A Review on state-of-the-art Violence Detection Techniques

ABSTRACT

Surveillance systems are playing a significant role in law enforcement and city safety. It is important to detect violent and suspicious behaviors automatically in video surveillance scenarios, for instance, railway stations, schools, hospitals to avoid any casualties which could cause social, economic, and ecological damage. Automatic detection of violence for quick actions is very significant and can efficiently help law enforcement departments. So, researchers are doing a lot of research on different techniques for detecting violence. This research study reviews various techniques and methods for detecting violent or anomalous activities from surveillance video that have been proposed by many researchers in recent years. The method of detection is divided into three categories. These categories are based on the classification techniques used. These categories are: traditional violence detection using machine learning, Support Vector Machine (SVM) & Deep Learning. Feature extraction & Object detection techniques are also described for each category. Moreover, dataset & video features that help in the recognition process are also discussed. The overall research finding has been discussed which will help the researcher in their future work in this field.

Keywords: *Surveillance system, Violence detection, Machine learning, Deep learning, Anomaly Detection, Object Detection*

1. INTRODUCTION

In different countries around the world, violence is just an everyday incident. As per the ActionAid survey in 2016 [1][3], street harassment is a prevalent problem and observed that 79% of Indian women, 86% of Thailand women, 89% of Brazil women, and 75% of London women are subjected to harassment or violence publicly. Law enforcement can't stop these incidents from happening as they are unable to take immediate action. The government has already installed a huge amount of surveillance cameras in public places like roads, stations, ports, and schools. But monitoring these surveillance feeds 24 hours is a tiresome job for humans.

Real-world anomalous [4] events are complicated and diverse. It's quite difficult to list all of the possible anomalous events and violence. If it's possible to detect this violence automatically using machine learning and alert the authorities in real-time then the crime rate can be controlled effectively. The loss of lives and properties can be minimized if violence can be detected right away when it happens. The effectiveness of anomalous event detectors is measured by the speed of response and the accuracy and the generality over different kinds of video sources with different formats.

Generally, anomalous events rarely happen as compared to normal day-to-day activities. Therefore, to alleviate the waste of labor and time, developing intelligent computer vision algorithms [4] for automated anomaly detection from video may be a pressing need. As per the researcher, 99% of the generated footage is never watched [6] and hence, detection of suspicious activity is a very difficult task. Once an anomaly is detected, the system should be able to categorize it using classification [7] techniques.

There are a lot of different techniques and methodologies available to detect violent and suspicious behaviors. These methods work with different attributes of videos. The process is done in several steps. The first step is extracting frames from the video, secondly detecting objects, then extracting features of those objects, after that classifies those features [11], and lastly detects the anomaly. These steps vary on different methods and techniques. The basic steps of detecting violence are shown in fig. 1. This process is different among the researchers to increase the accuracy and efficiency of the detection process. We have discussed different methods of violence detection using computer vision.

A lot of research has been done in this field over the past years. We aim to explore this field and present a summarized report in this paper. So, we need to gather all the relevant studies in this field. After that we can conduct a comprehensive research study; classify, analyze & summarize all the proposed methodology. The anomalies [12] activity can be different such as: Abuse, Burglar, Explosion, Shooting, Fighting, Shoplifting, Road Accidents, Arson, Robbery, Stealing, Assault, Vandalism, etc. So, detecting one anomaly activity process can be different from another.

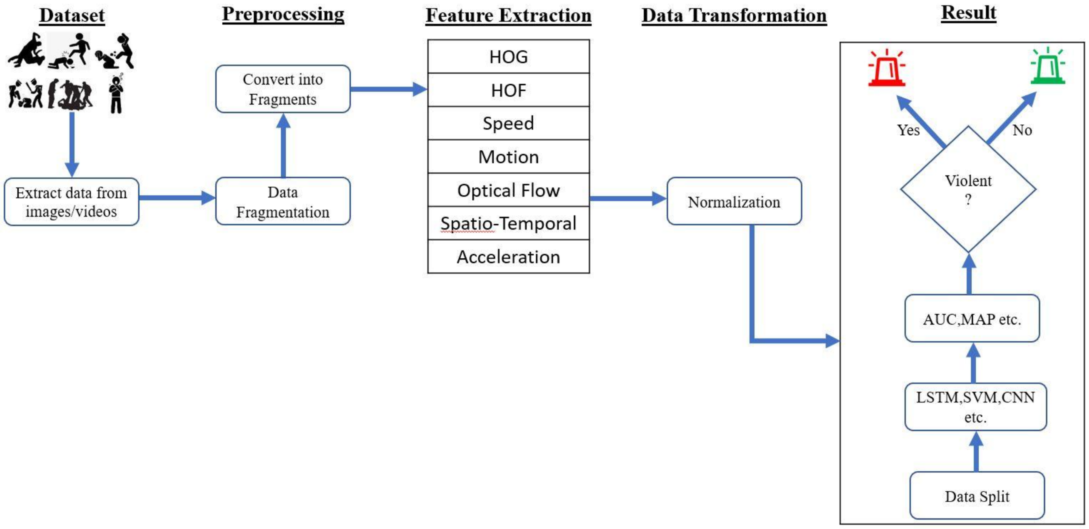

Fig 1: Basic steps of violence detection

A lot of detection model has been used over the past years. But the overall summarized detection model of all methodology is similar. The general model of violence and anomaly detection technique is shown in **Fig. 2.**

