Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

Detection of Unusual Activity in Surveillance Video Scenes Based on Deep Learning Strategies

Muthana S. Mahdi [(a)], Amer Jelwy Mohammed [(b)], Abdulghafor waedallah Abdulghafour [(c)]

[a] Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq, muthanasalih@uomustansiriyah.edu.iq

[b] Dewan Al-Waqf Al-Sunni, Baghdad, Iraq, amerjelewy@gmail.com

[c] Presidency of Mustansiriyah University, Baghdad, Iraq, vip.mostansirya@gmail.com

ARTICLEINFO

Article history:
Received: 02 /11/2021
Rrevised form: 19 /11/2021
Accepted : 05 /12/2021
Available online: 07 /12/2021

Keywords:
Abnormal activity,
Human-computer interaction,
deep-learning strategies,
Automated detection,
activity analysis,
surveillance scenes.

ABSTRACT

In today's world, abnormal activity indicates threats and risks to others. An anomaly can be defined as something that deviates from what is expected, common, or normal. Because it is difficult to continuously monitor public spaces, intelligent video surveillance is necessary. When artificial intelligence, machine learning, and deep learning were introduced into the system, the technology had progressed much too far. Different methods are in place using the above combinations to help distinguish various suspicious activities from the live tracking of footage. Human behavior is the most unpredictable, and determining whether it is suspicious or normal is quite tough. In an academic setting, a deep learning Technique is utilized to detect normal or abnormal behavior and sends an alarm message to the appropriate authorities if suspicious activity is predicted. Monitoring is frequently carried out by extracting successive frames from a video. The framework is split into two sections. The features are calculated from video frames in the first phase, and the classifier predicts whether the class is suspicious or normal in the second part based on the obtained features. This paper proposes an effective method to design a system that automatically detects any unexpected or abnormal circumstance and alerts the appropriate authority and it can be used in both indoor and outdoor settings in an academic area. The proposed system was able to achieve an accuracy rate of 95.3 percent.

MSC. 41A25; 41A35; 41A36

DOI : https://doi.org/10.29304/jqcm.2021.13.4.858

1. Introduction

In the actual world, human behavior recognition has a wide range of applications, such as intelligent video monitoring. Video surveillance offers a wide range of applications, particularly for indoor and outdoor environments. Surveillance is an important aspect of safety. Nowadays, security cameras have become an integral aspect of life for the sake of security and safety. Effective monitoring, minimal labor requirements, cost-effective

∗ Corresponding author: Muthana S. Mahdi

Email addresses: muthanasalih@uomustansiriyah.edu.iq

Communicated by: Dr. Rana Jumaa Surayh aljanabi.

auditing capability, and the adoption of new security trends are all advantages of video surveillance. It is now difficult to manually monitor all occurrences on a camera of CCTV (Closed Circuit Television). Even if the event had previously occurred, it takes a long time to manually search the recorded video. In the field of surveillance systems of automated video, analyzing abnormal occurrences is a new topic [1]. In a video surveillance system, human behavior detection is an automatic means of intelligently recognizing any suspicious conduct. For the automatic identification of body activity in general locations such as stations of railway, airports, offices, banks, and rooms of examination, a variety of efficient algorithms are available. Video surveillance is a new area in which Machine Learning, Artificial Intelligence, and Deep Learning are being used. Artificial intelligence enables a machine to think in the same way as a people [2]. Learning from data of training and making predictions about future data are fundamental components of machine learning. Because the processors of GPU (Graphics Processing Unit) and large datasets are now available, deep learning is being applied. Computer vision and scene surveillance will be used in tandem to ensure public safety and security. Modeling of environments, moving objects classification, tracking, behavior interpretation and description, detection of motion, and synthesis of input from many cameras are all stages of computer vision methods. To extract features from different scene sequences, this method necessitates a lot of pre-processing [3].

There are three types of classification techniques: unsupervised, supervised, and Semi-supervised. Classification of unsupervised is computer-controlled and does not need any human participation. Classification of supervised is needed to manual training and labeled data. Semi-supervised learning falls among unsupervised learn (with no labeled training data) and supervised learn (with labeled training data). Deep Neural Networks are one of the most effective architectures for tackling tough learning problems. The models of deep Learning extract features and generate high-level representations of picture data automatically. Because the feature extraction procedure is automated, this is more generic. Convolutional neural networks (CNNs) can learn visual patterns directly from picture pixels. The models of long short-term memory (LSTM) are capable of learn long-term dependencies in the case of video streams. The LSTM network is capable of remembering information. The proposed system will monitor human behavior and gently alert when any suspicious events occur using footage gathered from CCTV cameras [4]. Event detection and person behavior recognition are two key components of intelligent video surveillance. Understanding human behavior automatically is a difficult challenge [5].

