing violent and non-violent activities, a confusion matrix will be displayed.

Implementation in Real Time

The learned model will be included into a live video feed for practical uses:

- **Real-Time Processing**: The system will be set up to continually watch live video streams and look for indications of violent activity in every frame.
- **Alert System:** The system will automatically sound an alarm when it detects aggressive activity, which may include auditory and visual cues or alerts to security staff.

Testing and Deployment

Following completion of development, the system will be implemented and evaluated in actual settings:

- **Pilot Testing:** To make sure the system works properly in real-time and can manage different lighting situations, angles, and crowd densities, it will be tested in controlled settings.
- **Feedback Loop:** In order to increase accuracy and responsiveness, the system will be improved by retraining the model or modifying parameters based on input from the testing phase.

DATA:

Real Life Violence Situations Dataset: This dataset is made up of videos that were gathered from several sources and depict violent incidents that actually happened. It seeks to offer a thorough resource for creating and evaluating algorithms for identifying violence in authentic situations. For balanced training, the dataset usually consists of both labeled violent and non-violent clips.

- **Violent-Flows Dataset:** This dataset contains a variety of violent behavior examples and focuses on violent behaviors in films. It includes video recordings labeled with temporal segments showing violent activities and is frequently used for action detection tasks. The dataset is helpful for motion analysis since it highlights the motion flow in violent incidents.
- **Movie Fight Dataset:** This collection includes snippets of violence and staged battles from action films. To assist in recognizing different kinds of fight scenes, including choreography, combat tactics, and other aspects of staged violence, each clip has been annotated. This dataset is especially helpful for researching how violence is portrayed in movies.
- **Crowd Violence dataset:** This collection includes video footage showing crowd dynamics that may lead to violent incidents. It is intended to help people understand how violence may escalate in group

situations and contains examples of both violent and non-violent crowd behaviors. Usually, the annotations concentrate on identifying violent incidents and evaluating crowd dynamics.

- **UCF101 Dataset** is a widely used standard for action recognition, even though it is not solely centered on violence. Among its 101 action categories are those that are associated with violence (e.g., punching, fighting). Since the dataset includes a diverse range of YouTube videos, it may be used to train models for broad action detection, including violent behaviors.

Tools and Technology:

- Python as a programming language;
- TensorFlow, Keras, PyTorch, and other deep learning libraries;
- OpenCV and FFmpeg as video processing libraries;
- GPU-accelerated computation (such as NVIDIA GPUs) for real-time deployment and model training.

LITERATURE REVIEW

S. N O	Title	Authors	Year of Published	Accuracy	Gaps	Approaches
1	Violence Detection in Videos using Deep Recurrent and Convolutional Neural Networks	Abdarahmane Traoré, Moulay A. Akhloufi	2024	99% on Hockey Dataset, 93.53% on Violent Flow Dataset	Performance on optical flow datasets is not sharp enough; Lacks handling for large motion in videos	Combined 2D CNN and RNN (GRU or LSTM) with optical flow. Used EfficientNet-B0 with RGB frames for spatial-temporal feature extraction
2	Cover the Violence: A Novel Deep-Learning-Based Approach Towards Violence-Detection in Movies	Samee Ullah Khan, Ijaz Ul Haq, Seungmin Rho, Sung Wook Baik, Mi Young Lee	2019	Not Specified	Limited by manual processing of movie data; only frames with maximum saliency were used	Fine-tuned MobileNet model for frame selection and violence detection in movies
3	State-of-the-Art	Milon Biswas,	2022	97.23%	Models are computationall	Machine Learning and