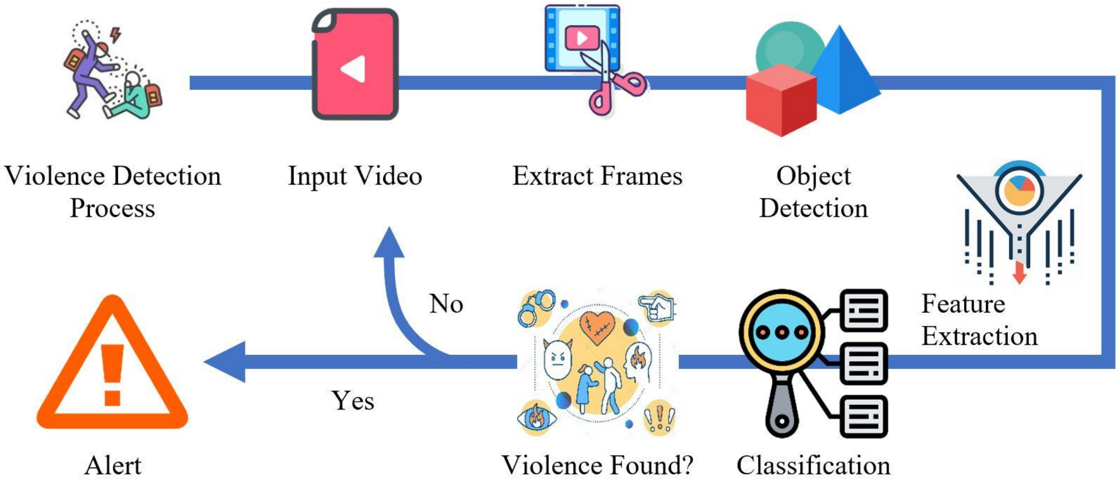

Fig 2: General model of the violence detection technique

Our research contribution and work can be summarized as follows:

- Classification of existing models into various categories.
- Describing the main features, limitations of each model.
- Reviewing the importance of video features for violence detection.
- Discussion of different datasets used mostly by the researchers.

The rest of the paper contains five chapters. Chapter 2 is about the basic concept of violence detection. The research methodology and selection process has been presented in chapter 3. Processes of violence detection are discussed briefly in chapter 4. Chapter 5 contains video features and finally chapter 6 presents the discussion of various datasets. Last chapter presents the conclusion.

2. BASIC CONCEPT

To understand the process of violence detection, we need to know the basics of violence, computer vision, and machine learning. Some basics and their descriptions are presented in **Table 1.**

S. No.	Feature	Description
1	Anomalous Events	Anomalous Events can be identified by the irregularities or deviation of the object behavior from the normal behavior including an object in an unusual location, unusual motion patterns such as movement in the wrong direction, object entry or access in restricted-area, illegal turns in traffic, violence or fighting among the human, sudden movements, dropped object or any sort of unusual event.
2	Computer Vision	A process to understand the visual world using videos and images, then identifying the object or event with the help of deep learning models.
3	Video Frame	A still image of a video is known as a frame. There are 24 frames or more in a one-second-long video.
4	Optical Flow	It's a method for calculating the motion of image intensities. To track an individual feature of a video, optical flow is used.
5	Movement	Changing the position of an object in the video.
6	Spatiotemporal	Spatial means space and temporal refers to time. Spatiotemporal refers to the time and space of an object.
7	RGB	RGB or Red, Green & Blue are three basic colors that form all the different colors.
8	Violence	Activity that is different from normal activity & contains fighting, beating, stealing is known as violence.
9	Acceleration of Images	Change of speed or velocity over the time unit. The field of acceleration consists of two directions x & y.

Table 1: Basic concept of violence detection

Researchers use a wide variety of processes and models to recognize an object from a video stream. The method could be different but the basic structure is always the same.

3. RESEARCH METHODOLOGY

This paper is about presenting the most efficient and effective methods or techniques of violence recognition process available. Classifying the process and review the result of those methods. This study is based on the research paper of researchers collected from various publication sites. These papers had to fill some predefined criteria before being included in this study.

A. Data Acquisition & Selection:

The basic target of this paper is to summarize the available methodology in the field of violence and anomalies detection from videos. Then find out the most effective process by comparing the efficiency of each method. To collect relevant studies and papers, a well-organized and sorted search has been done to extract only the relevant and meaningful information from a huge amount of data.

All kinds of irrelevant studies are being filtered out to focus on the relevant field of violence detection and removing the accurate knowledge for a well-organized literature review. To find out accurate and meaningful studies, a strategic plan for searching available studies is one of the most important steps. Two types of searches should be conducted for this kind of study. We have applied both automatic and manual searches.

To find out the relevant study from the digital library, we conducted an automatic search and then a manual search. Automatic search is done by entering the basic strings of these research fields. A manual search was needed for collecting more information in the field of violence and anomaly detection. Then manually searched papers from references to present an accurate review of the study.

These searches are conducted with some major keywords of this field which results in returning the relevant and practical information from the digital library. Those keywords were researched properly to perform the most accurate and reliable research on violence and anomaly detection. These keywords and possible alternate words and the keywords research summary are presented in **Table 2.**

((violen*) OR (fight*) OR (anomal*) AND (activity) OR (event) OR (scene) OR (sequence) AND (detect*) OR (recogni*) AND (from) AND (surveillance) OR (cctv*) AND (vi*) OR (motion) AND (using) OR (through) OR (by) OR (via) AND (machine learning) OR (computer vision) OR (deep learning) OR (artificial intelligence)).

Keywords	**Postfix & possible alternative words**
Violen*	Violence, Violent, etc.
Fight*	Fight, Fighting, etc.
Anomal*	Anomalies, Anomaly, Anomalous, etc.
Activity	Activity, scene, event, occurrence, sequence, etc.
Detect*	Detection, Detect, Detected, etc.
Recogni*	Recognition, Recognized, etc.
CCTV*	CCTV, Security Camera, Surveillance, etc.
Vi*	Video, Visual, Visualization, etc.
Using	Using, by, through, via, etc.
Artificial Intelligence	Artificial Intelligence, Machine learning, Computer vision, Deep learning, etc.

Table 2: Keyword researching summary

These keywords are used to perform an automated and manual search in the digital data library and websites to write an appropriate review and discover comprehensive data. Various journal publications, conference papers, and other information are collected from the following websites:

I. IEEE Xplore
II. Google Scholar
III. Science-Direct
IV. ACM

Some prerequisites were set to extract the most relevant study in the field of violence detection techniques. The following prerequisites were set to find out the publication from different digital data libraries:

I. Publication year range should be January 2019 to April 2021.
II. Papers that match the search string are included only.
III. Full-length papers are selected only.
IV. Paper written in English is selected for this study and all the papers are excluded which are written in English.