The complete surveillance system training process may be broken down into 3 stages: The preparation of data, model training, and inference. Two neural networks (CNN), and Recurrent Neural Network (RNN), make up the framework. The CNN algorithm is used to extract features of high level from the images to minimize the input's complexity. For categorization, RNN is utilized, which is ideally suited for video stream processing. The presented system makes use of a VGG16 (Visual Geometry Group) pre-trained model that was developed on the data set of ImageNet [6]. In this time, the model is being trained to predict behavior based on the film. In the footage utilized to enhance the monitoring process, the model can predict typical or suspicious human behavior. The majority of the present system relies on footage acquired from surveillance cameras. If a crime or act of violence occurs, this scene will be utilized to aid in the investigation. However, a system that automatically detects any unexpected circumstance in advance and alerts the appropriate authority is more appealing, and it can be used in both indoor and outdoor settings [7].

The proposed approach is to create a system to detect abnormal behavior in an academic setting. The following is how the paper is structured: The second section summarizes the related works in the field of analysis of behavior for the detection of unusual activity. Section three provides an overview of the proposed approach. In section four the specifics of the implementation are provided, then in section five the conclusion and future work.

2. Related Works

Different ways for recognizing human activities from the video are suggested in the associated publications. The work's goal was to detect any unusual events in a video scene surveillance system. An unauthorized entry into a restricted location was detected using the Advance Motion Detection (AMD) method [8]. The object was discovered using background subtraction in the first phase, and the object was retrieved from frame sequences in the second phase. The discovery of questionable activity was the second phase. The system's advantage was that the method worked in real time and had a low computational complexity. However, the system's storage capacity was limited, and it might be combined with a high-tech manner of video capture in monitoring regions. In [9], a semantic-based strategy was developed. Background subtraction was used to identify the foreground items from the collected video

data. After subtracting the items, a Hear-like technique is used to classify them as living or non-living. The Real-Time blob matching technique was used to track the objects. In this paper, there was also fire detection.

The system was tested and verified using a variety of datasets, including the UMN dataset and PETS. In [10], People tracking could be used to discover unexpected happenings in video footage. Using the background subtraction method, human persons are spotted in the footage. CNN was used to extract the features, which were then fed into a DDBN (Discriminative Deep Belief Network). The DDBN is also given labeled footage of some questionable incidents and extracts their features. Then, using a DDBN, features derived using CNN were compared to features recovered from a labeled sample video of categorized suspicious actions, and numerous suspicious activities were discovered from the video.

Videos of daily human activities were taken and classified into four categories: domestic, work-related, caring, and helping [11]. Deep learning is used for sports-related tasks. For obtaining input features, CNN was employed, and RNN was used for classification. They employed the Inception v3 model and datasets from UCF101 and Activitynet. On UCF101, 85.9% accuracy was attained, and on Activitynet, 45.9% accuracy was achieved. Suspicious behaviors were found in [12] based on the motion attributes between the objects. To define suspicious events, a semantic approach was applied. To track objects, the object detection and correlation approach was utilized. Motion features and temporal information are used to classify the events. The given framework had a lower computational complexity. In [13], the optical flow in every zone was evaluated utilizing the method of Lucas-Kanade after abnormal occurrences from a university were separated into zones. The magnitude histogram of optical flow vectors was then produced. The content of a video is analyzed using software algorithms to classify events as unusual or usual. Based on the analysis of movement data from video scenes sequences, an approach was created to identify aberrant events from regular events [14]. The histogram of optical flow orientation in the video frames was learned using the HMM approach. It compares the acquired video frames to the existing usual frames and determines whether they are similar.

The VID data set was employed, and the accuracy for detecting violence in football stadiums was 94.5 percent. In [15], the abnormal event detection system is made up of several modules that process video data. Human activity was detected using deep architectures. The UT Interaction dataset was employed in the proposed CNN and LSTM models. One of the system's flaws was that it was difficult to distinguish between identical human behaviors such as punching or pointing.

