International Journal of Advanced Technology and Engineering Exploration, Vol 9(92)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2021.875907

Research Article

Anomaly detection in surveillance videos based on H265 and deep learning

Zainab K. Abbas* **and Ayad A. Al-Ani**
Department of Information and Communication Engineering, College of Information Engineering / Al-Nahrain University / Baghdad, Iraq

Received: 03-May-2022; Revised: 24-July-2022; Accepted: 26-July-2022

Abstract
This paper discusses anomaly detection, which is one of the most well-known applications of human activity recognition. Due to the ever-increasing activities posing risks ranging from planned aggression to harm caused by an accident, providing security to an individual is a major issue in any community today. Traditional closed-circuit television does not suffice since it necessitates a human being to be awake and always watch the cameras, which is costly. This necessitates the creation of an automated security system that detects anomalous activity in real-time and provides rapid assistance to victims. However, identifying activity from long surveillance footage takes time. Hence, in this research, we study the effect of the down-sampling concept of the challenging database, namely the university of central Florida (UCF Crime) using high efficiency video coding (HEVC)-H265 before feeding them into the anomaly detection system. This step reduced the size of the data, making it easier to store and transfer, and highlights the unique properties of each video clip. In the proposed work, first, we are down-sampling each video's frame into half by using H265 on the fast forward moving picture experts group (FFMPEG) platform, and then spatiotemporal features are extracted from a series of frames (frame level) using a pre-trained convolutional neural network (CNN) called Resnet50, then to boost the feature we are combining the features of every 15 video frames to generate a new feature vector that will be fed into the classifier model. The values of the new feature vectors represent the summation of the values of the original feature vectors obtained from Resnet50. Finally, the features obtained from a series of frames are fed to the bidirectional long short-term memory (BiLSTM) model, to classify the video as normal or abnormal. We conducted comprehensive tests on a different benchmark dataset for anomaly detection to verify the proposed framework's functionality in complex surveillance scenarios. The numerical results were carried out on the UCF crime dataset, with the proposed approach achieving an area under curve (AUC) score of 90.16% on the database's test set.

Keywords
Anomaly detection, Video surveillance system, BiLSTM, Deep learning, CNN.

1.Introduction
Shopping malls, banks, hospitals, markets, educational institutions, smart cities, and roadways are all places where video surveillance systems (VSS) are commonly used to improve public safety. The correctness and speed with which video anomalies are identified are usually the primary focus of security applications. Recently, many surveillance cameras have been installed at various locations around the world for public safety purposes. Massive volumes of video data are continuously generated by these cameras [1].

Real-time video analysis and anomalous case detection necessitate many human resources and are subject to mistakes due to a loss of human attention over time. Automatic anomaly detection technologies based on artificial intelligence (AI) mechanisms are required in surveillance systems because human observation is ineffective [2].

Different methodologies in the literature describe anomaly activities as "the occurrence of variance in regular patterns"[1]. Abnormal event recognition in surveillance recordings has a wide range of applications, including crime prevention, automated intelligent visual monitoring, and traffic security [3]. Video anomalous identification was formerly thought to be a one-class classification challenge due to the scarcity of real-world anomalous incidents [4−6] i.e., the classifier is trained on normal videos, and when

*Author for correspondence
This work was partially supported by the Department of Information and Communication Engineering, College of Information Engineering / Al-Nahrain University / Baghdad, Iraq.

abnormal patterns appear in the test set, a video is classed as anomalous [3]. Hence, it's impossible to collect all the typical events of real-world monitoring in a single dataset. As a result, different, normal behaviors may deviate from normal events in the training set, resulting in false alarms. Because of its importance, intelligent video anomaly detection systems have sparked a rush in research and applications in recent years. Anomaly detection in video surveillance, on the other hand, still faces several difficulties, including ambiguousness where anomaly detection is defined as the process of detecting events that are not expected to occur in each situation. However, in real-life settings, the line between normal and aberrant items is not always clearly defined. Some normal samples, for example, will exhibit unusual properties that are shared by abnormal occurrences, reducing model detection accuracy and dependency. Even though the term "anomaly" is mentioned in a lot of literature, there hadn't been a standard definition of it yet. Even the same occurrence is likely to have distinct qualities and differ greatly depending on the situation, sparsity, and diversity.