	Violence Detection Techniques: A Review	Afjal Hossain Jibon, Mim Kabir, Khandokar Mohima, et al.			y expensive; Limited real-time applications	Deep Learning methods, including CNN and SVM for violence detection
4	Violence Detection in Surveillance Video Using Low-Level Features	Peipei Zhou, Qinghai Ding, Haibo Luo, Xinglin Hou	2018	Superior performance compared to previous methods on benchmark datasets	Struggles with varied human body movements and dynamic backgrounds	Uses Local Histogram of Oriented Gradient (LHOG) and Local Histogram of Optical Flow (LHOF), with SVM classifier (5)
5	Detection of Violence Behavior using Deep Learning Technique	Mohamed Safaa M. Shubber, Ziyad Tariq Mustafa Al-Ta'i	2022	94% (AvdDS dataset) and 99% (SfDS dataset) (68c7269b59919c36)	Classification errors when friendly movements resemble violent actions	Modified CNN-VGG16 model using transfer learning (68c7269b59919c36)
6	Threshold Active Learning for Physical Violence Detection on Video Frames	Itzel M. Abundez, Roberto Alejo, Francisco Primero, Everardo Gutiérrez, Otniel Portillo-Rodríguez, Juan Alberto Antonio Velázquez	2024	Precision up to 96% (algorithms-17-00316-v2)	Struggles with ambiguous images, requiring manual intervention	Hybrid model combining pre-trained neural networks and active learning

7	Novel Deep Feature Fusion Framework for Multi-Scenario Violence Detection	Sabah Abdulazeez Jebur, Khalid A. Hussein, Haider Kadhim Hoomod, Laith Alzubaidi	2023	97.66% (RLVS dataset), 92.89% (Hockey dataset)	Struggles with generalization across diverse environments	Feature fusion of Xception, Inception, and InceptionResNet models
8	Deep Learning for Automatic Violence Detection: Tests on the AIRTLab Dataset	Paolo Sernani, Nicola Falcionelli, Selene Tomassini, Paolo Contardo, Aldo Franco Dragoni	2021	99.2% on Crowd Violence dataset	99.2% on Crowd Violence dataset	3D CNN, ConvLSTM, and transfer learning-based models
9	Human Violence Detection Using Deep Learning Techniques	S. A. Arun Akash et al.	2022	74%	Focuses on detection using YOLO-v5 and Inception-v3 but lacks testing on diverse environments and real-time deployment challenges.	Deep learning using YOLO-v5 and Inception-v3 for object detection and classification
10	Violence Detection in Social Media-Review	U. Dikwatta, T.G.I. Fernando	2019	Not Provided	Focuses only on social media violence. Lacks real-world testing on large-scale datasets and	Review of machine learning and NLP methods for social media violence detection

					multimodal detection (image and text combined).	
11	Violence Detection in Videos by Combining 3D Convolutional Neural Networks and Support Vector Machine	Simone Accattoli, Paolo Sernani, Nicola Falcionelli, et al.	2020	High performance on benchmark datasets (e.g., Hockey Fight)	The model combines CNN and SVM but lacks performance testing in real-world surveillance scenarios and doesn't incorporate audio features which could improve detection accuracy.	3D Convolutional Neural Network (CNN) combined with Support Vector Machine (SVM)
12	Human Pose Augmentation for Facilitating Violence Detection in Videos	Ivan Calzavara	2020	Performance varies by context	DensePose may not always correctly segment violent scenarios due to overlapping body parts in fights, affecting its contribution to detection accuracy.	DensePose model for human pose augmentation combined with VioNet
13	CrimeNet: Neural Structured Learning using Vision Transformer for	Fernando J. Rendón-Segador, Juan A. Álvarez-García, Jose L. Salazar-González,	2023	99.98%	Does not reach above 90% precision on challenging datasets; false positives in production environments	Vision Transformer (ViT) and Neural Structured Learning (NSL) with adversarial

