Real-time Anomaly Detection and Multi-class Classification in Surveillance Videos Using Deep Learning

Deepika Varshney
Department of Computer Science & Information Technology, Jaypee Institute of Information Technology, Noida, India
E-mail: deepikavarshney06@gmail.com

Keywords: crime, machine learning, deep learning, anomaly detection, surveillance videos, C3D

Received: January 26, 2025

With the escalating crime rates in the school and colleges, surveillance cameras have become instrumental in monitoring and detecting anomalies happened in the campus to reduce crime rates. This study proposes a novel deep learning-based approach for real-time anomaly detection and multi-class classification in surveillance videos. We employ a Convolutional 3D Neural Network (C3D) model for extracting features and detecting probabilistic anomalies within video frames. Subsequently, a Simple Recurrent Neural Network (RNN) is trained to classify detected anomalies into one of thirteen distinct categories, including Abuse, Arson, Assault, among others. Our approach outperforms previous methods, achieving an accuracy of 83.67% for anomaly classification on the UCF Crime dataset. This work demonstrates the feasibility of automatically detecting and classifying anomalies in real-world surveillance scenarios, potentially enhancing public safety and well as school/college campus safety and facilitating prompt response measures.

Povzetek: Študija predlaga realnočasno zaznavanje in razvrščanje anomalij v nadzornih videih (C3D za zaznavo, RNN za 13 razredov), ki na podatkovni zbirki UCF Crime doseže visoko točnost.

1 Introduction

In the era of information technology, the Surveillance cameras are widely being integrated into public locations to improve public safety, such as residential buildings, offices, malls, crossroads, banks, shops, market area and other places. Even there is a wide use of surveillance camera in these places, it does not tend to be beneficial for the government law enforcement agencies for monitoring abnormal activities. Due to this, there is a noticeable lack in the deployment of surveillance camera. The important key role in video surveillance is to detect abnormal occurrences such as traffic accident, crimes or unlawful activities. Anomalous activities are uncommon as compared to typical events. To overcome this, it is required to incorporate deep learning algorithms for the automatic detection of anomalous video to reduce time and manpower. The purpose of a realistic anomaly detection system is to detect and communicate activity that deviates from typical patterns in a timely manner, as well as to determine the time window in which the abnormality occurs. As a result, anomaly detection may be thought of as high-level video knowledge that separates abnormalities from typical patterns.

Once an anomaly has been identified, classification techniques may be used to categorize it into one of the specialized activities. Developing algorithms to identify a specific anomalous event, such as a violence detector [1] or a traffic accident detector [2, 3], is a tiny step toward tackling anomaly detection. However, such solutions are evident in that they cannot be applied to identify other abnormal occurrences, therefore they are only useful in theory.

Anomaly events in the real world are difficult and varied. It's impossible to make a comprehensive list of all conceivable anomalous events. As a result, it is best if the anomaly detection method doesn't rely on any prior knowledge of the occurrences. In other words, anomaly detection should be carried out with the utmost care. Approaches based on sparse coding [4, 5] are regarded as exemplary methods for achieving state-of-the-art anomaly detection outcomes. These approaches presume that just a tiny fraction of the beginning of a film has normal events, hence the beginning is utilized to construct the normal event dictionary. The primary notion behind anomaly detection is that abnormal occurrences can't be reliably recreated using the standard event dictionary. However, because the environment collected by security cameras might change dramatically over time (for example, during different times of the day), these systems result in significant false alarm rates for various routine actions. The Figure 1 Alt Text shows the examples concerning to normal videos and anomalous videos with their description. Although the ideas described above are tempting, they are predicated on the notion that any behavioral pattern that differs from the learned normal usual patterns is an aberration. This assumption, however, may not be correct because defining a typical event that encompasses all conceivable normal patterns/behaviors is extremely difficult or impossible [6]. What's more, the line between normal and abnormal conduct is frequently blurred. Furthermore, given realistic circumstances, the same behavior might be considered normal or abnormal depending on the circumstances. As a result, it is believed that normal and anomalous event training data can aid an anomaly detection system's learning. We present an anomaly detection system based on weakly labelled training movies in this research.