We have got 170 relevant publications that match all the prerequisites from Google Scholar, 72 from IEEE Xplore, 22 from Science-Direct, and 7 results from ACM. After analyzing the search result and title, abstract and result of those papers, and then filtering out those papers that don't match with the prerequisite. These papers are selected based on the abstract, proposed methodology, video is used as input for detection techniques, different feature extraction techniques, and accuracy of the detection. After going through every step, we have finalized 20 papers for the final study.

Our search techniques & paper selection steps are shown in **Figure 3.**

Search results, paper filtering steps, and selection of paper in every step are presented in **Table 3.**

Searching Source	**Automated Search**	**1st Step Filter**	**2nd Step Filter**	**3rd Step Filter**	**Manual Search**	**Finalized**
IEEE Xplore	72	22	13	6	1	7
Google Scholar	170	33	18	9	1	10
Science-Direct	22	6	4	2	0	2
ACM	7	4	2	1	0	1
Total	271	65	36	18	2	20

Table 3: Paper Selection process

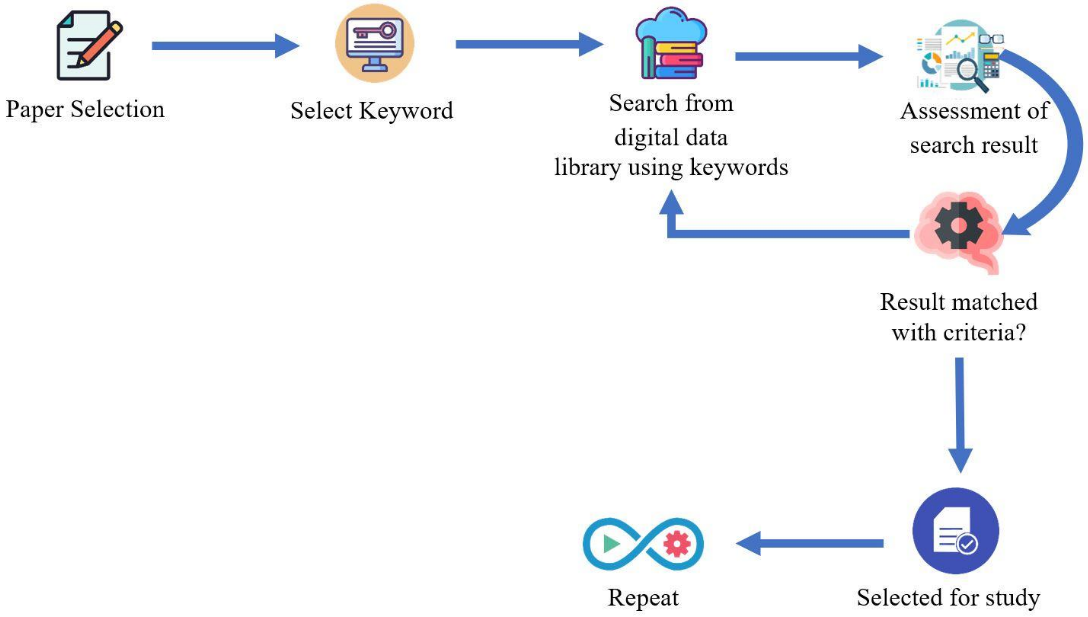
Paper Selection
Select Keyword
Search from digital data library using keywords
Assessment of search result
Result matched with criteria?
Selected for study
Repeat

Fig. 3: Paper Searching strategy

Figure 4. is about the summarized paper selection process, steps, and filtering criteria of every paper.

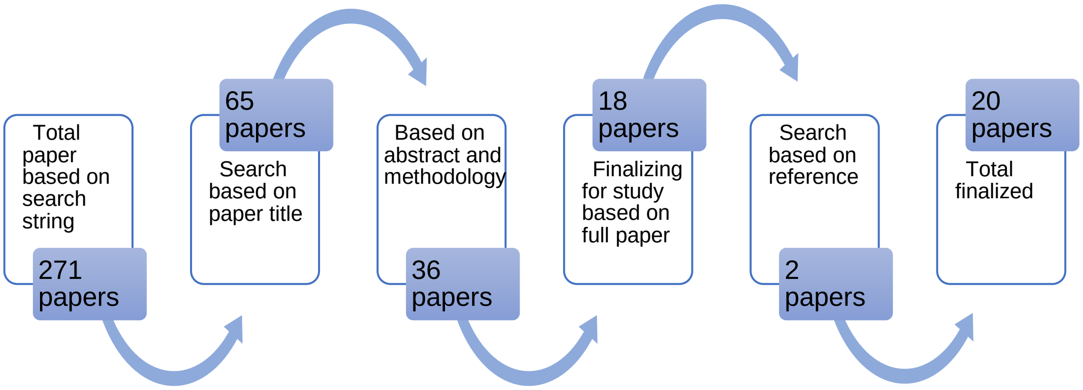
Total paper based on search string
271 papers
65 papers
Search based on paper title
Based on abstract and methodology
36 papers
18 papers
Finalizing for study based on full paper
Search based on reference
2 papers
20 papers
Total finalized

Fig. 4: Paper Selection Process

4. REVIEW OF VIOLENCE DETECTION TECHNIQUES

4.1. Multiple Anomalous Activity Detection:

To locate unusual violent behaviors in surveillance cameras is the main task but it requires manpower that works 24/7. That's why this proposed framework consists of two methods that can ease the process. First is determining object direction by measuring if the objects in videos show any kind of movements and the second one is to locate if those movements are usual or non-usual. It is capable of detecting multiple activities in a single video and also can perform behavior understanding. If any kind of anomaly is detected, it will give an alert to the system. Gaussian Mixture Model (GMM) [20] is used for object detection in this framework. The main components that are used in this method are preprocessing phase, feature extraction phase, and recognition phase. In the first phase-the preprocessing, the moving object is detected and noise removal is done. Different features like centroid, movement, speed, direction and dimensions are calculated in the feature extraction phase, and in the last phase, the Rule-based classification method is used to classify the activities of the input video and the alarm is generated for every suspicious activity. This framework suggests that the Rule-Based approach works very handy in detecting multiple activities in a single video.