In real-world anomaly detection databases, positive samples (i.e., abnormalities) are substantially fewer than negative samples, in contrast to typical classification tasks. Supervised models are challenging to train when there is a data imbalance. Furthermore, real-world anomalous behaviors are varied and can't be fully depicted; in some cases, they may have yet to occur. As a result, considering all forms of probable anomalies in one model are unfeasible, the privacy of an individual will be violated if video surveillance data is made available for public access, especially if it includes facial and activity information. Because of this privacy feature, there aren't many open-source datasets available. Noise with the wide use of surveillance videos, cameras are routinely seen in places like crosswalks, elevators, restaurants, shopping malls, and even some private residences to improve safety. While existing imaging facilities may easily gather video surveillance data, manually labeling this information is a time-consuming and error-prone operation. Data noise will likely have an impact on model accuracy in the long run [7, 8]. Due to their wide intra/inter-class adaptability, lack of annotated data, and low resolution, recognizing anomalous events in surveillance cameras is particularly challenging, as these events are rarely correlated to normal appearance. Humans can recognize common and uncommon events based on intuition, whereas machines must rely on visual cues to do so. In general, stronger visual features outperform weaker visual features in terms of activity detection and recognition [1]. The majority of existing approaches suffer from a high percentage of false alarms. Additionally, while these strategies perform well on basic datasets, their effectiveness is restricted when applied to real-world circumstances. To address these challenges, we first coded the video before entering it into the anomaly detection model by using high efficiency video coding (HEVC) -H265 which is a video coding standard that was created to improve the display requirements of its predecessor H264. The primary goal of the HEVC project is to enable significantly improved compression performance in the range of 50% bit rate reduction for equal perceptual video quality compared to existing standards [9]. Then, instead of feeding one feature frame at a time, we combine the features of fifteen consecutive frames by taking the sum of their values and feeding them into our classifier model [10]. To train our classifier model, we use a weakly supervised strategy based on spatiotemporal data and bidirectional long short term memory (BiLSTM). BiLSTMs have proven to be very useful when the context of the input is required. Information moves from backward to forward in a unidirectional long short term memory (LSTM). On the other hand, BiLSTM uses two hidden states to flow information not only backward to forward but also forward to backward. As a result, BiLSTMs have a greater understanding of the context [11].

The main contributions of the current research are:

- Coding (down-sampling) each video using HEVC-H265.
- Extracting the spatial-temporal features by using pre-trained convolutional neural networks (CNN) namely Resnet50.
- For anomaly identification, we adopt the framework BiLSTM architecture.
- The evaluation of our work was done on the university of central Florida (UCF crime) dataset, which is a challenging benchmark dataset.

The remaining part of this document is arranged as follows: The second section looks at a literature assessment of existing approaches. The general recommended framework is explained in section 3. Section 4 assesses our research's experimental outcomes and compares them to existing methodologies. Section 5 provides a conclusion and recommendations for future research directions.

2.Related work

A variety of strategies have been explored to detect various types of abnormalities. All the strategies adopted, however, were tailored to a specific situation. A lot of models previously utilized for anomaly detection have been discussed in the following sections.

Chaudhary et al. [12] proposed a technique for automatically detecting multiple irregular activities in videos. Important features including centroid, direction, speed, and dimensions are identified throughout the feature extraction process. They introduced a new database with 45 films of different people walking, crawling, and running (without overlapping) for three different actions.

Bhagyalakshmi et al. [13] presented a technique for detecting weapons, live surveillance videos have been utilized to track and identify anomalies using approaches for real-time image processing. This system has 3 modules for processing: the first detects objects using CNN, the second identifies weapons, and the third manages monitoring and alert operations.

Sultani et al. [14] offered a framework that can detect unusual attitudes and inform the user of the kind of behavior. The deep multiple instance learning (MIL) systems are proposed in this research for learning anomalies from weakly labeled training movies, where the training labels (normal or anomalous) are applied at the video level rather than the clip level. This research utilized the UCF crime database, and the area under the curve (AUC) score was 75.41%.