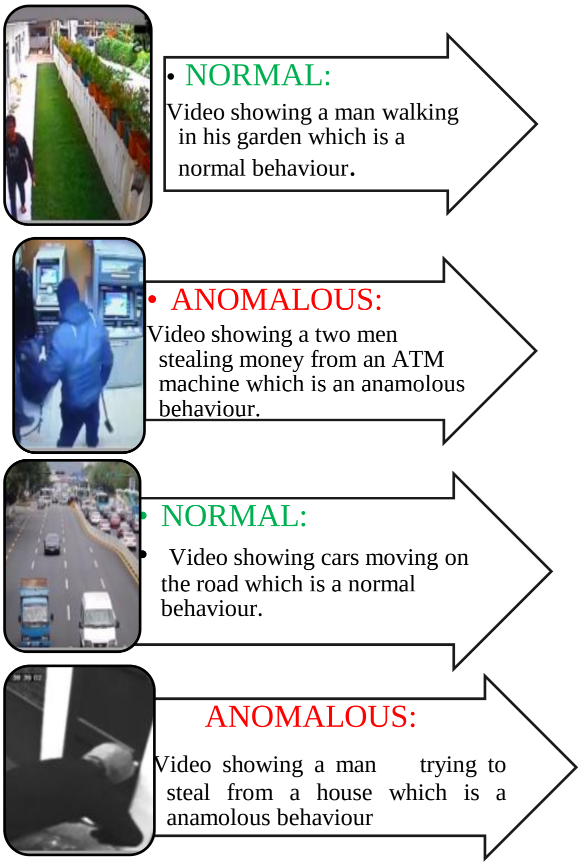
• NORMAL:
Video showing a man walking in his garden which is a normal behaviour.
• ANOMALOUS:
Video showing a two men stealing money from an ATM machine which is an anamolous behaviour.
• NORMAL:
• Video showing cars moving on the road which is a normal behaviour.
ANOMALOUS:
Video showing a man trying to steal from a house which is a anamolous behaviour

Figure 1: Normal vs anomalous videos

That is, we just know the video-level labels, which means that a video is normal or includes an abnormality someplace, but we don't know where. This is fascinating since adding video-level labels allows us to simply annotate a huge number of films [7, 8]. This proposed work is divided into two major tasks: first is to detect the anomaly in the video by measuring the probability of anomalous activity inside the frame, and the second task is to classify the type of anomalous activity (one among the 13 anomalous activities listed in dataset description i.e., Abuse, Arrest, Arson, Assault, Accident, Burglary, Explosion, Fighting, Robbery, Shooting, Stealing, Shoplifting, and Vandalism) in the video. We automatically learn an anomaly ranking algorithm based on training movies that predicts high anomaly scores for anomalous regions in a video. During testing, a lengthy untrimmed video is broken into parts and fed into our deep network, which provides an anomaly score to each video segment in order to discover anomalies. The following are the key contributions made by this study.

- The first task is to detect the anomaly in the video by measuring the probability of anomalous activity inside the frame,
- The second task is to classify the type of anomalous activity (one of the 13 anomalous activities listed in the dataset description) in the video.
- The comparative analysis has been performed with the baseline approach C3D [9] and C3D+RNN [10], on detecting 13 distinct abnormal behaviors concerning anomaly detection and the experimental findings on the UCF dataset reveal that the suggested strategy outperforms others methods.

The rest of the work discussed in the paper is structured as follows. The upcoming Section 2 discusses the detailed state of the art work that are focused on the crime rate prediction and classification. Whereas, Section 3, incorporates the detailing of dataset used, the methods adopted, and the technology used to build the crime rate detection model shown in Section 4. In Section 5, the comparative and experimental analysis as well as the results have been shown. Section 6 shows the conclusion and future scope.

2 Related work

Many Researchers have put their keen interest towards videos to solve the real time problems [30]. Recent works presented a deep learning strategy based on picture normalisation and CNN architecture for automated pedestrian detection. Their suggested architecture adaptively learns pedestrian representation to accomplish efficient recognition [11] with improved accuracy and shorter pre-processing times. Their work swiftly and properly wins distinctive pedestrian features. Images from the data set are shrunk to a fixed scale, adjusted with a unity standard deviation, and fed into the proposed DCNN to distinguish pedestrians from non-pedestrians.

In [12], authors described a technique based on theory of queuing and computer vision for detecting pedestrians and computing essential macroscopic statistics of pedestrian traffic. They used an Aggregated Channel Feature based detector to detect pedestrians in a given zone of interest in a traffic video. Their unique approach was then applied in-situ at traffic monitoring camera locations to create a distributed traffic control system. Whereas, in [13], the author has added a revertive link to the pedestrian re-recognition detector, making it more like the human cognitive process by turning a single image into an image sequence, which is then reidentified by the memory image sequence pattern. Deep learning-based pedestrian re-recognition algorithms can now recall visual sequence patterns and re-identify pedestrians in photos, thanks to their technique. Simultaneously, their study offered a shallow learning strategy based on selective dropout. The other studies incorporate quick approach for recognising pedestrians in surveillance systems with less memory size and processing units. As even compressed architecture of deep neural net needed large memory, the proposed work by [14], added a compression technique based on a teacher student framework to a random forest classifier instead of a wide and deep network.

In [15], the authors proposed that processing in device scenarios used to identify crowd anomalies. In this paper, an abnormality analysis method based on improved k-means is proposed. This approach combines mean shift with the k-means classification procedure to provide quick and accurate detection of crowd anomaly.

In [16], authors examined three aspects of this issue. To begin with, a novel global factor defines the moving crowd. This function might very well characterise the specifics of the point-of-interest in terms of spatial and temporal motion. Then, a method is used that clusters the feature point first and then measures it collectively, allowing for more coherent and efficient computation of individual groups.