4.2. Real-Time Anomaly Recognition Using Neural Networks:

This framework proposes to use different Deep Learning models to detect high movement in the videos. This framework categorized the videos into segments. Then use deep learning models like CNN and RNN to identify high movement in the frame. This model applies transfer learning [21] as a widely used object identification model. It has trained 6 variations of the approach by comparing different parameters and refining the dataset. The output layer of the RNN classifies the entire dataset into two categories such as threat and safe. The anomalies which are being considered for this model are Abuse, Arrest, Assault. For increasing the accuracy of this model, eight more classes of anomalies: Road Accidents, Burglar, Explosion, Shooting, Fighting, Shoplifting, Robbery, Stealing, and Vandalism are also added to the dataset. The final model has been tested on the self-collected dataset to look at its application in real-time scenarios. The overall accuracy of this model is about 97.23% with reduced overfitting.

4.3. Autocorrelation of gradients-based violence detection:

This framework utilized the Spatio-temporal autocorrelation of gradient-based features to effectively detect violent activities in crowded scenes. To recognize violent behavior a discriminative classifier is used. This method has used Spatio Temporal Autocorrelation of Gradients (STACOG) as handcrafted features to find out the characteristics of violent or anomalous activities from surveillance videos. It consists of two phases: (1) Extraction of Gradient-based auto-correlation (STACOG) [22] features; (2) Discriminative learning of violent/non-violent activities with an SVM Classifier. The Crowd Violence Dataset [23] is used here. The presented methodology reduces the dimension of the activity representation which inevitably reduces the computation time of features. When experiments are being performed on the Crowd Violence and Hockey Fights benchmark data sets, the efficiency, and effectiveness of the proposed methodology have been proven in comparison to any other feature-based representation of state-of-the-art approaches.

4.4. Anomaly detection using bidirectional prediction:

The approaches made in this video are categorized into two sections: probability estimation and frame generation. Anomaly detection approaches based on frame prediction [24,25,30,31] usually use a few previous frames to predict the target frame. Compared with the frame reconstruction approaches [26–31], the frame prediction approaches consider the anomaly not only in appearance and location but also in motion. This method proposes a new loss function that combines the cross-mean square error MSEM as well as the traditional mean square error MSEF and MSEB. The proposed anomaly detection model proposes a simple and elegant structure with fewer parameters, which proves beneficial for network training and easy to be combined with other algorithms and training tricks. Experimental results showed that this model achieves great performance on different datasets, outperforming the most state-of-the-art competing algorithms in detection accuracy, and has good adaptability to different datasets.

4.5. Brutality Identification using Haar Cascade Algorithm:

In this framework, the Haar Cascade algorithm is used. This method trains first motion vector images by using a machine learning algorithm. The OpenCV can be trained the Haar-cascade utility. Using multithreading on the training process reduces the time taken for the training of the data set. This method can discriminate fighting scenes with high true positive rates. The final step is to create a fragment in a straightforward technique and to apply the filter to reduce the noise and to get the final accurate detection of the fragment. As a result, the early detection method is 84.6% accurate when the fight is started.

4.6. Violence Detection using Computer Vision and Machine Learning Techniques:

A framework has been proposed in this methodology that can detect violent action in sensitive areas by using machine learning and computer vision techniques. It applies six different modules upon the captured videos- Slicing the frames according to motion tracker, Optical flow calculation, mean calculation of magnitude change vector, Histogram formation, ViF descriptor calculation, and Final decision on violence. The violent flow descriptors for each video are applied to different machine learning techniques to make a final decision about violent events. Different machine learning techniques are used here. Linear Support Vector Machine (Linear SVM), Cubic SVM, Quadratic SVM, Logistic Regression, Random Forest (RF), Bagging Trees, and Adaboost. For further accuracy, it uses a weighted averaging method on feasible machine learning techniques. Also, a face detection module can be used to identify the people involved in the violent act.

4.7. Real-world Anomaly Detection:

A small step towards addressing anomaly detection is to develop algorithms to detect a specific anomalous event, for example, a violence detector [32] and a traffic accident detector [33, 34]. The framework proposes an approach that divided the videos into a fixed number of segments during the dataset training. These segments then make those instances into two categories one positive (anomalous) and second negative (normal) and train the anomaly detection model using the proposed deep MIL ranking loss. To check the performance of the proposed approach, a new large-scale anomaly dataset that is consisting a variety of real-world anomalies is introduced. The experimental results on this dataset show that the proposed anomaly detection approach performs significantly better than baseline methods.

4.8. Fight Detection in Video Sequences:

This study uses a deep learning model based on multi-stream and high-level hand-crafted descriptors. In this framework, the main focus is to use a multi-stream of VGG-16 streams to cope with the binary problem and uses the deep neural network to verify the existence of fight detection in videos. A four-stream model can provide the information of which features are relevant to be considered during a binary fight detection problem. In this framework, the method is evaluated on Hockey Fight Dataset [35] and Movie Fight Dataset [35].

4.9. Violence Detection using ConvNets:

Most of the time 99% of the generated footage is never watched [36] and because of that detection of suspicious activity is a very difficult task [37], [38]. Researchers are now using computer vision techniques to recognize human activity [39]– [41] and monitor the crowd automatically. The main aim is to present state-of-the-art research in the field of violence detection specifically using CNN. The advantages and shortcomings are pointed out by researching the various CNNs used for violence detection in videos. CNN. The advantages and shortcomings of various CNNs are shown here. The popular available datasets are used to evaluate the proposed models.