Shine and CV [15] proposed a real-time automated system for identifying motorcycle riders not wearing helmets in videos of traffic surveillance and generating a new dataset because there are no datasets available. The system's experimentation and development dataset is made up of videos shot on roadways.

Shreyas et al. [16] proposed a new execution strategy in which films before being sent to the activity identification system are adaptively compressed. The study was conducted on the UCF101-crime dataset, and the AUC value obtained was 79.8%.

Combining the advantages of handcrafted and hierarchical feature learning to identify video outliers was suggested by Ramchandran and Sangaiah [17]. Raw and edge image sequences are mixed and fed into a convolutional auto-encoder and convolution LSTM model to detect the irregularity. The experiment was done at UCSD ped1, UCSD ped2, and Avenue datasets.

Anala et al. [18] identify a framework that can detect anomalous behavior and notify the user based on the kind of abnormality. The detection of anomalies is regarded as a regression problem. The performance of this technique was evaluated just on normal videos. The experiment was conducted on the UCF crime dataset, with an AUC value of 85%.

In addition to annotating the largest benchmark with more than 28K bounding boxes, Liu and Ma [19] provide an anomaly-region-guided framework that explicitly motivates the network to search for the area that represents the core of the anomaly. A meta-learning module is included in the training program to enhance generalization skills. The experiment was conducted on the UCF crime dataset, with an AUC value of 82%.

Weakly supervised anomaly detection using a temporal convolutional network (TCN) and inner bag loss, is proposed by Zhang et al. [20]. For the MIL method objective function optimization, outer bag ranking loss and inner bag loss work in tandem. By using the previous neighboring segment in the video without considering the future, the TCN directly encodes the temporal context and is appropriate for real-time anomaly identification. The experiment was conducted on the UCF crime dataset, with an AUC value of 78.66%.

To train a motion-aware feature, Zhu and Newsam [21] suggest a temporal augmented network, however, the model still struggles in some well-known difficult situations, such as low resolution, people grouping, fast motion, and dark images. The testing was performed on UCF crime with an AUC value of 79%.

Zhong et al. [22] take a new approach to weakly supervised anomaly detection by framing it as a supervised learning task with noisy labels. In addition, they use a graph convolutional network (GCN) to clean the labels to train an action classifier. The testing was performed on UCF crime, ShanghaiTech, and UCSD-Peds2 databases. The AUC value on UCF-Crime was equal to 82.12%.

Hao et al. [23] presented a two-stream convolutional network model that was unique. The proposed model

consists of two-stream flow and (red-green-blue) RGB networks, The completed anomaly activity recognition score is determined by the combined score. The detection of anomalies is regarded as a regression problem. The experiment was conducted using the UCF crime dataset, with an AUC value of 81.22%.

Venkatesh et al. [24] discussed a criminal detection method that could be used for on-device crime surveillance utilizing deep learning (DL). By making decisions on the device, they can lower the cost of gathering information into a centralized unit, reduce latency, and reduce the invasion of privacy. The testing was done on UCF Crime to provide reduced inference time using the notion of early-stopping–multiple instances learning and LSTM for anomaly identification (in addition to the normal video class, there are eight classes for anomalies).

The real world fighting (RWF)-2000 database, which consists of 2,000 films gathered by security cameras in actual situations, was presented by Cheng et al. [25]. They also suggest a novel strategy called flow gated network that merges the advantages of optical flow and 3D-CNNs.

Zaheer et al. [26] present a method for detecting anomalous events that can be trained using simply video-level labels that are weakly supervised. Using a batch-based training approach. Depending on the duration of the video, a batch may contain numerous chronologically ordered video segments, and a movie may be split into several batches. These batches are chosen in random order, which breaks inter-batch correlation and significantly improves performance. A normality suppression technique is also put forth, which works in combination with the backbone network to detect anomalies by training the network to learn to suppress the features that correlate to the video's normal sections. Additionally, a clustering distance-based loss is developed, which enhances the network's ability to more accurately represent both normal and abnormal events. The testing was performed on the UCF-Crime and ShanghaiTech databases with an AUC value of 83.03% for the UCF crime database.