The optical flow was employed as supplemental information for anomaly identification. They created a deep WCAELSTM network [17] that captures spatial variances with CAE and temporal variances with three convolutional LSTM (Conv-LSTM) units, and proposed a weighted Euclidean loss that focuses on moving foregrounds, thus limiting the influence of complex backgrounds, as well as a global-local analysis to achieve anomaly detection and localization jointly.

In this study, authors presented three contributions [18]: (1) a new structural context descriptor is proposed to characterise the structural properties of individuals in crowd scenes; (2) a self-weighted multi-view clustering method is proposed to cluster feature points by incorporating their orientation and context similarities; and (3) a novel framework for group detection is introduced.

Researchers proposed a crowd anomaly detection framework that satisfied continuous feed in of Spatio-temporal [19], information from live CCTVs. Firstly, an extraction algorithm for the spatial-temporal texture is built. Their method will efficiently strip textures from the video with ample information about crowd motion. It is by the adoption of Gabor-filtered textures with the maximum entropy values of knowledge.

Introduced a new system for global abnormality identification while using context location (CL) as well as motion-rich STVs (MRSTVs) [20] through block-level function extraction. In order to record motion properties of regular and irregular cases, the histogram of optical movement direction as well as motion magnitude features in Spatio-temporal volumes (STVs) is being used as a regional object descriptor.

The researchers concentrated on deep learning methodology for crime detection in videos. Deep learning techniques are evaluated based on their algorithms and models. The main model [21] purposed to use deep learning algorithms to identify the precise list, the individuals engaged, and the behaviour that has developed in a large crowd under any weather situation.

In this paper, Aravindan, S., E. Anusuya, and M. Ashok Kumar developed a GUI model for Chennai and (SMLT) supervised machine learning technique model [22], which finds the best model among SML algorithms by getting more than 90% accuracy for determining crime rate by PSA and year. The dataset is taken from the online Indian police department, split into training and testing in 7:3, and applied to Random Forest, SVC, KNN, Decision tree, and logistic regression algorithms. The research gap is not comparing with deep learning models and not incorporating this model into a real-time system for crime rate prediction.

Wang and Bao developed deep learning ST-Resnet model which outperforms the fully-ternary ST-Resnet, ARIMA, KNN, and HA model. The authors used the well-represented data to adjust the spatial-temporal residual network to forecast the distribution of crime in Los Angeles at the hourly scale in neighbourhood-sized blocks. These trials, along with comparisons with other prediction methods, show that the suggested model is more accurate [23]. Finally, the authors offer a ternary approach to solve the issue of resource consumption for real-world deployment. Researchers Tabedzki and Christian studied developed a machine learning technique for predicting crime-related statistics in Philadelphia, United States [24]. The problem was broken down into three sections: identifying whether or not a crime happens, the likelihood of a crime occurring, and the most likely crime occurring. To generate comprehensive quantitative crime predictions with increased relevance, algorithms including logistic regression, KNN, ordinal regression, and tree techniques were utilized to train the datasets. They also showed a map that showed different crime categories in different areas of Philadelphia for a certain time period, with different colours representing each sort of crime. Various sorts of crimes, ranging from assaults to cyber fraud, were included in order to mirror the overall trend of crime in Philadelphia over a period of time. Their system predicted the likelihood of a crime with a remarkable 69% accuracy, as well as the number of crimes ranging from one to 32% with a 47% accuracy.

The authors of [25], concentrated primarily on the study and construction of machine learning algorithms to lower crime rates in India. To discover the pattern relations between a huge amount of data, machine learning techniques were used. The study's major goal was to provide a prediction of future crime based on the prevalence of prior crime areas.

The dataset was analysed and interpreted using approaches such as Bayesian neural networks, the Levenberg Marquardt algorithm, and a scaled algorithm, with the scaled algorithm outperforming the other two techniques. A statistical analysis based on correlation, analysis of variance, and graphs revealed that the crime rate may be decreased by 78% using the scaled method [26] with an accuracy of 0.78. On the other hand, Hossain and Sohrab's study proposes a method for predicting crime based on an analysis of a dataset comprising records of already committed crimes and their trends. The suggested system is based on two machine learning algorithms: decision trees and KNNs. Many of the earlier studies also applied video surveillance based deep learning model for the detection of criminal activities [27,28]. The authors of [29], detect anomaly by examining separate memory modules for normal and anomalous patterns, we initiate the process by generating pseudo-anomalies with a temporal pseudo-anomaly synthesizer.

The prediction model's accuracy was improved using techniques like the random forest algorithm and adaptive boosting. The offenses were classified into common and rare classes to improve the model's results. The most common crimes were in the frequent class, while the least common crimes were in the rare class. The suggested system was given data on criminal activities in San Francisco during a 12-year period. The accuracy was raised to 99.16% by combining under sampling and oversampling approaches with the random forest algorithm.