Using the approach of deep spatiotemporal to crowd behavior, the films are divided into three categories: destination estimation, prediction of the pedestrian future path, and behavior of holistic crowd [16]. A convolutional layer was used to extract spatial information from video frames. The architecture of LSTM was utilized to learn or grasp the dynamics of temporal motion sequences. PWPD, ETH, UCY, and CUHK data sets were employed in the proposed system. By employing deeper architectures, the system's accuracy can be increased. Using a neural network and a Gaussian distribution, a system was created to track students' exam conduct [17]. Face detection, suspicious state detection, and anomaly detection are the three stages of the process. The trained model determines whether or not the student is suspicious, and the Gaussian distribution determines whether or not the student engages in any unusual behavior. An accuracy of 93 percent was achieved in this method. To prevent audience or player violence in sports, a detection system of real-time violence based on deep learning was proposed [18]. Frames from real-time videos were retrieved in a spark environment. If there is any football violence detected by the system, security personnel will be notified. To prevent violence from occurring in the first place, the system recognizes camera actions in real-time and warns security forces. In this method, 94.7% accuracy was achieved.

It was addressed how to use surveillance of intelligent video for analysis of crowd in [19]. This was a review paper that discussed the importance of video surveillance analysis in nowadays society, as well as the numerous deep learning models, datasets, and algorithms that are employed in video surveillance analysis. For recognizing human behavior analysis from movies, the majority of the publications described above-used computer vision employing different algorithms. To comprehend the evolution of features in a scene series, computer vision systems need a lot of preprocessing to trajectories extraction of motion patterns [20]. Furthermore, background subtraction is based on a static hypothesis of background that is frequently inapplicable in real-world circumstances. In the real world,

the majority of problems arise in crowds. When it comes to dealing with crowds, the solutions mentioned above are inefficient. Based on the literature review, a deep architecture for unusual activity prediction can be developed using LSTM and 2D CNN to increase the system's accuracy. The majority of articles using a deep learning approach merely detect questionable activities. As a result, an effective method is required to notify security in the event of any questionable behavior.

3. The Proposed System

When any suspicious occurrence occurs, the suggested system will use footage gathered from CCTV cameras to monitor students' behavior and deliver a message to the appropriate authority.

3.1 Architecture of the System

Video capture, scene preprocessing, feature extraction, events classification, and prediction are all phases of the architecture. Figure (1) depicts the overall layout of the system architecture. The videos are divided into two categories by the algorithm.

1) On campus, students fainting or fighting - unusual class.
2) Normal class: running and walking.