Dubey et al. [27] suggested a deep-network with multiple ranking methods (DMRMs). The abnormal event identification trouble was likewise treated as a regression problem in this study. They employed 3D ResNet-34 to detect anomalies, and the testing was done on the UCF crime database, with an AUC value of 81.91%.

Ullah et al. [3] offered a lightweight CNN-based anomaly identification system that may be used in a surveillance system with low time difficulty. For outlier detection, they used a pre-trained lightweight CNN-multilayer BiLSTM, and the execution was done on UCF crime, with an AUC value of 78.43%.

Ullah et al. [1] offered an intelligent abnormal detection system based on deep features that can work in surveillance networks with minimal time difficulty. CNN ResNet50 with Multilayer BiLSTM was employed to detect anomalies in this study. The deployment was carried out on the UCF crime database, with an AUC of 85.53%.

Öztürk and Can [2] suggested using an anomaly detection network (ADNet), to find anomalies in videos by using temporal convolutions. The internet model operates by accepting a series of video clips. To accurately locate irregularities in videos, ADNet collects features from video clips and utilizes them in a window. The anomaly detection (AD) loss function is suggested to enhance the segment identification capability of the AD-anomalous Net. Additionally, they suggest using "F1@k," features extracted by inflated three-dimensional (I3D), and temporal shift module (TSM), the second feature extractor, for temporal anomaly identification. The implementation was carried out on UCF crime, and they also added the anomaly categories "Molotov bomb" and "protest" to the data set.

Wu et al. [28] suggested extracting the movable targets in the movie using a Gaussian model and then using a pre-trained model called VGG16 for feature extraction purposes. The UCSD dataset was used to train and forecast MIL models at the pixel level using normalized set kernel and multiple instance support vector machines approaches.

Aziz et al. [29] describe how to identify motion-based abnormal events when a one-class support vector machine is utilized. Using a mask recurrent convolution neural network (RCNN), which delivers object masks at pixel level as well as object class identification and regression of bounding box, the suggested technique identifies abnormalities at the frame level and then locates them at the pixel level in the identified anomalous frames. The Avenue and UMN datasets were used to test the suggested framework.

Boekhoudt et al. [30] developed a human related crime (HR-crime) dataset, a subgroup of the UCF crime database, suited for tasks involving the recognition of human-related anomalies, and constructed the feature extraction pipeline for such tasks using cutting-edge methods. Additionally, they provide the HR-crime baseline outlier detection analysis.

Wan et al. [31] provide a new large-scale anomaly detection (LAD) dataset as a baseline for identifying anomalous video sequences. It has 2000 video sequences with both abnormal and normal film clips, as well as 14 categories of anomalies like crashes, fire, and violence. To detect anomalies, it also contains annotation data, such as video and frame level labels (normal and abnormal). The anomaly detection problem is formulated as a fully supervised learning problem, and a multi-task deep neural network is suggested to tackle it.

Zaheer et al. [32] offered a weakly supervised abnormality identification technique that relies on video level labels to train a feature extractor model like convolution three dimensional (C3D), k-mean, or the fully connected network. The testing was done on UCF crime, and the AUC value achieved was 78.27%.

Majhi et al. [33] suggested a technique that uses a weakly supervised learning model to handle abnormality detection and classification in a unified model. For outlier detection, I3D many to many LSTM was employed, and the development was performed on UCF crime, with an AUC value of 82.12%.

Wu et al. [34] proposed a dual branch network using multi-detail concepts in both the temporal and spatial dimensions as input. C3D was used to extract the features, and namely spatio-temporal (ST-UCF-crime) was used to implement them on a new dataset (ST-UCF-crime) that annotates spatial-temporal bounding boxes for unusual occurrences in UCF crime. The ST-UCF-Crime dataset had an AUC of 87.65%.

To significantly increase the robustness of the MIL strategy to the negative instances from anomalous videos, Tian et al. [35] propose the new method known as Robust Temporal Feature Magnitude (RTFM) Learning. RTFM involves training a feature magnitude learning function to effectively recognize the positive instances. the video snippets' temporal feature magnitude was used to perform RTFM, where features with low magnitude imply normal (i.e., negative) snippets and high magnitude features denote aberrant (i.e., positive) snippets. The testing was performed on (ShanghaiTech, UCF-Crime, XD-Violence, and UCSD-Peds. The AUC score on the UCF crime database was equal to 84.30% and 83.28% by using I3D and C3D respectively.