	violence detection	Tatiana Tommasi				training for violent activity recognition in videos
14	Literature Review of Deep-Learning-Based Detection of Violence in Video	Pablo Negre, Ricardo S. Alonso, Alfonso González-Briones, Javier Prieto, Sara Rodríguez-González	2024	Various accuracy levels based on different algorithms	Challenge in defining violence due to its ambiguity, infrequent occurrences, and limited real-case recordings	Review categorizing AI-powered violence detection methods, covering CNN, LSTM, transformer, and audio-based approaches
15	Crowd Violence Detection Using Global Motion-Compensated Lagrangian Features and Scale-Sensitive Video-Level Representation	Tobias Senst, Volker Eiselein, Alexander Kuhn, Thomas Sikora	2017	Not explicitly mentioned, but state-of-the-art performance on public benchmarks	Issues in CCTV footage analysis due to poor video quality and challenges in detecting individual behaviors in crowds	Lagrangian theory for analyzing non-local, long-term motion using motion-compensated features
16	Violence-YOLO: Enhanced GELAN Algorithm for Violence Detection	Wenbin Xu, Dingju Zhu, Renfeng Deng, Kai-Leung Yung, Andrew W. H. Ip	2024	92.6% mAP@0.5	Limited robustness in extremely complex scenes, computational costs on lightweight devices, and further improvement	YOLOv9 architecture, GELAN-C network, GhostConv, DySample, SimAM, and Focaler-IoU for real-time detection

					on specific cases of missed detection.	
17	A Novel Multi-stream Method for Violent Interaction Detection	Hongchang Li, Jing Wang, Jianjun Han, Jinmin Zhang, Yushan Yang, Yue Zhao	2020	Effective on multiple datasets, but exact value not provided.	Lacks real-time detection, struggles in varied lighting and motion conditions in real-world surveillance scenarios.	Multi-stream deep learning using RGB stream, optical flow, block-based spatial stream for feature fusion
18	Deep Neural Network for Gender-Based Violence Detection on Twitter Messages	Carlos M. Castorena, Itzel M. Abundez, Roberto Alejo, Everardo E. Granda-Gutiérrez, Eréndira Rendón, Octavio Villegas		80% AUC	Dataset is not extensive, limited focus on the nuances of context in GBV-related messages.	Deep learning neural network applied to minimally pre-processed Twitter data; CountVectorizer, TfidfVectorizer, HashingVectorizer used for feature extraction
19	A Sensor Network Approach for Violence Detection in Smart Cities Using Deep Learning	Marius Baba, Vasile Gui, Cosmin Cernazanu, Dan Pescaru	2019	State-of-the-art performance on low computational resources	Limited computational power and communication bandwidth in distributed sensor networks	Deep neural network fed with motion vector features, cascaded approach, time domain classifier for temporal and spatial information
20	Violence Detection by Pretrained Modules	Shakil Ahmed Sumon, Raihan Goni, Niyaz	2019	97.06% (ResNet50 + LSTM)	Dataset limited to specific regions (Bangladesh)	Pretrained ImageNet models (VGG16, VGG19,

	with Different Deep Learning Approaches	Bin Hashem, Tanzil Shahria, Rashedur M. Rahman				ResNet50), LSTM, spatial transformer network, CNN
21	Toward Fast and Accurate Violence Detection for Automated Video Surveillance Applications	Viktor Dénes Huszár, Vamsi Kiran Adhikarla, Imre Négyesi, Csaba Krasznay	2023	2% improvement over state-of-the-art	High computational cost for some spatio-temporal feature extraction methods	3D convolutions, pre-trained action recognition model, computationally efficient deep learning architecture (X3D-M)
22	Violence Detection in Video Using Computer Vision Techniques	E. Bermejo, O. Deniz, G. Bueno, R. Sukthankar	2011	Near 90%	High computational cost, lack of real-world generalization	Bag-of-Words, STIP, MoSIFT, SVM, Histogram Intersection Kernel
23	State-of-the-art Violence Detection Techniques in Video Surveillance Security Systems: A Systematic Review	B. Omarov, S. Narynov, Z. Zhumanov, A. Gumar, M. Khassanova	2022	Varies depending on methods, SVM 97%	Lack of real-time evaluation, insufficient dataset variety	Deep learning, conventional ML, video feature descriptors, datasets comparison
24	Real Time Violence Detection in Video	V. Machaca Arceda, K. Fernández		90.9% (Hockey Dataset),	High computational cost for complex	Violent Flow (ViF), Horn-Schunck