The best way to highlight the novelty in our study is by comparing it with the work that was done by others and pointing out the things that our study does which was never done before. It is described in the following format for each research, first conduct a thorough literature search to identify what is already known in your field of research and what are the gaps to be explored. In some researches, the research gap is not identifying other abnormal activity like which is considered in our research and not incorporating this model into a real-time system for crime rate prediction. The authors of [31], developed the Composite Recurrent Bi-Attention (CRBA) model to detect anomalies in surveillance footage. It leverages DenseNet201 for comprehensive spatial feature extraction and utilizes BiLSTM networks to model temporal relationships between video frames. To further enhance its performance, a multi-attention mechanism is incorporated, guiding the model's attention to key spatiotemporal regions. This combination of methods allows the CRBA model to effectively tackle both spatial and temporal aspects of video data, resulting in more accurate anomaly detection. Experimental results indicate strong performance on both the University of Central Florida (UCF) dataset and the newly introduced Road Anomaly Dataset (RAD).They achieves an F1-score of 83.8% when trained on the UCF dataset. While, some other research is using different methods to classify and detect and getting lower accuracy in comparison to ours. In our study, it is clearly shows that we are detecting the abnormality in the video by measuring the probability of anomalous activity using (Convolutional 3D neural network (or C3D) model) and detection of abnormal activity using simple RNN, which has not been done before in any other research done on this dataset of UCF Crime. Also, we are testing our model on the custom videos created by us for real-time detection.

In the end our research will were able to detect crime using C3D model and detect abnormal behaviour using simple RNN model. Beyond the scope of other researches, we are also testing on our custom dataset which makes our model uniquely precise in comparison to other existing models. Real-Time Surveillance videos are able to capture a variety of realistic criminal activities. In order to minimize crime, we need to be able to detect it as soon as it happens. Crime prediction and detection is the unique ability to predict future crime by studying the trends of the past. Through this paper, the crime rate can be forecasted and hence minimized. In case any abnormal activity is detected during the live stream, there will either be a notification issued to the authorities. This notification will either be linked to mobile devices of nearby police personnel or may be linked to a hardware IOT device that we plan to develop.

3 Data description

The dataset contains videos of UCF crime dataset[1] used for Real-world Anomaly Detection in Surveillance Videos taken from Kaggle.

The dataset contains videos of UCF Crime Dataset. In data preprocessing every 10th frame is extracted from each full-length video and combined for every video in that class. All the images are of size 64*64 and in ".png" format. The dataset consists of 13 realistic anomalies including Abuse, Arrest, Arson, Assault, Road Accident, Burglary, Explosion, Fighting, Robbery, Shooting, Stealing, Shoplifting, and Vandalism. The total image count for the train subset is 1,266,345. The total image count for the test subset is 111,308. The average number of frames in this dataset is 7247. The dataset consists of 1900 long and untrimmed real-world surveillance videos. As an example of dataset Figure 2 depicts Abuse class video, Accident class and Robbery class video respectively. The complete size of dataset is 82.9 GB. With a total of 128 hours of videos from CCTV cameras.

Figure 2: (a) Abuse class video (b) Accident class video (c) Robbery class video

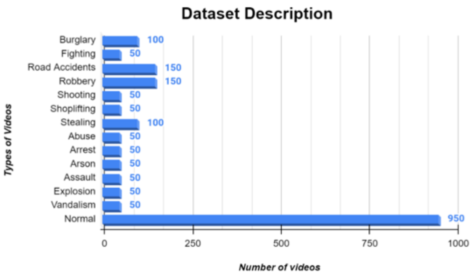

Figure 3: Bar chart depicting distribution of videos in dataset

The dataset is constructed as a new large-scale dataset, called UCF-Crime, to evaluate various method. The Figure 3 **Alt Text** gives brief description of the dataset . It consists of long untrimmed surveillance videos which

[1] https://www.kaggle.com/datasets/odins0n/ucf-crime dataset

cover 13 real world anomalies, including Abuse, Arrest, Arson, Assault, Road Accident, Burglary, Explosion, Fighting, Robbery, Shooting, Stealing, Shoplifting, and Vandalism along with their distribution. These anomalies are selected because they have a significant impact on public safety.

4 Methodology

4.1 Proposed methodology

In our study, it is clearly shows that we are detecting the abnormality in the video by measuring the probability of anomalous activity using (Convolutional 3D neural network (or C3D) model) and detection of abnormal activity using simple RNN, which has not been done before in any other research done on this dataset of UCF Crime. The flow diagram of our proposed methodology is summarized in the Figure 4. This proposed work is divided into two major tasks: first is to detect the anomaly in the video by measuring the probability of anomalous activity inside the frame, and the second task is to classify the type of anomalous activity (one among the 13 anomalous activities listed in dataset description i.e., Abuse, Arrest, Arson, Assault, Accident, Burglary, Explosion, Fighting, Robbery, Shooting, Stealing, Shoplifting, and Vandalism) in the video. The videos in the dataset have variable lengths so variable number of frames, and computing on all the available frames is a very costly task. Hence, initially we tried to reduce the frames so as to maintain the necessary details as well as reduce the cost. For training the neural networks as shown in Figure 4, we considered one-tenth of the total number of frames available in each video taken at regular interval.