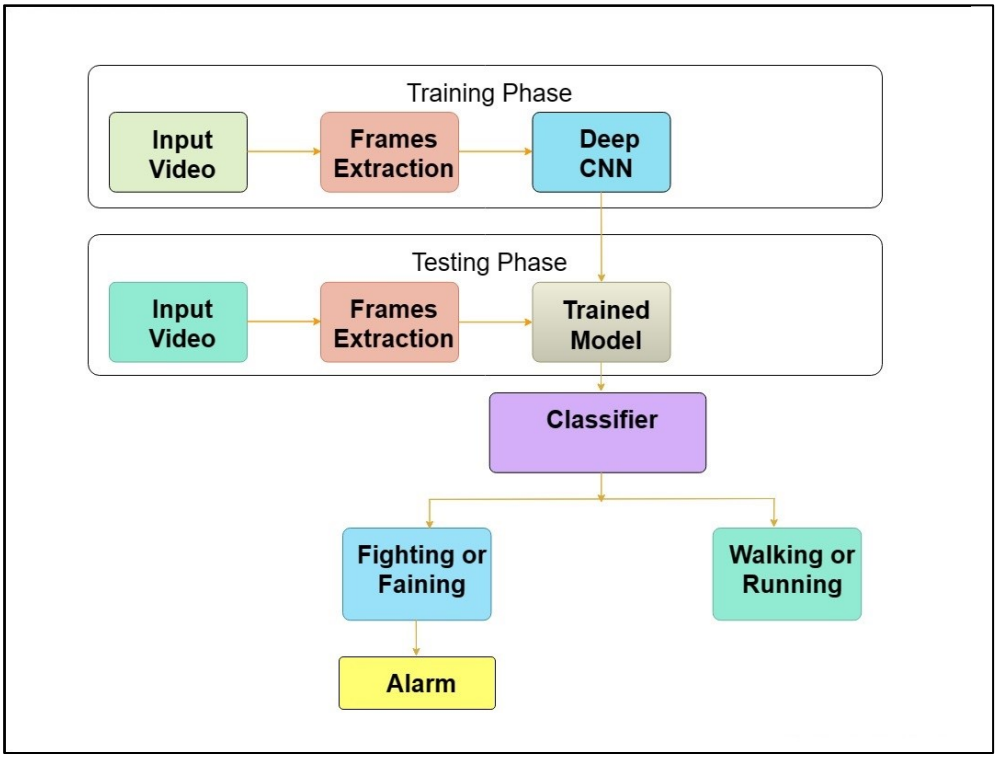

Fig1 - The overall layout of the system architecture

3.2 Capturing Video

The first stage in a video surveillance system is to install a CCTV camera and monitor the footage. Various types of videos are taken from various cameras, which cover the entire surveillance region. Because our implementation uses frames for processing, the videos are transformed into frames.

3.3 Description of the data set

The KTH data set is a standard data set that contains a collection of sequences representing six activities, with 100 sequences for each action type. Each segment comprises about 600 frames, and the movie is shot at a speed of 25 frames per second [21]. This dataset is used to train the model for typical behavior (running and walking). Suspicious behavior is trained using the CAVIAR dataset and YouTube videos (fainting or fighting). The dataset is manually labeled and divided into two sets: 20 percent for validation and 80 percent for training. The structure of the dataset is given in figure (2). In our system, we employ a collection of CAVIAR, KTH, and YouTube scenes.

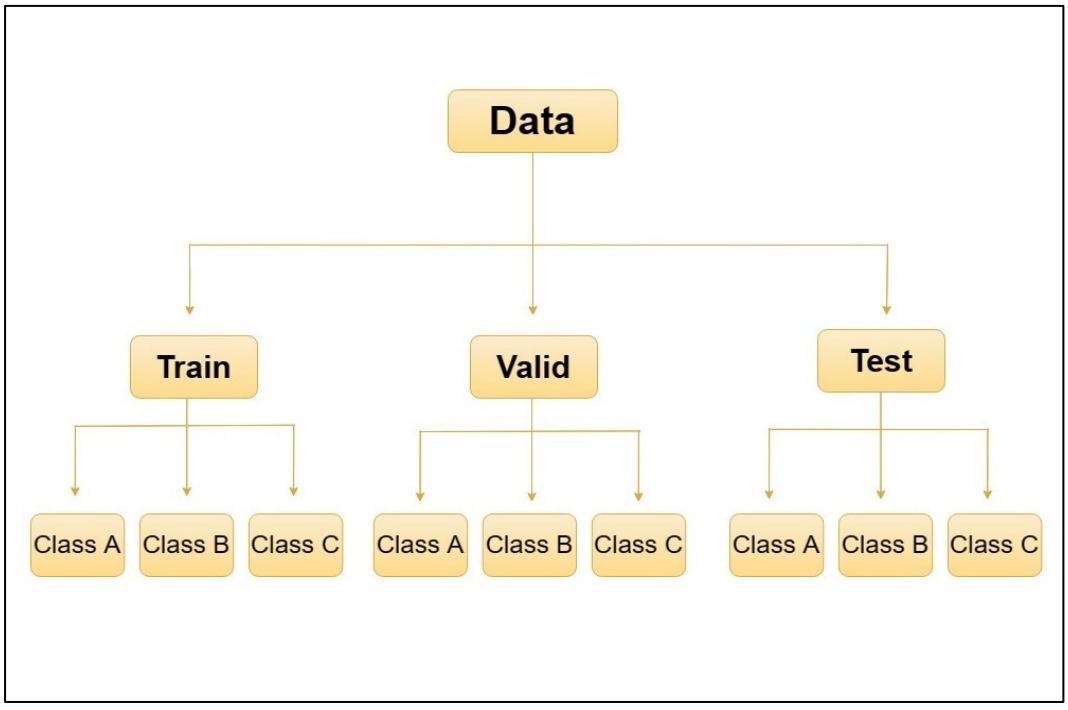

Fig2 - The structure of the Dataset directory

3.4 Video preprocessing

In the proposed system for detecting unusual activities from scenes surveillance, a deep learning network is used. The accuracy attained with deep learning architectures can be bigger, and it also performs better with big data sets. The videos for the input are drawn from both existing and newly developed datasets. Frames are retrieved from collected videos as part of the pre-processing process. Three labeled folders are formed based on the videos, and the frames are placed in them. The entire movie is changed to (7035) frames, which are saved as jpeg files. Every frame is then scaled to (224 x 224) pixels to fit inside the 2DCNN architecture and saved. The test video is also saved in a folder after being converted to frames and scaled to 224 x 224. For video pre-processing, the Python package Open CV is used.