		Fabián, J. C. Gutiérrez		89.5% (Movies)	scenes, real-time detection challenges	optical flow, SVM classifier
25	Fast Violence Detection in Video	Oscar Deniz, Ismael Serrano, Gloria Bueno, Tae-Kyun Kim	2014	Near 90%	High computational cost; Lack of real-time applications	Extreme acceleration patterns using Radon transform on the power spectrum
26	A Comprehensive Review on Vision-based Violence Detection	Fath U Min Ullah, Mohammad S. Obaidat, Amin Ullah, et al.			Lack of focus on real-time surveillance; Limited coverage on crowded scenes	Neural networks (RNN, CNN, CNN-LSTM) and machine learning for automatic violence detection
27	A Review on State-of-the-Art Violence Detection Techniques	Muhammad Ramzan, Adnan Abid, Hikmat Ullah Khan, et al.			Lack of efficient real-time processing for embedded systems; Limited datasets	Comparison of traditional machine learning (SVM), deep learning, feature extraction techniques
28	Violence Detection using Oriented Violent Flows	Yuan Gao, Hong Liu, Xiaohu Sun, Can Wang, Yi Liu		87.5% (Hockey Fight)	Not suitable for crowded scenes	ses Oriented Violent Flows, AdaBoost, SVM
29	Detecting Violence in Video using Subclasses	Xirong Li, Yujia Huo, Jieping Xu, Qin Jin		55% (P100 on MediaEval)	Motion features dispensable, subclass-based detection challenging in unseen data	Combines visual, audio, and motion features
30	Violence Detection	Fillipe D. M. de			Misclassification of non-	Spatio-temporal

	in Video Using Spatio-Temporal Features	Souza, Guillermo C. Cháves, Eduardo A. do Valle Jr., Arnaldo de A. Araújo		Not specified	violent human actions as violent; context dependency	features, SVM, bag of visual words
31	Toward Subjective Violence Detection in Videos	Bruno Peixoto, Bahram Lavi, João Paulo Pereira Martin, Sandra Avila, Zanoni Dias, Anderson Rocha		Not explicitly mentioned	Not capturing fine-grained subjective aspects of violence	Uses two deep neural networks (3D CNN, CNN-LSTM) to classify violence based on semantic concepts like explosions, fights, and blood
32	Deep-violence: Individual Person Violent Activity Detection in Video	Anuja Naik, M. T. Gopalakrishna		Weizmann (73.1%), KTH (93.4%), Own Dataset (86.5%)	Low accuracy in real-world scenarios due to background and camera angles	Ensemble model based on Mask RCNN, Key-point detection, and LSTM for recognizing violent actions like punching and kicking
33	An Overview of Violence Detection Techniques: Current Challenges and Future Directions	Nadia Mumtaz, Naveed Ejaz, Shabana Habib, Syed Muhammad Mohsin, Prayag Tiwari, Shahab S. Band,		N/A	Limited real-time performance and lack of suitable datasets	Overview of traditional machine learning and deep learning approaches for violence detection, focusing on spatio-temporal models

		Neeraj Kumar				
34	Weapon Operating Pose Detection and Suspicious Human Activity Classification Using Skeleton Graphs	Anant Bhatt, Amit Ganatra	2022	89.09%	Needs improvements in real-time accuracy and crowd-specific scenarios	LSTM-RNN Network with Kalman filter
35	Children's Safety on YouTube: A Systematic Review	Saeed Ibrahim Alqahtani, Wael M. S. Yafooz, Abdullah Alsaeedi, Liyakathunisa Syed, Reyadh Alluhaibi	2023	N/A	Lack of comprehensive protection mechanisms for children's online safety	Systematic review on protecting children from inappropriate YouTube content
36	Real-Time Video Anomaly Detection for Smart Surveillance	Manal Mostafa Ali	2022	94.94% (AUC)	Semi-supervised methods show potential but require refinement in reducing false positives	Deep learning with background subtraction and object detection
37	Data augmentation for fairness-aware machine learning	Ioannis Pastaltzidis, Nikolaos Dimitriou, Katherine Quezada-Tavárez,	2022	Unspecified	Overrepresentation of minority subjects in crime detection datasets	Proposes data augmentation to rebalance datasets and address algorithmic bias issues in