4.1 Anomaly detection in the video

This task involves finding the probability of any anomalous behaviour in the video at different instances. This task was break down into supervised approach of finding whether a particular video clip has anomalous behaviour or not. For this we trained a general Convolutional 3D neural network (or C3D) model on provided labelled videos in the dataset. The C3D model was capable of extracting the features and training them against the labels. The extracted features of all the videos were saved for further classification of anomaly.

Algorithm 1: Anomaly Detection in Video

```
set m = 10
for each video in dataset:
    set num_frames = number of frames
    set req_frames = num_frames/10
    set first_frame = 1
    set last frame = 10*m+1
    while (number of clips possible):
        clip = create a clip window with m number of frames
        pass the clip to trained c3d model
        probability = model.predict(clip)
        plot probability against the frame number
        first_frame += 10
        last_frame += 10
    end while
end for
```

Now, to find the probabilities of anomalous behaviour at different instances in the video, the complete algorithm is summarized in the Algorithm 1. Each video is segmented into multiple small clips (in this proposed methodology we used 10 frames per clip). Hence, in a video having n number of frames we created n-m+1 number of clips having m frames in each clip. Each of these clips is passed as an input to the trained C3D model, which provide the probability of any anomalous behaviour in that clip.

4.2 Anomaly classification in the video

Now after the anomaly is detected in the video, the task is to classify the type of anomaly detected. Classifying the anomaly is again a supervised approach, were any machine learning or deep learning algorithm can train on the features and label. This will be a multi-class classification problem. There could be multiple possible ways of doing it like using C3D itself for multi-lass classification, or using combination of 2DCNN+RNN or features extracted using C3D+RNN. Here we used the features extracted using C3D model to train a Simple RNN model. For each video, features of all the considered frames are sequenced in order to form a vector of features.

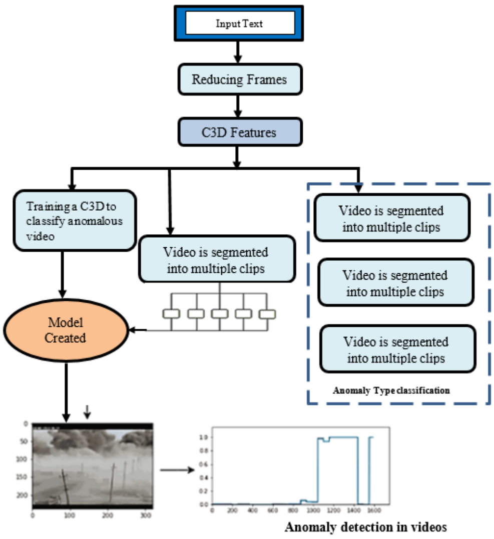
Input Text
Reducing Frames
C3D Features
Training a C3D to classify anomalous video
Video is segmented into multiple clips
Model Created
Video is segmented into multiple clips
Video is segmented into multiple clips
Video is segmented into multiple clips
Anomaly Type classification
0
50
100
150
200
0
100
200
300
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
1200
1400
1600
Anomaly detection in videos

Figure 4: The Flow Diagram of the proposed methodology

This feature vector of all the videos are grouped into a single matrix. Strings serve as the labels for the videos. Because neural networks cannot grasp string values, they must be transformed to a numerical format before being given to the model. The Simple RNN trained on features and numerical labels to perform a multiclass classification task.

5 Experimental analysis and results

With increasing number of criminal offences all across the world, it is important to monitor and keep a track of these cases so that they can be prevented in future. The paper revolves around development of a surveillance system to detect anomalous activities and inform the required authorities about them so that they can be prevented and crime rates can be controlled. To address this problem UCF dataset was used for training the C3D model to detect anomalous activities in surveillance videos. Further a simple RNN was trained to classify the anomaly into one of the 13 types as present in dataset.

5.1 Probabilistic anomaly detection results

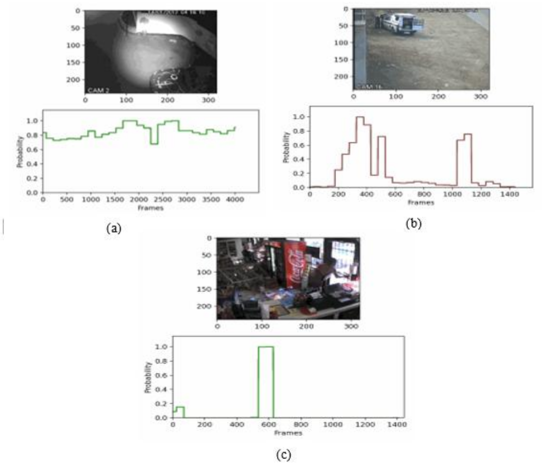

Figure 5: Detection of probabilistic anomaly

After the C3D model was trained, the model was tested on the testing set and the results obtained are depicted in the Figure 5. The X axis in the given graphs depicts the frame of the video while the Y axis depicts the probability that an anomaly has occurred in some frames of that segment. In case of normal videos, the probabilistic anomaly is found to be zero since no anomalous activity takes place in any frame. This is shown in Figure 6. Figure 5 (a), (b) and (c) Alt Text demonstrates anomalous activity and the probabilistic anomaly detected corresponding to it as shown in the graphs. The C3D model has been applied and from the experimental analysis it is depicted the anomaly with an accuracy of 84.5%. The demo outcome of the experiment you can check on the link[1]. The results are presented in three forms, first in anomalous, normal and customized videos in Figure 5, 6 and 7 respectively.