To detect video anomalies, Cao et al. [36] suggested an adaptive GCN. The suggested method builds a global graph while taking the similarity of features and temporal differences into account. Additionally, a graph learning layer is used to build connections between video segments adaptively, which may effectively capture spatial-temporal correlations between video segments and improve existing temporal characteristics. The testing was performed on UCF-Crime and ShanghaiTech databases. The AUC value on the UCF crime was 83.14%.

A comprehensive review of the techniques and databases that have recently been utilized for anomaly detection purposes, was introduced by Abbas and Al-Ani [8]. Besides, a comparative study on various techniques that have been used for anomaly detection. Through this study, we find that the DL has surpassed alternative techniques in this area.

3.Methods

From a comparison study on the various approaches that have been performed which are utilized for anomaly detection, it was found that DL has surpassed alternative techniques in this area [8]. The proposed methodology followed in this work can be shown in *Figure 1*.

The proposed methodology is divided into three stages:

- Video coding using HEVC-H265
- Feature extraction using Resnet50
- BiLSTM training and classification for anomaly event detection.

This work is focused on the evaluation of the DL algorithms for anomaly detection purposes on the UCF-Crime dataset after down-sampling using HEVC. *Figure 2* illustrates the suggested framework. *Figure 3* shows the comprehensive workflow of the proposed system.