4.10. Video Anomaly Detection using Inflated 3D Convolution Network:

Computer vision [42] is the technology for building artificial systems [55] and it contains methods for data analysis and pattern recognition. It is the process of extracting meaningful information from images and videos and making decisions based on this information. This framework suggests a video surveillance anomaly detection algorithm that works on a weekly labeled dataset. For feature extraction, I3D-Resnet-50 [43] (Two-Stream Inflated 3D ConvNet (I3D)-Resnet-50) [56,57] model is used, which is pre-trained on the Kinetics dataset. Kinetics [10] is a huge top-quality video dataset that is extracted from YouTube with approximately 650,000 videos that cover around 7000 classes. I3D-Resnet-50 [44] is a Spatio-temporal feature extractor that gives it a better performance.

4.11. Deep Network using Transfer Learning:

This research work selected GoogleNet due to deep network architecture with 12 times fewer parameters than AlexNet as a pre-trained model. It uses Hockey and Movies datasets by using transfer learning for creating a deep representation classifier [58] for violent scenes detection. It is also trained using a 10-fold cross-validation scheme, by developing a dataset images pipeline with images resizing, as input to fine-tuning network for each distinct fold. Results show the highest accuracies of 99.28% and 99.97% on Hockey and Movies datasets respectively. The proposed strategy specifically improved Hockey dataset accuracy by learning to generalize deep features for abrupt camera motion sequences, as compared to the old techniques. It can be seen that the proposed approach is outperforming on both datasets as compared to all competitive state-of-the-art published approaches from hand-crafted and deep learning domains parameters.

4.12. Real-World Fight detection:

This framework proposed a pipeline for fight detection and the results show that the use of explicit motion information (Optical Flows) has a major positive impact on performance. It is significantly superior to the RGB-only methods. Also, this framework can leverage the information coming from non-CCTV fights, through a 2-tiered model which can generalize better for the CCTV source. Better use of the sequential information at the prediction stage is another interesting aspect since the LSTM [59] failed to leverage this information. Also, it is possible to design Early Detection methods for this scenario as well, considering the importance of quickly detecting that the fight has started.

4.13. Hough Forests and 2D Convolutional Neural Network:

This framework proposed a hybrid approach. The rich Spatio-temporal voting information from Hough Forests classifier is used to leverage the representative image for each sequence, which is fed with BRISK features that capture motion and appearance from a video sequence. It demonstrates superiority over different ‘handcrafted” features [60] and 3D Convolutional Neural Network approaches for this binary recognition task. It provides the best absolute accuracy in two of the three considered datasets with 99%, 94.6%, and 91.4% accuracies in the Movies, Hockey, and Behave datasets.

4.14. Big Data Analysis and Deep Learning through Bidirectional LSTM:

This proposed framework uses the HOG function to extract the features from the input video in the Spark environment. The frames are labeled into three models. A bidirectional LSTM network is trained to recognize the anomalous events in those models. This system has a 94.5% of accuracy in violence detection rate.

4.15. Local Distinguishability Aggrandizing Network for Human Anomaly Detection:

The researcher of Local distinguishability aggrandizing network for human anomaly detection analyses how a local distinguishability aggrandizing network (LDA-Net) detects and locates human anomalies in a supervised manner. To validate the effectiveness of LDA-Net, it compares the method with state–of–the–art methods on the UCSDped2 and subway datasets. For the LDA-Net analysis result, this framework uses frame-level and pixel-level evaluations, and both of them comparisons with state-of-the-art methods on UCSD Ped2 in terms of EER and AUC. The accuracy of Frame-level 97% and Pixel-level 92%.

4.16. Deep Learning-based Methods for Video Anomaly Detection:

A comprehensive review of deep learning-based methods [61] for video anomaly detection has been presented in this work. The existing deep learning methods are evaluated in terms of datasets, performance metrics, qualitative analysis. The challenges available in deep learning approaches are outlined here.

4.17. Violent Event Detection in Visual Surveillance:

In this proposed framework, both multiple and monocular cameras were used to detect violence. BEHAVE datasets were used to compare single-temporal frameworks with multi-temporal frameworks. The experiments show that the proposed multi-temporal framework produces highly accurate results. The semantic-modeling bases (SGT) were used for visual surveillance.

4.18. Multi- stream CNN for violence detection using handcrafted features:

A novel multi-stream CNN is used in this proposed framework to detect abnormal behavior between persons. Various feature descriptors using texture motion [52] and shapes are extracted by handcrafted methods. This proposed architecture is consisted of two handcrafted and deep learning parts for feature extraction and data classification. The CNN network predicts all the input frames of datasets and the accuracy is approximately 100% for both crowded and uncrowded environments.

4.19. CNN- BiLSTM Model for Violence Detection:

A Convolutional Neural Network Bidirectional LSTM [53] model (CNN-BiLSTM) architecture is used to predict violence from video feed. Accuracy obtained in this proposed framework is 99.27% for Hockey Fight dataset, 100% for Movie dataset & 98.64% for Violent-Flows dataset.

Object detection method, feature extraction [54] method, scene type, and accuracy of every selected literature are presented in **Table 4.** To understand the methodology and result more effortlessly.

Ref	**Object Detection Method**	**Feature Extraction Method**	**Scene Type**	**Accuracy**
[2]	Haar-Cascade Algorithm	Motion vector images are trained for the consecutive frames.	Both crowded & less crowded	84.6% using the dataset
[3]	Optical Flow method	Violent Flow (ViF) descriptors	Less crowded	90%
[4]	Pre-trained 3D ConvNet	C3D network	Crowded	23%
[5]	Optical Flow	Multi-stream of VGG-16 networks with CNN	Less Crowded	88.62%
[7]	CNN	I3D-Resnet-50	Crowded	84.28%
[8]	GoogleNet	Transfer learning	Less Crowded	99.97 using dataset
[9]	Temporal stream,3D CNN	2D-CNN architecture	Crowded	
[10]	2D Convolutional Neural Network	Hough Forests classifier	Less Crowded	99%, 94.6% and 91.4% on three dataset
[11]	Gaussian Mixture Model	Apply different formulas on the consecutive frame to extract the required feature	Less Crowded	90.71%
[12]	pre-trained model inceptionV3	Convolution Neural Network	Crowded	97.23%
[13]	Support vector machine	Gradient-based autocorrelation (STACOG)	Crowded	90% and 91% on two dataset
[14]	BDLSTM network	Histogram of Oriented Gradients	Crowded	94.5%
[15]	Bidirectional prediction network	Prediction subnetwork with U-Net	Less Crowded	90.39%
[16]	YOLO object detection	3D CNN	Less Crowded	97.88%