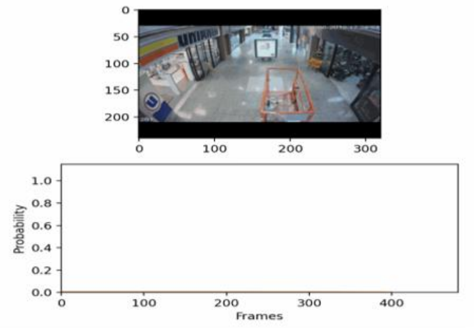

Figure 6: Detection of probabilistic anomaly in normal videos

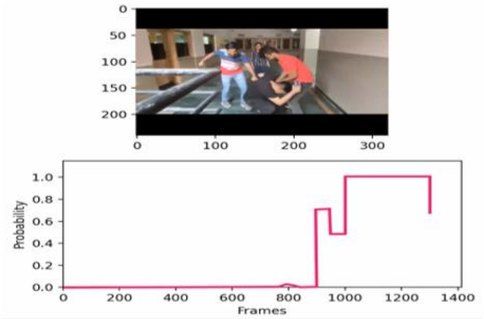

Figure 7: Detection of probabilistic anomaly in customized videos

The model was also tested on some customized videos wherein anomalous activity was performed and its detection was checked.

5.2 Multi-class classification results

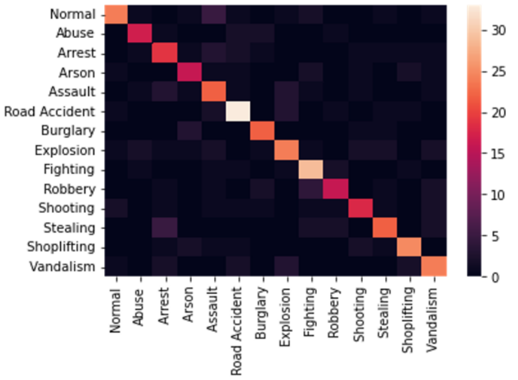

Figure 8: Depicts the confusion matrix for the

Table 2: The performance analysis of different typ

After the detection of anomalous activity, multi class classification performed using simple RNN so as to classify the videos into the type of anomaly detected in the video.

The Table 2 depicts the performance measures scores for multi- class classification of anomalies. As depicted clearly, the accuracy obtained for classification was around 83.67%. The anomalous activities were classified into either of the 13 anomaly categories considered or into the category of normal videos. Further, Some false results were also obtained in case of anomaly detection where, either the normal videos were classified as anomalous or anomalous activity was not detected and the videos were considered normal. Compared with the previous work

	Precision	**Recall**	**F1-Score**	**Support**
Normal	0.844444	0.883721	0.863636	43
Abuse	0.736842	0.823529	0.777778	34
Arrest	0.780488	0.8	0.790123	40
Arson	0.821429	0.793103	0.807018	29
Assault	0.861111	0.794872	0.826667	39
Road Accident	0.904762	0.883721	0.894118	43
Burglary	0.864865	0.864865	0.864865	37
Explosion	0.852941	0.763158	0.805556	38
Fighting	0.885714	0.794872	0.837838	39
Robbery	0.891892	0.891892	0.891892	37
Shooting	0.8	0.818182	0.808989	44
Stealing	0.825	0.846154	0.835443	39
Shoplifting	0.808511	0.863636	0.835165	44
Vandalism	0.885714	0.911765	0.898551	34
Accuracy			**0.83666**	
Macro Avg	0.840265	0.838105	0.838403	540
Weighted Avg	0.840576	0.838889	0.838961	540

done on this problem, our model outperforms other methods and has obtained an accuracy of 83.67%. The Table 3, depicts the comparative analysis of our approach with the methodology adopted by others. The proposed method involved the latest idea of deep learning concepts. As per the comparative study shown in the Table 3, our proposed models work well for complex cases and achieve the best results in many of the past studies.