		Stergios Aidinlis, Thomas Marquenie, Agata Gurzawska, Dimitrios Tzovaras				law enforcement
38	A Critical Analysis on Machine Learning Techniques for Video-Based Human Activity Recognition	Shahriar Jahan, Roknuzzaman, Md Robiul Islam	2024	95.9% (CNN), varies across methods	Traditional approaches rely on handcrafted features, which are time-consuming and inefficient in dynamic environments	Review of various deep learning and machine learning models for human activity recognition
39	Multimodal Violence Detection in Videos	Bruno Peixoto, Bahram Lavi, Paolo Bestagini, Zanoni Dias, Anderson Rocha	2023	Unspecified	Subjective nature of violence makes consistent detection difficult	Uses fusion of visual and auditory features with deep learning to detect violence in video streams

FUTURE WORK

Given the promising results across various models, future research directions can focus on enhancing violence detection accuracy in complex real-world settings. A potential direction includes integrating audio cues with visual data, creating a multimodal system that can provide contextual understanding for violence detection. Additionally, optimizing models to handle large-scale, high-resolution video feeds in real-time could improve applicability in surveillance and public safety. Future studies could also explore more diverse datasets that include non-standard camera angles, low lighting, and crowded environments, which pose unique challenges in video analysis. Lastly, developing lightweight, efficient algorithms that can operate on mobile or edge devices may expand practical applications in public surveillance and safety measures.

CONCLUSION

This study highlights various advanced deep learning and machine learning methods for detecting violent actions in video content, demonstrating significant progress in accuracy and efficiency. Approaches utilizing CNNs, RNNs, and hybrid models have shown high performance in experimental settings, yet there remains a gap in practical deployment. Despite achieving high accuracy in controlled datasets, models often struggle with real-world complexities, such as variable lighting and non-uniform camera positions. The integration of advanced preprocessing techniques, enhanced feature extraction, and real-time adaptability represents a promising pathway for future research. This paper underscores the critical role that automated violence detection can play in enhancing public safety and operational efficiency in surveillance systems.

REFERENCES

1. Y. Gao, H. Liu, X. Sun, C. Wang, and Y. Liu, "Violence detection using Oriented VIolent Flows," *Image and Vision Computing*, vol. 48–49, pp. 37–41, 2016. doi: 10.1016/j.imavis.2016.01.006 (12m)
2. X. Li, Y. Huo, J. Xu, and Q. Jin, "Detecting Violence in Video using Subclasses," *arXiv preprint*, arXiv:1604.08088, 2016(13m).
3. F. D. M. de Souza, G. C. Chávez, E. A. do Valle Jr., and A. A. Araújo, "Violence Detection in Video Using Spatio-Temporal Features," in *Proceedings of the SIBGRAPI Conference on Graphics, Patterns and Images*, 2010, pp. 224–231. doi: 10.1109/SIBGRAPI.2010.38(14m).
4. E. Bermejo, O. Deniz, G. Bueno, and R. Sukthankar, "Violence Detection in Video Using Computer Vision Techniques," E.T.S.I.Industriales, Universidad de Castilla-La Mancha, Ciudad Real, Spain, 2011.
5. B. Omarov, S. Narynov, Z. Zhumanov, A. Gumar, and M. Khassanova, "State-of-the-art violence detection techniques in video surveillance security systems: a systematic review," PeerJ Comput. Sci., vol. 8, p. e920, 2022.
6. V. E. Machaca Arceda, K. F. Fabián, and J. C. Gutiérrez, "Real Time Violence Detection in Video," Conference Paper, 2016.
7. O. Deniz, I. Serrano, G. Bueno, and T.-K. Kim, "Fast Violence Detection in Video," Proc. 9th Int. Conf. on Computer Vision Theory and Applications (VISAPP-2014), pp. 478-485, 2014, doi:10.5220/0004695104780485.
8. F. U. M. Ullah, M. S. Obaidat, A. Ullah, K. Muhammad, and S. W. Baik, "A Comprehensive Review on Vision-based Violence Detection in Surveillance Videos," ACM Comput. Surv., vol. 54, no. 1, pp. 1-35, Jan. 2022, doi:10.1145/3561971.
9. M. Ramzan et al., "A Review on State-of-the-Art Violence Detection Techniques," IEEE Access, vol. 7, pp. 107560-107575, Aug. 2019, doi:10.1109/ACCESS.2019.2932114.
10. F. J. Rendón-Segador, J. A. Álvarez-García, J. L. Salazar-González, and T. Tommasi, "CrimeNet: Neural Structured Learning using Vision Transformer for violence detection," Neural Networks, vol. 161, pp. 318–329, 2023, doi: 10.1016/j.neunet.2023.01.048.
11. P. Negre, R. S. Alonso, A. González-Briones, J. Prieto, and S. Rodríguez-González, "Literature Review of Deep-Learning-Based Detection of Violence in Video," Sensors, vol. 24, no. 4016, 2024, doi: 10.3390/s24124016.