In the image features extraction stage, a model pre-trained CNN called VGG16 is trained on the image Net data set. Figure (3) depicts the VGG16 architecture. The architecture of deep learning utilized here was the neural network of VGG16 [22], which contains convolution layers of (3x3) size, max-pooling layers of (2x2) size, and in final there are the fully connected layers, where the total are 16 layers. The input image should be RGB 224 x 224 x 3 pixels in size. Layers of Convolution, a layer of ReLU (Rectified Linear Unit) (activation function), Layers of max pooling, Layers of fully connected dense, and Layers of normalizing are all represented. The model can be fine-tuned to our specifications, and the model's last layer is deleted. The model is then trained using LSTM (Long Short-Term Memory) architecture. In sequence prediction problems, the networks of LSTM are a type of RNN that can learn order dependency. Now

there is a ReLU activation function, a layer of dropout, and layers of dense that are fully coupled. The number of neurons in the end layer is equal to our classes number, thus there are three neurons here.

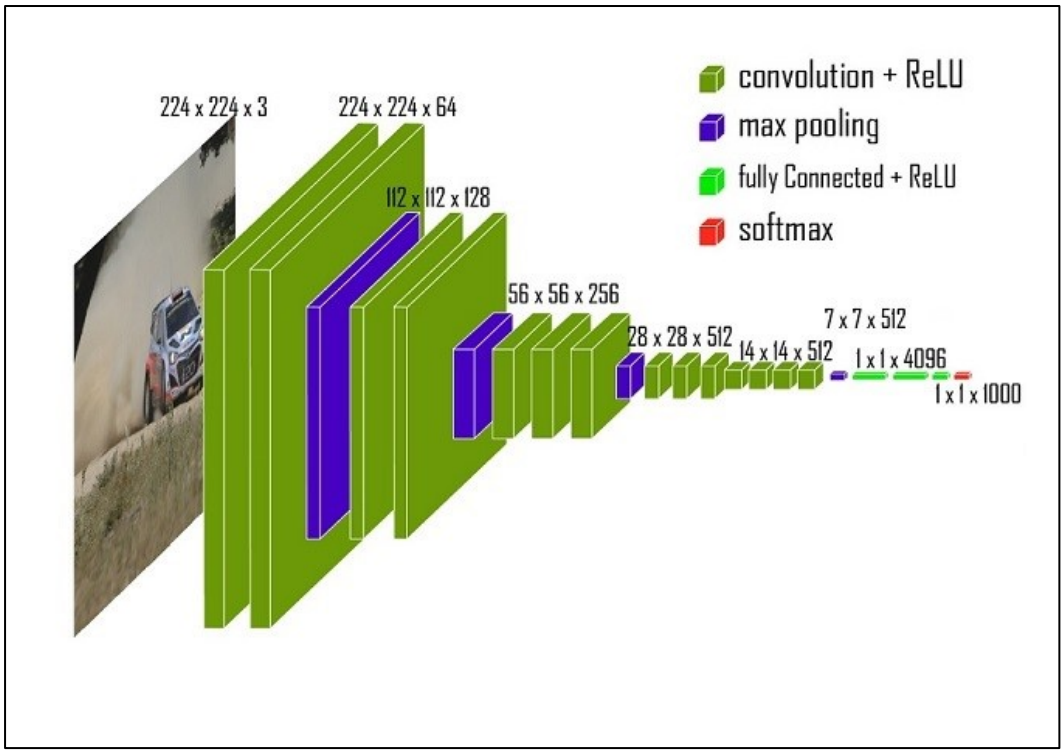

Fig3 - The VGG16 architecture.

The system categorizes the videos as either unusual (fainting or fighting) or regular (running, walking). An SMS (Short Message Service) will be sent to the appropriate officials in the event of suspicious conduct.