Table3: Comparative Analysis

Methodology adopted	**Accuracy Obtained**
2D CNN	72%
C3D (2 stream)	75.9%
I3D (RGB)	79.9%
C3D on Segmented videos (Our Approach)	**83.6%**
ResNET 50	45.20
MobileNet V2	41.90
[34]	52.70
ConvGRU-CNN [32]	82.22%
Pushpajit Khaire[33]	91.90%

6 Conclusion and future scope

To detect real-world abnormalities in surveillance films, we present a deep learning technique. because of the intricacy of these realistic anomalies, utilizing simply normal data for anomaly identification may not be the best option. We try to take advantage of both conventional and unusual surveillance recordings. We proposed a twofold approach for the detection and classification of anomalies. We made use of C3D for feature extraction and detection of probabilistic anomaly in surveillance videos. Further we trained a simple RNN for multi- class classification to determine the anomaly types. We achieved an accuracy was around 83.67% for multi- class classification. We further plan on developing a real time surveillance system to detect anomalous activity and notify the required authorities as and when an anomaly is suspected in any area. We wish to come up with IoT based solutions to raise alarm before crime incident takes place. Raising awareness amongst citizens about their rights and encourage them to raise their voice against crime remains a major concern. In future we are planning to add the upcoming.

Constraints of the study:

Given the fact that this paper is done with great care and attention to detail, there are certain constraints we encountered are as follows.

- First, the entire system must be built correctly and completely in the near future in order for it to be implemented quickly and correctly. Furthermore, because such technologies cannot be directly deployed in the open world, implementation is a major problem, wrt cost.
- The system must first be evaluated in a small region of a metropolitan area, and only then can its use be expanded up with continual improvements (first-model revisions). As a result, the obstacles are more of a help in improving the model and producing a flawless model that can be applied to the actual world over time.
- Additionally, there are a few technological challenges with the model, since the amount of the learning data will be vast, requiring days, if not weeks, to analyze it.

References

[1] Mohammadi, Sadegh, et al. "Angry crowds: Detecting violent events in videos." *European Conference on Computer Vision*. Springer, Cham,

2016. https://doi.org/10.1007/978-3-319-46478-7_1.

[2] Kamijo, Shunsuke, et al. "Traffic monitoring and accident detection at intersections." *IEEE transactions on Intelligent transportation systems* 1.2 (2000): https://doi.org 108-118.W. 10.1109/6979.880968.

[3] Sultani, Waqas, and Jin Young Choi. "Abnormal traffic detection using intelligent driver model." *2010 20th International Conference on Pattern Recognition*. IEEE, 2010. 10.1109/ICPR.2010.88

[4] Lu, Cewu, Jianping Shi, and Jiaya Jia. "Abnormal event detection at 150 fps in matlab." *Proceedings of the IEEE international conference on computer vision*. 2013. DOI: 10.1109/ICCV.2013.338

[5] B. Zhao, L. Fei-Fei, and E. P. Xing. Online detection of unusual events in videos via dynamic sparse coding. In CVPR, pages 3313–3320, 2011. DOI: 10.1109/CVPR.2011.5995524

[6] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey." ACM computing surveys (CSUR) 41.3 (2009): 1-58. https://doi.org/10.1145/1541880.1541882

[7] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Perez. Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence, 89(1):31–71, 1997. https://doi.org /10.1016/s0004-3702(96)00034-3.

[8] Andrews, Stuart, Ioannis Tsochantaridis, and Thomas Hofmann. "Support vector machines for multiple-instance learning." *Advances in neural information processing systems* 15 (2002). https://doi.org/10.1007/s10479-012-1241-z.

[9] Tran, Du, et al. "Learning spatiotemporal features with 3d convolutional networks." *Proceedings of the IEEE international conference on computer vision*. 2015. https://doi.org/10.1109/ICCV.2015.510.

[10] Hou, Rui, Chen Chen, and Mubarak Shah. "Tube convolutional neural network (T-CNN) for action detection in videos." *Proceedings of the IEEE international conference on computer vision*. 2017. https://doi.org/10.48550/arXiv.1703.10664.

[11] Al-Hazaimeh, Obaida M., and M. Al-Smadi. "Automated pedestrian recognition based on deep convolutional neural networks." *International Journal of Machine Learning and Computing* 9.5 (2019): 662-667. https://doi.org/10.18178/ijmlc.2019.9.5.855.

[12] Ritwika Chowdhury, Kinjal Bhattacharyya, Sudipta Mukhopadhyay and Bhargab Maitra, "Multi-Directional Pedestrian Detection and Flow Monitoring from Traffic Videos", Proceedings of International Conference on Transportation Research, pp. 34-41, 2018.

[13] Feng-Ping An, "Pedestrian Re-Recognition Algorithm Based on Optimization Deep Learning-Sequence Memory Model", Complexity, Vol. 2019, pp. 1-6, 2019. https://doi.org/10.1155/2019/5069026

[14] Kim, Sangjun, Sooyeong Kwak, and Byoung Chul Ko. "Fast pedestrian detection in surveillance video based on soft target training of shallow random forest." *IEEE Access* 7 (2019): 12415-12426. https://doi.org/10.1109/ACCESS.2019.2892425

[15] Shuqiang Guo, Qianlong Bai, Song Gao, Yaoyao Zhang and Aiquan Li, "An Analysis Method of Crowd Abnormal Behavior for Video Service Robot", IEEE Access, Vol. 7, pp. 169577-169585, 2019. DOI: 10.1109/ACCESS.2019.2954544

[16] Shuqiang Guo, Qianlong Bai, Song Gao, Yaoyao Zhang and Aiquan Li, "A Deep Spatiotemporal Perspective for Understanding Crowd Behavior", IEEE Access, Vol. 20, pp. 3289-3297, 2018. DOI: 10.1109/TMM.2018.2834873.