	algorithm			
[17]	Object Detection	HOMO descriptor	Crowded	89.3%
[18]	Temporal stream, CNN Network	Multi stream CNN	Both crowded & less crowded	Approximately 100%
[19]	Multitemporal Perception Layers	Situation graph trees (SGT) and support vector machines (SVMs)	Both crowded & less crowded	78.2%
[20]	Bidirectional LSTM	Convolutional neural network	Both crowded & less crowded	99.27%, 100% & 98.64% on three different datasets

Table 4: Summary of violence & anomalies detection techniques of selected papers

5. VIDEO FEATURES

Detecting any activity from video requires some elements. The video feature is one of the basic elements for detecting any activity from that video. The detecting process and methodology directly depend on the video features that are extracted from a video. Those features are used to analyze the pattern of the activity. In a video that contains a fight scene, the movement of the objects is faster than the normal video. Analyzing those extracted features classifies the activity of the video. In **Table 5**, we have presented all the features that are used in the selected study.

Ref.	**Extracted Video Features**
[2]	Motion and movement
[3]	Optical flow, motion tracker &magnitude change vector.
[4]	Positive & negative bag of features.
[5]	Spatial, temporal, rhythmic, and depth information.
[7]	Spatiotemporal
[9]	RGB information & temporal stream.
[10]	Spatiotemporal, motion & time.
[11]	Centroid, movement, speed, direction, and dimensions.
[13]	Spatio Temporal Autocorrelation of Gradients & Motion information.
[14]	HOG
[15]	Temporal and spatial information.
[17]	Optical flow magnitude and orientation changes.
[18]	Spatial, temporal, and spatiotemporal streams.
[19]	Feature vector and multi-temporal feature.
[20]	Temporal & spatial features.

6. DATASET

6.1. Ethical Issues

All of the datasets and information are publically available and ethically sourced. There is no ethical issue in this paper.

6.2 Dataset Used

Different researchers used different datasets. These datasets are used to evaluate the performance of the methodology. **Table 5** represents the summary of the datasets.

Dataset Name	No. of Images/Clips	Year of Release	Ref	Dataset Used in articles
Hockey	1000 Clips	2011	[45]	[3] [5] [8] [10] [13] [17] [18] [20]
Crowded	246 Clips	-	[45]	[3]
Movies	200 Clips	2011	[46]	[3] [5] [8] [10] [18] [20]
UT Interaction	20 Clips	2017	[23]	[3]
UCF	2109 Clips	2012	[47]	[7]
CCTV Fight	1000 Clips	-	[9]	[9]
Behave	4 Clips	2007	[48]	[10] [19]
Crowd Violence Dataset	246 Clips	2012	[23]	[13] [17] [18] [20]
Violent Interaction	2314 Clips	-	[49]	[14]
CUHK	30662 Frames	-	[50]	[15]
UCSD	70 Clips	-	[51]	[15]

Table 5: Summary of used datasets

7. CONCLUSION

These days the crime rate has been increased a lot. Using surveillance cameras is a must for the law enforcement section. But watching these video feeds 24/7 is almost impossible for a human being. So, detecting crime using AI has opened a new era for researchers. Throughout the last decade, a lot of research has been done in the field of violence and anomaly detection. Researchers have used a lot of detection techniques and methodologies. Many researchers also proposed some new methodologies. The basic goal of this study is to present a systematic review of violence detection techniques. Explore the methodologies used by the researchers and explain the result of each technique. Dataset, video feature, accuracy, etc. which plays important roles in detection techniques are also listed in comprehensive tables. Feature extraction, object detection, classification techniques used on different methodologies are also explored in our study. Our study Contributes to highlighting the violence and anomaly detection methods and techniques from surveillance videos.

REFERENCE

[1] Wilkinson, S., 2016. Meet The Heroic Campaigners Making Cities Safer For Women. *Global Street Harassment-Making Street Safer: Action Aid*.

[2] Teja, M.S., Reddy, M.R. and Aishwarya, R., 2020, May. Man-on-Man Brutality Identification on Video data using Haar Cascade Algorithm. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 274-278). IEEE.

[3] Singh, K., Preethi, K.Y., Sai, K.V. and Modi, C.N., 2018, December. Designing an Efficient Framework for Violence Detection in Sensitive Areas using Computer Vision and Machine Learning Techniques. In *2018 Tenth International Conference on Advanced Computing (IcoAC)* (pp. 74-79). IEEE.

[4] Sultani, W., Chen, C. and Shah, M., 2018. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6479-6488).

[5] Carneiro, S.A., da Silva, G.P., Guimaraes, S.J.F. and Pedrini, H., 2019, October. Fight detection in video sequences based on multi-stream convolutional neural networks. In *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)* (pp. 8-15). IEEE.

[6] Jain, A. and Vishwakarma, D.K., 2020, July. State-of-the-art Violence Detection using ConvNets. In *2020 International Conference on Communication and Signal Processing (ICCSP)* (pp. 0813-0817). IEEE.

[7] Koshti, D., Kamoji, S., Kalnad, N., Sreekumar, S. and Bhujbal, S., 2020, February. Video Anomaly Detection using Inflated 3D Convolution Network. In *2020 International Conference on Inventive Computation Technologies (ICICT)* (pp. 729-733). IEEE.

[8] Mumtaz, A., Sargano, A.B. and Habib, Z., 2018, December. Violence Detection in Surveillance Videos with Deep Network using Transfer Learning. In *2018 2nd European Conference on Electrical Engineering and Computer Science (EECS)* (pp. 558-563). IEEE.