12. T. Senst, V. Eiselein, A. Kuhn, and T. Sikora, "Crowd Violence Detection Using Global Motion-Compensated Lagrangian Features and Scale-Sensitive Video-Level Representation," IEEE Transactions on Information Forensics and Security, vol. 12, no. 7, pp. 1604-1616, 2017, doi: 10.1109/TIFS.2017.2725820.
13. A. Traoré and M. A. Akhloufi, "Violence Detection in Videos Using Deep Recurrent and Convolutional Neural Networks," *arXiv preprint arXiv:2409.07581*, 2024.
14. S. U. Khan, I. U. Haq, S. Rho, S. W. Baik, and M. Y. Lee, "Cover the Violence: A Novel Deep-Learning-Based Approach Towards Violence-Detection in Movies," *Applied Sciences*, vol. 9, no. 4963, pp. 1-11, Nov. 2019.
15. M. Biswas, A. H. Jibon, M. Kabir, K. Mohima, R. Sinthy, and M. S. Islam, "State-of-the-Art Violence Detection Techniques: A Review," *Asian Journal of Research in Computer Science*, vol. 13, no. 1, pp. 29-42, Jan. 2022.
16. P. Zhou, Q. Ding, H. Luo, and X. Hou, "Violence detection in surveillance video using low-level features," *PLOS ONE*, vol. 13, no. 10, p. e0203668, 2018.
17. M. S. M. Shubber and Z. T. M. Al-Ta'i, "Detection of Violence Behavior using Deep Learning Technique," *Academic Science Journal*, vol. 1, no. 2, pp. 188-192, 2022.
18. I. M. Abundez, R. Alejo, F. Primero, E. E. Granda-Gutiérrez, O. Portillo-Rodríguez, and J. A. A. Velázquez, "Threshold Active Learning Approach for Physical Violence Detection on Images Obtained from Video (Frame-Level) Using Pre-Trained Deep Learning Neural Network Models," *Algorithms*, vol. 17, p. 316, 2024.
19. S. A. Jebur, K. A. Hussein, H. K. Hoomod, and L. Alzubaidi, "Novel Deep Feature Fusion Framework for Multi-Scenario Violence Detection," *Computers*, vol. 12, no. 9, p. 175, 2023.
20. P. Sernani, N. Falcioni, S. Tomassini, P. Contardo, and A. F. Dragoni, "Deep Learning for Automatic Violence Detection: Tests on the AIRTLab Dataset," *IEEE Access*, vol. 9, pp. 160580-160593, 2021
21. S. A. Arun Akash, S. S. Moorthy, E. K., and N. Nathiya, "Human Violence Detection Using Deep Learning Techniques," *Journal of Physics: Conference Series*, vol. 2318, no. 1, 2022.
22. U. Dikwatta and T.G.I. Fernando, "Violence Detection in Social Media-Review," *Vidyodaya Journal of Science*, vol. 22, no. 2, pp. 7-16, 2019.
23. S. Accattoli, P. Sernani, N. Falcionelli, D. N. Mekuria, and A. F. Dragoni, "Violence Detection in Videos by Combining 3D Convolutional Neural Networks and Support Vector Machines," *Applied Artificial Intelligence*, vol. 34, no. 4, pp. 329-344, 2020.
24. I. Calzavara, "Human Pose Augmentation for Facilitating Violence Detection in Videos," Mid Sweden University, 2020.
25. B. Peixoto, B. Lavi, J. P. P. Martin, S. Avila, Z. Dias, and A. Rocha, "Toward Subjective Violence Detection in Videos," in Proc. ICASSP, 2019, pp. 1-5. doi: 10.1109/ICASSP.2019.8682833.
26. Naik and M. T. Gopalakrishna, "Deep-violence: individual person violent activity detection in video," Multimedia Tools Appl., vol. 80, no. 13, pp. 1-22, May 2021. doi: 10.1007/s11042-021-10682-w.
27. N. Mumtaz, N. Ejaz, S. Habib, S. M. Mohsin, P. Tiwari, S. S. Band, and N. Kumar, "An Overview of Violence Detection Techniques: Current Challenges and Future Directions," arXiv preprint arXiv:2209.11680, Sep. 2022. Available: https://arxiv.org/abs/2209.11680.
28. M. Baba, V. Gui, C. Cernazanu, and D. Pescaru, "A Sensor Network Approach for Violence Detection in Smart Cities Using Deep Learning," Sensors, vol. 19, no. 7, p. 1676, 2019, doi: 10.3390/s19071676.