[17] Biao Yang, Jinmeng Cao, Nan Wang and Xiaofeng Liu, "Anomalous Behaviors Detection in Moving Crowds based on a Weighted Convolutional Auto Encoder-Long ShortTerm Memory Network", IEEE Transactions on Cognitive and Developmental Systems, Vol. 11, No. 4, pp. 473-482, 2018. https://doi.org/10.1109/TCDS.2018.2866838

[18] Qi Wang, Mulin Chen, Feiping Nie, Xuelong Li, "Detecting Coherent Groups in Crowd Scenes by Multi view Clustering", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 42, No. 1, pp 46-58, 2018. https://doi.org/10.1109/TPAMI.2018.2875002.

[19] Yu Hao, Zhi-Jie Xu, Ying Liu, Jing Wang, Jiu-Lun Fan, "Effective Crowd Anomaly Detection through Spatiotemporal Texture Analysis", International Journal of Automation and Computing, Vol. 16, pp 27-39, 2019. https://doi.org/10.1007/s11633-018-1141-z.

[20] G. Sreenu and M.A. Saleem Durai, "Intelligent Video Surveillance: A Review through Deep Learning Techniques for Crowd Analysis", Journal of Big Data, Vol. 4, No. 48, pp. 1-13, 2019. https://doi.org/10.1186/s40537-019-0212-5.

[21] S. Manjula and K. Lakshmi, "Detection and Recognition of Abnormal Behaviour Patterns in Surveillance Videos using SVM Classifier", Proceedings of International Conference on Recent Trends in Computing, Communication and Networking Technologies, pp. 1-7, 2019. https://dx.doi.org/10.2139/ssrn.3453542.

[22] Aravindan, S., E. Anusuya, and M. Ashok Kumar. "GUI BASED PREDICTION OF CRIME RATE USING MACHINE LEARNING APPROACH." (2020).

[23] Wang, Bao, et al. "Deep learning for real-time crime forecasting and its ternarization." Chinese Annals of Mathematics, Series B 40.6 (2019): 949-966. https://doi.org/10.1007/s11401-019-0168-y.

[24] Tabedzki, Christian, et al. "Yo home to Bel-Air: predicting crime on the streets of Philadelphia." University of Pennsylvania, CIS 520: machine

learning. 2018. https://doi.org/10.1145/3348445.3348483

[25] Bandekar, Shraddha Ramdas, and C. Vijayalakshmi. "Design and analysis of machine learning algorithms for the reduction of crime rates in India." Procedia Computer Science 172 (2020): 122-127. https://doi.org/10.1016/j.procs.2020.05.018

[26] Hossain, Sohrab, et al. "Crime prediction using Spatio-temporal data." International Conference on Computing Science, Communication and Security. Springer, Singapore, 2020. https://doi.org/10.1007/978-981-15-6648-6_22.

[27] Tong, Kang, Yiquan Wu, and Fei Zhou. "Recent advances in small object detection based on deep learning: A review." *Image and Vision Computing* 97 (2020): 103910. https://doi.org/10.1016/j.imavis.2020.103910.

[28] Gandapur, Maryam Qasim. "E2E-VSDL: End-to-end video surveillance-based deep learning model to detect and prevent criminal activities." *Image and Vision Computing* (2022): 104467.https://doi.org/10.1016/j.imavis.2022.104467.

[29] Zhang, L., Li, S., Luo, X. *et al.* Video anomaly detection with both normal and anomaly memory modules. *VisComput* (2024).https://doi.org/10.1007/s00371-024-03584-z.

[30] Sheng, Bin, et al. "Depth-Aware Motion Deblurring Using Loopy Belief Propagation." IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 4, 2019, pp. 955-969. https://doi.org/10.1109/TCSVT.2019.2901629.

[31] Natha, Sarfaraz, et al. "Deep BiLSTM Attention Model for Spatial and Temporal Anomaly Detection in Video Surveillance." *Sensors* 25.1 (2025): 251. https://doi.org/10.3390/s25010251.

[32] Qasim Gandapur, Maryam, and Elena Verdú. "ConvGRU-CNN: Spatiotemporal deep learning for real-world anomaly detection in video surveillance system." (2023). https://doi.org/10.1109/TII.2019.2938527.

[33] Pushpajit Khaire, Praveen Kumar, “A semi-supervised deep learning based video anomaly detection framework using RGB-D for surveillance of real-world critical environments”, Forensic Science International: Digital Investigation", Volume 40, 2022. https://doi.org/10.1016/j.fsidi.2022.301346.

[34] Saleem, Gulshan, et al. "Efficient anomaly recognition using surveillance videos." *PeerJ Computer Science* 8 (2022): e1117. https://doi.org/10.7717/peerj-cs.1117.