[9] Perez, M., Kot, A.C. and Rocha, A., 2019, May. Detection of real-world fights in surveillance videos. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2662-2666). IEEE.

[10] Serrano, I., Deniz, O., Espinosa-Aranda, J.L. and Bueno, G., 2018. Fight recognition in video using hough forests and 2D convolutional neural network. *IEEE Transactions on Image Processing*, *27*(10), pp.4787-4797.

[11] Chaudhary, S., Khan, M.A. and Bhatnagar, C., 2018. Multiple anomalous activity detection in videos. *Procedia Computer Science*, *125*, pp.336-345.

[12] Singh, V., Singh, S. and Gupta, P., 2020. Real-Time Anomaly Recognition Through CCTV Using Neural Networks. *Procedia Computer Science*, *173*, pp.254-263.

[13] Deepak, K., Vignesh, L.K.P. and Chandrakala, S., 2020. Autocorrelation of gradients based violence detection in surveillance videos. *ICT Express*, *6*(3), pp.155-159.

[14] Fenil, E., Manogaran, G., Vivekananda, G.N., Thanjaivadivel, T., Jeeva, S. and Ahilan, A., 2019. Real-time violence detection framework for football stadium comprising big data analysis and deep learning through bidirectional LSTM. *Computer Networks*, *151*, pp.191-200.

[15] Chen, D., Wang, P., Yue, L., Zhang, Y., and Jia, T., 2020. Anomaly detection in surveillance video based on bidirectional prediction. *Image and Vision*

	Computing, *98*, p.103915.
[16]	Gong, M., Zeng, H., Xie, Y., Li, H. and Tang, Z., 2020. Local distinguishability aggrandizing network for human anomaly detection. *Neural Networks*, *122*, pp.364-373.
[17]	Mahmoodi, J. and Salajeghe, A., 2019. A classification method based on optical flow for violence detection. *Expert systems with applications*, *127*, pp.121-127.
[18]	Mohtavipour, S.M., Saeidi, M. and Arabsorkhi, A., 2021. A multi-stream CNN for deep violence detection in video sequences using handcrafted features. *The Visual Computer*, pp.1-16.
[19]	Song, D., Kim, C. and Park, S.K., 2018. A multi-temporal framework for high-level activity analysis: Violent event detection in visual surveillance. *Information Sciences*, *447*, pp.83-103.
[20]	Qu, J., Du, Q., Li, Y., Tian, L. and Xia, H., 2020. Anomaly Detection in Hyperspectral Imagery Based on Gaussian Mixture Model. *IEEE Transactions on Geoscience and Remote Sensing*.
[21]	Liu, K., Zhu, M., Fu, H., Ma, H. and Chua, T.S., 2020, October. Enhancing anomaly detection in surveillance videos with transfer learning from action recognition. In *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 4664-4668).
[22]	Deepak, K., Srivathsan, G., Roshan, S. and Chandrakala, S., 2021. Deep Multi-view Representation Learning for Video Anomaly Detection Using Spatiotemporal Autoencoders. *Circuits, Systems, and Signal Processing*, *40*(3), pp.1333-1349.
[23]	Hassner, T., Itcher, Y. and Kliper-Gross, O., 2012, June. Violent flows: Real-time detection of violent crowd behavior. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1-6). IEEE.
[24]	Chen, D., Wang, P., Yue, L., Zhang, Y. and Jia, T., 2020. Anomaly detection in surveillance video based on bidirectional prediction. *Image and Vision Computing*, *98*, p.103915.
[25]	Tang, Y., Zhao, L., Zhang, S., Gong, C., Li, G. and Yang, J., 2020. Integrating prediction and reconstruction for anomaly detection. *Pattern Recognition Letters*, *129*, pp.123-130.
[26]	Wang, D. and Wang, S., 2021. Abnormal event detection algorithm based on dual attention future frame prediction and gap fusion discrimination. *Journal of Electronic Imaging*, *30*(2), p.023009.
[27]	Sabokrou, M., Khalooei, M., Fathy, M., and Adeli, E., 2018. Adversarially learned one-class classifier for novelty detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3379-3388).
[28]	Ravanbakhsh, M., Nabi, M., Sangineto, E., Marcenaro, L., Regazzoni, C., and Sebe, N., 2017, September. Abnormal event detection in videos using generative adversarial nets. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 1577-1581). IEEE.
[29]	Akcay, S., Atapour-Abarghouei, A. and Breckon, T.P., 2018, December. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian conference on computer vision* (pp. 622-637). Springer, Cham.
[30]	Wang, L., Zhou, F., Li, Z., Zuo, W., and Tan, H., 2018, October. Abnormal event detection in videos using hybrid Spatio-temporal autoencoder. In *2018 25th IEEE International Conference on Image Processing (ICIP)* (pp. 2276-2280). IEEE.
[31]	Luo, W., Liu, W. and Gao, S., 2017, July. Remembering history with convolutional lstm for anomaly detection. In *2017 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 439-444). IEEE.
[32]	Gkountakos, K., Ioannidis, K., Tsikrika, T., Vrochidis, S. and Kompatsiaris, I., 2020, June. A Crowd Analysis Framework for Detecting Violence Scenes. In *Proceedings of the 2020 International Conference on Multimedia Retrieval* (pp. 276-280).

[33] Huang, X., He, P., Rangarajan, A. and Ranka, S., 2020. Intelligent intersection: two-stream convolutional networks for real-time near-accident detection in traffic video. *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, *6*(2), pp.1-28.

[34] Rajesh, G., Benny, A.R., Harikrishnan, A., Abraham, J.J. and John, N.P., 2020, July. A Deep Learning based Accident Detection System. In *2020 International Conference on Communication and Signal Processing (ICCSP)* (pp. 1322-1325). IEEE.

[35] Accattoli, S., Sernani, P., Falcionelli, N., Mekuria, D.N. and Dragoni, A.F., 2020. Violence detection in videos by combining 3D convolutional neural networks and support vector machines. *Applied Artificial Intelligence*, *34*(4), pp.329-344.