4. Results and Analysis

The project's goal is to use CCTV footage to track suspicious activity on campus and to warn security when anything suspect happens. This was accomplished by utilizing CNN to extract features from the frames. After the frames have been extracted, the LSTM architecture is utilized to identify them as unusual or usual. The unusual or usual video sequences are shown in figure (4).

The collect scene sequences from CCTV footage, extracting frames from movies, preprocessing and preparing train, and validation collection from data sets, train, and test are the phases in developing the entire system. When the system detects questionable activities, it sends an SMS to the appropriate officials. The system was created in Python on a platform (open-source). Setting an account with Twilio and installing the library of Twilio in Python allows you to send SMS. Twilio allows you to make and receive phone calls, as well as send and receive text messages, programmatically.

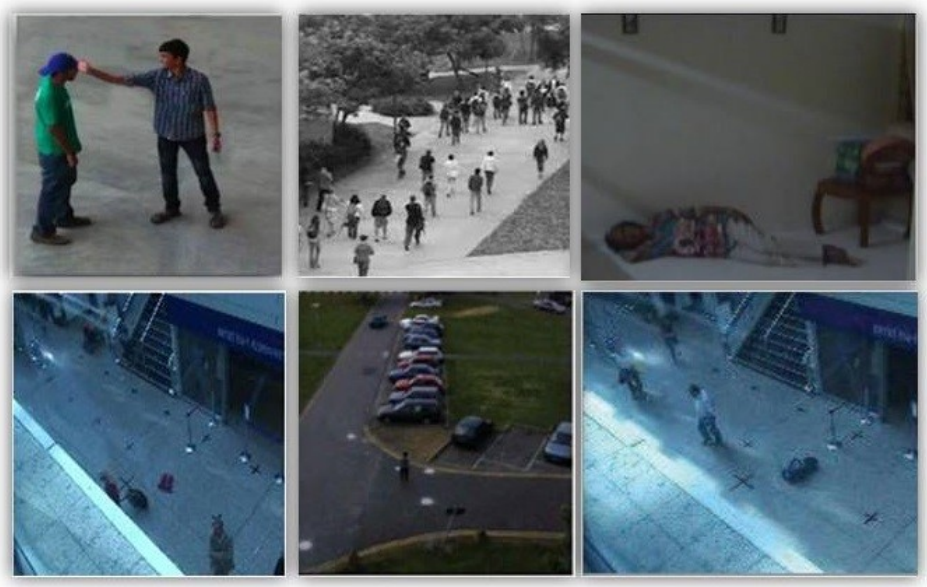

Fig4 - The suspicious and normal video sequences.

4.1 Training & Testing

The video scenes for the input have been obtained from the CAVIAR data set, the KTH data set, and YouTube scenes. A total of 300 scenes showing suspicious and regular behavior have been gathered. Frames are retrieved from collected videos as part of pre-processing. VGG16 is the pre-trained model and it is used to solve our problem. Based on our requirements, the last layer of this model is eliminated, and classification is performed using the LSTM architecture. The dataset used is trained on it. For testing, CCTV video footage from various circumstances is captured and translated into frames. The trained model receives the stored frames, and the classifier then categorizes the video as suspicious or usual activity. The dataset used for training was 80% and 20% for testing.

4.2 The Results

For the first 10 epochs, the training phase's accuracy is 86 percent. By increasing the iterations number, the model's accuracy can be enhanced. For testing purposes, the frames are taken from scenes and kept in one folder. The algorithm classifies the frames as suspicious (fainting or fighting) or regular (running or walking) using our trained model. In the event of suspicious activity, a notice with the expected class will be forwarded to the appropriate authority. 95.3 percent accuracy was attained. To evaluate the accuracy of the proposed algorithm, it was compared with three algorithms mentioned in the relevant works. Table (1) presents the final results. Figure (5) shows the overall accuracy of the unusual activity detection algorithms.

Table (1) shows the final results.

True positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
161	125	6	8
$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} = 95.3$			

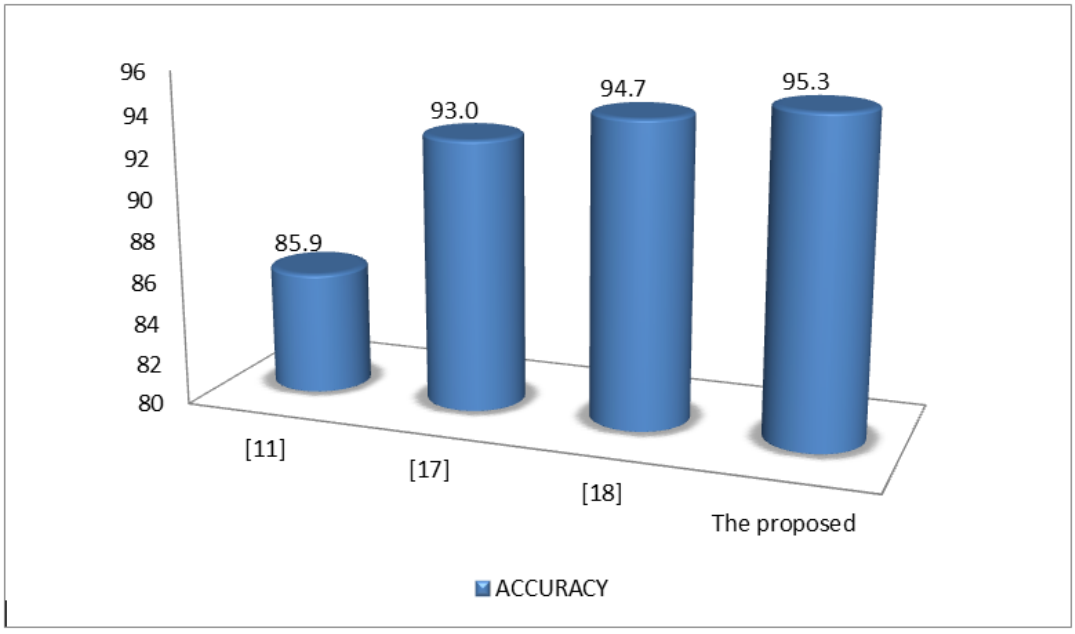

Fig5 - The comparison process between the proposed work and other relevant works.

5. *Conclusions*

Nowadays, practically everyone understands the importance of CCTV footage, yet in most circumstances, this footage is only utilized for purposes of investigation after an incident or crime has occurred. The proposed system offers the advantage of sending alerts when an accident occurred. CCTV footage is being tracked and analyzed in real-time. The analysis' outcome is a directive to appropriate officials to decide if the result shows that an undesirable event is likely to occur. As a result, this can be prevented. Although the proposed approach is dedicated to the academic realm, it can be utilized to forecast more unusual actions in private or public settings. The system can be utilized in any location where training should be delivered in conjunction with the unusual activity that is appropriate for that location.

References

[1] Ruff, L., Vandermeulen, R. A., Gornitz, N., Binder, A., Muller, E., & Kloft, M. Deep support vector data description for unsupervised and semi-supervised anomaly detection. In Proceedings of the ICML 2019 Workshop on Uncertainty and Robustness in Deep Learning, Long Beach, CA, USA,2019.

[2] Fan, Y., Wen, G., Li, D., Qiu, S., Levine, M. D., & Xiao, F. Video anomaly detection and localization via Gaussian mixture fully convolutional variational autoencoder. Computer Vision and Image Understanding, 102920, 2020.

[3] Gkountakos, K., Ioannidis, K., Tsikrika, T., Vrochidis, S., & Kompatsiaris, I. A Crowd Analysis Framework for Detecting Violence Scenes. In Proceedings of the 2020 International Conference on Multimedia Retrieval (pp.276-280), 2020.

[4] Lin, W., Hasenstab, K., Cunha, G. M., & Schwartzman, A.Comparison of handcrafted features and convolutional neural networks for liver MR image adequacy assessment. Scientific Reports, 10(1), 1-11, 2020.

[5] Ramchandran, A., & Sangaiah, A. K. Unsupervised deep learning system for local anomaly event detection in crowded scenes. Multimedia Tools and Applications, 1-21, 2019.

[6] Krizhevsky, A., Sutskever, I., & Hinton, G. E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105, 2012.

[7] Kamoona, A. M., Gostar, A. K., Bab-Hadiashar, A., & Hoseinnezhad, R. Sparsity-Based Naive Bayes Approach for Anomaly Detection in Real Surveillance Videos. In 2019 International Conference on Control, Automation and Information Sciences (ICCAIS) (pp. 1-6), 2019.

[8] P.Bhagya Divya, S.Shalini, R.Deepa, Baddeli Sravya Reddy, "Inspection of suspicious human activity in the crowdsourced areas captured in surveillance cameras", International Research Journal of Engineering and Technology (IRJET), December 2017.