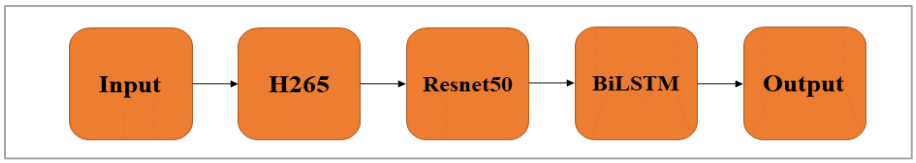

Figure 1 The methodology Stages

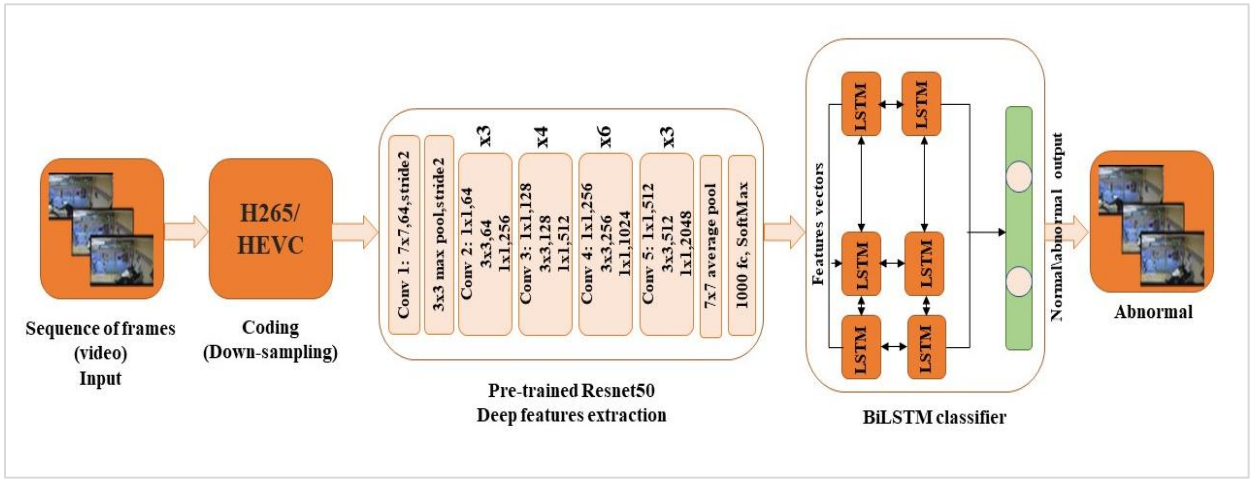

Figure 2 The suggested framework

3.1Input dataset UCF-Crime

The entire database in this work was the database of the UCF crime [37, 38]. The UCF-Crime database includes videos of both abnormal and normal events, as well as 13 various types of anomalies, such as fights, explosions, abuse, and accidents. There are 1900 surveillance movies in the collection, with an approximately equal amount of abnormal and normal videos. The training set contained 810 abnormal and 800 normal samples, while the testing set contains the remaining 140 abnormal and 150 normal movies [14, 39]. This collection has over 129 hours of videos at a 320x240 resolution, 13 million frames, these videos are different in their length [14,37,39]. The anomalies in this dataset have a significant impact on public safety, hence we chose it because it contains various abnormal event types. However, this dataset has two drawbacks the first one is video level labels, which means that we just know that each video contains an anomaly, but do not know which specific segment is an anomaly; the second is this dataset's anomalous class has enormous inter-class variances, which lead to overfitting [37]. We took the video with a length equal to or less than 2 min relying on this condition. We had 1324 videos: 1116 videos for the training stage (90% for training purposes and 10% for validation purposes) while the testing was done on 208 videos.

3.2High-efficiency video coding (HEVC)-H265

The joint collaborative team on video coding (JCTVC) is a collaboration between the telecommunications (ITU-Ts) and technology industries (ISO/IEC MPEG) to develop (HEVC) (JCTVC). It is referred to as H.265 in ITU-T recommendations and MPEG-H part 2 in ISO/IEC standards. HEVC was created to improve coding efficiency in high and ultra-high-resolution videos. In HEVC, this is accomplished through a computationally complicated encoding method with numerous possibilities for improving efficiency. One of the benefits of HEVC is the expanded range of block sizes available when splitting a frame into blocks, ranging from 4×4 to 64×64 [40]. In recent years, growth in video resolution, like 4K or 8K ultra-high-definition (UHD), has necessitated the development of a video coding standard that is far more efficient than H.264. The HEVC video coding standard is the most recent enhanced video coding standard, designed to improve the display requirements of its predecessor. According to H.265, the major purpose of the HEVC is to allow much-improved compression performance compared to existing standards, with up to a 50% bit-rate reduction for equal perceptual video quality [9, 40]. In this work, each video was coding by using HEVC by using the fast forward moving picture experts group (FFMPEG) platform and by using libx265 (the

encoding library for generating HEVC video streams with rate control mode conditional random field (CRF) equal to 28. This is the recommended rate control mode for most users to keep the best quality and care less about the file size. This step reduced the size of the data, making it easier to store and transfer, and highlights the unique properties of each video clip, which facilitates the discovery of unusual events, which is the primary goal of this research.

3.3Center crop

In this work, we used a pre-trained network for features extraction, namely ResNet50, the size of the input to ResNet50 is 224×224. For this reason, a center crop has been used in this work to crop the longest edges of a video and resizes them to have a size equal to the size of the input.