[36] https://vidimensions.com/, "Vi Dimensions | We Discover. Going Beyond Detection." [Online]. Available: https://vidimensions.com/. [Accessed: 10-Dec-2019].

[37] Serrano Gracia, I., Deniz Suarez, O., Bueno Garcia, G. and Kim, T.K., 2015. Fast fight detection. *PloS one*, *10*(4), p.e0120448.

[38] Deniz, O., Serrano, I., Bueno, G. and Kim, T.K., 2014, January. Fast violence detection in video. In *2014 international conference on computer vision theory and applications (VISAPP)* (Vol. 2, pp. 478-485). IEEE.

[39] Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A.M. and Wang, X., 2020. Computer vision techniques in construction: a critical review. *Archives of Computational Methods in Engineering*, pp.1-15.

[40] Verma, K.K., Singh, B.M., Mandoria, H.L. and Chauhan, P., 2020. Two-Stage Human Activity Recognition Using 2D-ConvNet. *International Journal of Interactive Multimedia & Artificial Intelligence*, *6*(2).

[41] Ladjailia, A., Bouchrika, I., Merouani, H.F., Harrati, N. and Mahfouf, Z., 2020. Human activity recognition via optical flow: decomposing activities into basic actions. *Neural Computing and Applications*, *32*(21), pp.16387-16400.

[42] Chai, J., Zeng, H., Li, A. and Ngai, E.W., 2021. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, *6*, p.100134.

[43] Guo, J., He, H., He, T., Lausen, L., Li, M., Lin, H., Shi, X., Wang, C., Xie, J., Zha, S. and Zhang, A., 2020. GluonCV and GluonNLP: Deep Learning in Computer Vision and Natural Language Processing. *J. Mach. Learn. Res.*, *21*(23), pp.1-7.

[44] Sharma, V., Gupta, M., Kumar, A. and Mishra, D., 2021. EduNet: A New Video Dataset for Understanding Human Activity in the Classroom Environment. *Sensors*, *21*(17), p.5699.

[45] Serrano Gracia I, Deniz Suarez O, Bueno Garcia G, Kim T-K (2015) Fast Fight Detection. PloS ONE 10(4): e0120448. https://doi.org/10.1371/journal.pone.0120448

[46] Ryoo, M.S., and Aggarwal, J.K., 2009, September. Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. In *2009 IEEE 12th international conference on computer vision* (pp. 1593-1600). IEEE.

[47] K. Soomro, A. R. Zamir, and M. Shah. UCF101: A dataset of 101 human actions classes from videos in the wild. arXiv preprint arXiv:1212.0402, 2012.

[48] S. Blunsden and R. B. Fisher. The BEHAVE video dataset: ground truthed video for multi-person behavior classification. Annals of the BMVA, 4(1-12):4, 2010.

[49] Zhou, P., Ding, Q., Luo, H. and Hou, X., 2017, June. Violent interaction detection in video based on deep learning. In *Journal of physics: conference series* (Vol. 844, No. 1, p. 012044). IOP Publishing.

[50]	Lu, C., Shi, J. and Jia, J., 2013. Abnormal event detection at 150 fps in matlab. In *Proceedings of the IEEE international conference on computer vision* (pp. 2720-2727).
[51]	Mahadevan, V., Li, W., Bhalodia, V. and Vasconcelos, N., 2010, June. Anomaly detection in crowded scenes. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1975-1981). IEEE.
[52]	Berlin, S.J. and John, M., 2020. Spiking neural network based on joint entropy of optical flow features for human action recognition. *The Visual Computer*, pp.1-15.
[53]	Ullah, W., Ullah, A., Haq, I.U., Muhammad, K., Sajjad, M. and Baik, S.W., 2021. CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. *Multimedia Tools and Applications*, *80*(11), pp.16979-16995.
[54]	Ma, W., Wu, Y., Cen, F. and Wang, G., 2020. Mdfn: Multi-scale deep feature learning network for object detection. *Pattern Recognition*, *100*, p.107149.
[55]	Biswas, M., Tania, M.H., Kaiser, M.S., Kabir, R., Mahmud, M. and Kemal, A.A., 2021. ACCU3RATE: A mobile health application rating scale based on user reviews. PloS one, 16(12), p.e0258050.
[56]	Hassan, M.K., Hassan, M.R., Ahmed, M.T., Sabbir, M.S.A., Ahmed, M.S. and Biswas, M., 2021. A Survey on an Intelligent System for Persons with Visual Disabilities. *Aust. J. Eng. Innov. Technol*, *3*(6), pp.97-118.
[57]	Biswas, M., Kaiser, M.S., Mahmud, M., Al Mamun, S., Hossain, M. and Rahman, M.A., 2021, September. An XAI Based Autism Detection: The Context Behind the Detection. In *International Conference on Brain Informatics* (pp. 448-459). Springer, Cham.
[58]	Biswas, M., Kaiser, M.S., Mahmud, M., Al Mamun, S., Hossain, M. and Rahman, M.A., 2021, September. An XAI Based Autism Detection: The Context Behind the Detection. In *International Conference on Brain Informatics* (pp. 448-459). Springer, Cham.
[59]	Tamanna, I., Mahi, M.J.N., Ahmed, S., Kader, M., Haque, A. and Biswas, M., 2021, August. Controlling Body Sources of Noise Generated by Niddle Electrogram Machines: A New EMG Idea for Skipping Traditional Approaches. In *2021 International Conference on Science & Contemporary Technologies (ICSCT)* (pp. 1-6). IEEE.
[60]	Mahbub, M., Biswas, M., Miah, A.M., Shahabaz, A. and Kaiser, M.S., 2021, July. Covid-19 detection using chest x-ray images with a regnet structured deep learning model. In *International Conference on Applied Intelligence and Informatics* (pp. 358-370). Springer, Cham.
[61]	Ray, B., Saha, K.K., Biswas, M. and Rahman, M.M., 2020, December. User Perspective on Usages and Privacy of eHealth Systems in Bangladesh: A Dhaka based Survey. In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)* (pp. 1-5). IEEE.