[9] Jitendra Musale, Akshata Gavhane, Liyakat Shaikh, Pournima Hagwane, Snehalata Tadge, "Suspicious Movement Detection and Tracking of Human Behavior and Object with Fire Detection using A Closed Circuit TV (CCTV) cameras ", International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 5 Issue XII December 2017.

[10] Elizabeth Scaria, Aby Abahai T and Elizabeth Isaac, "Suspicious Activity Detection in Surveillance Video using Discriminative Deep Belief Network", International Journal of Control Theory and Applications Volume 10, Number 29 -2017.
[11] Javier Abellan-Abenza, Alberto Garcia-Garcia, Sergiu Oprea, David Ivorra-Piqueres, Jose Garcia-Rodriguez "Classifying Behaviours in Videos with Recurrent Neural Networks", International Journal of Computer Vision and Image Processing, December 2017.
[12] U.M.Kamthe, C.G.Patil "Suspicious Activity Recognition in Video Surveillance System", Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018.
[13] Zahraa Kain, Abir Youness, Ismail El Sayad, Samih Abdul-Nabi, Hussein Kassem, " Detecting Abnormal Events in University Areas ", International Conference on Computer and Application,2018.
[14] Tian Wanga, Meina Qia, Yingjun Deng, Yi Zhouc, Huan Wangd, Qi Lyua, Hichem Snoussie, "Abnormal event detection based on analysis of movement information of video sequence", Article-Optik, vol- 152, January-2018.
[15] Kwang-Eun Ko, Kwee-Bo Sim"Deep convolutional framework for abnormal behavior detection is a smart surveillance system."Engineering Applications of Artificial Intelligence,67 (2018).
[16] Yuke Li "A Deep Spatiotemporal Perspective for Understanding Crowd Behavior", IEEE Transactions on Multimedia, Vol. 20, NO. 12, December 2018.
[17] Asma Al Ibrahim, Gabriel Abosamra, Mohamed Dahab "Real-Time Anomalous Behavior Detection of Students in Examination Rooms Using Neural Networks and Gaussian Distribution", International Journal of Scientific and Engineering Research, October 2018.
[18] Dinesh Jackson Samuel R, Fenil E, Gunasekaran Manogaran, Vivekananda G.N, Thanjaivadivel T, Jeeva S, Ahilan A, "Real-time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM", The International Journal of Computer and Telecommunications Networking,2019.
[19] G. Sreenu and M. A. Saleem Durai "Intelligent video surveillance: a review through deep learning techniques for crowd analysis", Journal Big Data,2019.
[20] Gurav, S. S., Godbole, B. B., & Sonale, M. S. Improved accuracy of suspicious activity detection in surveillance video. International journal of engineering and advanced technology, 9(3), 267-270, 2020.
[21] K. Kavikuil and Amudha, J., "Leveraging deep learning for anomaly detection in video surveillance", Advances in Intelligent Systems and Computing,2019.
[22] Sudarshana Tamuly, C. Jyotsna, Amudha J, "Deep Learning Model for Image Classification", International Conference on Computational Vision and Bio-Inspired Computing (ICCVBIC),2019.
[23] A. Ali, M. RASHEED, S. SHIHAB, T. RASHID, A. Sabri, and S. Abed Hamed, "An Effective Color Image Detecting Method for Colorful and Physical Images", JQCM, vol. 13, no. 1, pp. Comp Page 88 -, Mar. 2021.
[24] S. Hasen and A. Abdulhadi, "Influence of a Rotating Frame on the Peristaltic Flow of a Rabinowitsch Fluid Model in an Inclined Channel", JQCM, vol. 12, no. 1, pp. Math Page 21 -, Feb. 2020.
[25] A. Hameed Khaleel, "Automated ovarian masses extraction in CT images based on division of image", JQCM, vol. 6, no. 1, pp. 11-27, Aug. 2017.
[26] K. Neamah Hussein, "Video Frames Edge Detection of Red Blood Cells: A Performance Evaluation", JQCM, vol. 10, no. 1, pp. Comp Page 16 - 27, Jan. 2018.
[27] S. Turky, A. Ahmed AL-Jumaili, and R. Hasoun, "Deep Learning Based On Different Methods For Text Summary: A Survey", JQCM, vol. 13, no. 1, pp. Comp Page 26-, Mar. 2021.