29. S. A. Sumon, R. Goni, N. B. Hashem, T. Shahria, and R. M. Rahman, "Violence Detection by Pretrained Modules with Different Deep Learning Approaches," Vietnam Journal of Computer Science, vol. 7, no. 1, pp. 19-40, 2019, doi: 10.1142/S2196888820500013.
30. V. D. Huszár, V. K. Adhikarla, I. Négyesi, and C. Krasznay, "Toward Fast and Accurate Violence Detection for Automated Video Surveillance Applications," IEEE Access, vol. 11, pp. 18772-18785, 2023, doi: 10.1109/ACCESS.2023.3245521.
31. A. Bhatt and A. Ganatra, "Weapon Operating Pose Detection and Suspicious Human Activity Classification Using Skeleton Graphs," *Mathematical Biosciences & Engineering*, vol. 20, no. 2, pp. 2669–2690, 2022. doi: 10.3934/mbe.2023125.
32. S. I. Alqahtani, W. M. S. Yafooz, A. Alsaeedi, L. Syed, and R. Alluhaibi, "Children's Safety on YouTube: A Systematic Review," *Applied Sciences*, vol. 13, no. 6, p. 4044, 2023. doi: 10.3390/app13064044.
33. M. M. Ali, "Real-Time Video Anomaly Detection for Smart Surveillance," *IET Image Processing*, vol. 17, no. 5, pp. 1375-1388, 2022. doi: 10.1049/ipr2.12720.
34. I. Pastaltzidis et al., "Data augmentation for fairness-aware machine learning: Preventing algorithmic bias in law enforcement systems," in *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, Seoul, Republic of Korea, 2022, pp. 1-13. doi: 10.1145/3531146.3534644.
35. S. Jahan, R. Roknuzzaman, and M. R. Islam, "A critical analysis on machine learning techniques for video-based human activity recognition of surveillance systems: A review," *arXiv preprint*, vol. 2409.00731, Sep. 2024. [Online]. Available: https://arxiv.org/abs/2409.00731.
36. B. Peixoto, B. Lavi, P. Bestagini, Z. Dias, and A. Rocha, "Multimodal violence detection in videos," in *Proceedings of IEEE International Conference on Computer Vision*, 2023.
37. Machaca Arceda, V. E., Fabián, K. F., & Gutiérrez, J. C. "Real Time Violence Detection in Video," Conf. Paper, 2016.
38. Deniz, O., Serrano, I., Bueno, G., & Kim, T.-K. "Fast Violence Detection in Video," *Proc. VISAPP*, 2014. doi: 10.5220/0004695104780485.
39. Ullah, F. U. M., Obaidat, M. S., Ullah, A., & Baik, S. W. "A Comprehensive Review on Vision-based Violence Detection," *ACM Comput. Surv.*, vol. 54, no. 1, 2022. doi: 10.1145/3561971