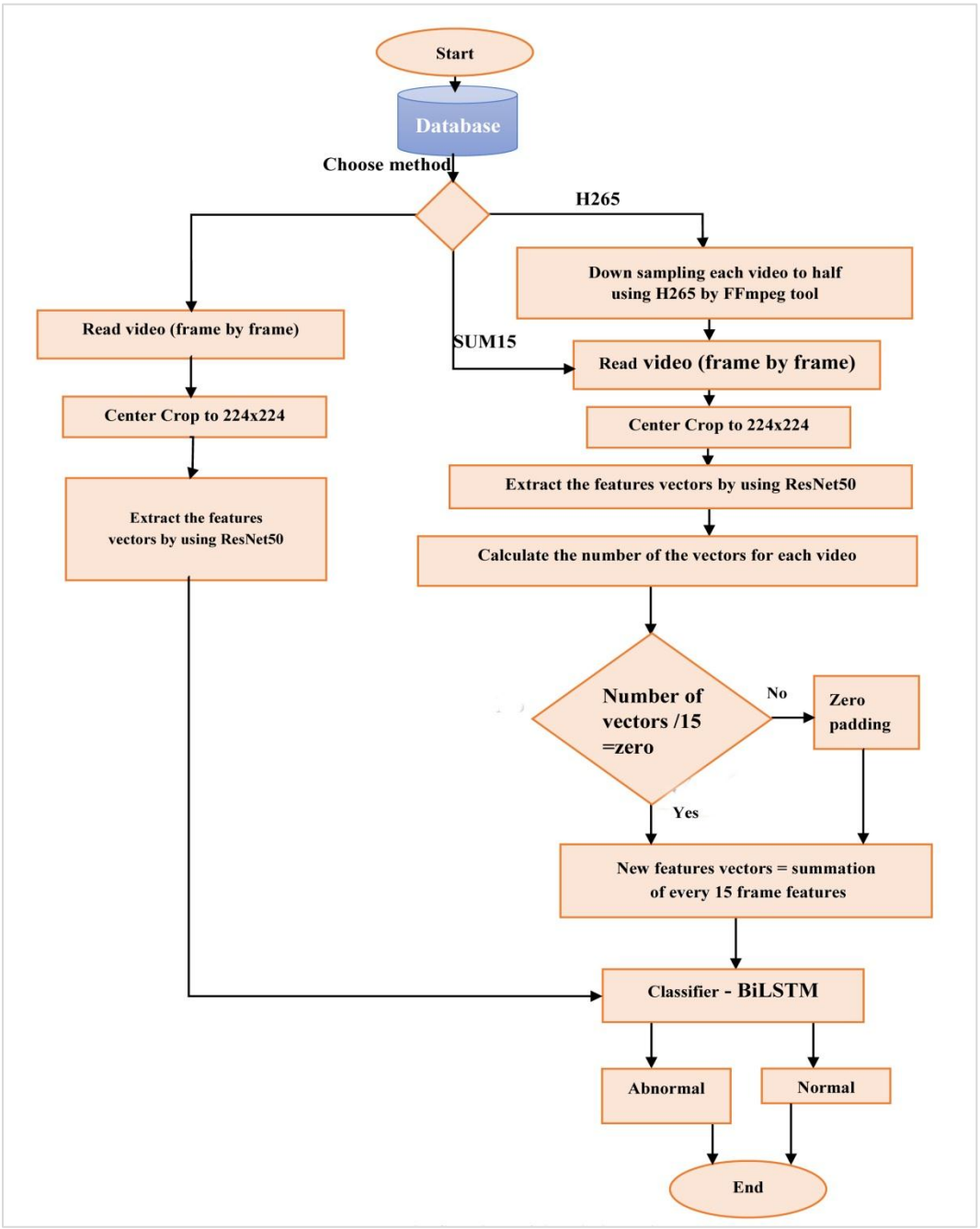

Figure 3 The flowchart of the whole work

3.4 Deep learning (DL)

DL is a subset of machine learning that succeeds in handling unstructured data. DL methods outperform existing machine learning methods. It enables computational models to know and understand the features from data at multiple levels, step by step. Deep learning gained popularity as the amount of data available increased, as did the improvement of hardware that allowed for powerful computers. The input is sent through several levels of the deep learning algorithm, each of which can extract features and pass them on to the next layer. Initial layers extract low-level information, which is then combined in subsequent layers to build a comprehensive representation. Different designs, such as recurrent neural networks (RNN), CNN, unsupervised pre-trained networks, and so on, can be used to achieve deep learning. When compared to standard learning approaches, the effectiveness of deep learning classifiers significantly improves as the number of data increases. When typical machine learning algorithms reach a certain amount of training data, their performance stabilizes, whereas deep learning improves as the amount of training data increases. Deep learning architectures outperform simple Artificial Neural Networks (ANNs), although deep structures take longer to train. However, approaches such as transfer learning and graphics processing unit (GPU) computing can reduce training time [8, 41]. CNN is a type of neural network that includes convolutional layers. Although CNN is good at processing spatial data, RNN is superior at handling sequential data. RNN uses state variables to store past data and uses it in conjunction with current inputs to determine present outputs [42]. CNN's are mostly employed in image processing. It assigns biases and weights to different objects in the image and distinguishes them. In comparison to other classification methods, it requires less preparation. To capture the spatial and temporal connections in an image, CNN employs relevant filters. Some types of CNN are AlexNet, ResNet, LeNet, VGGNet, GoogleNet, and ZFNet [41]. On the other hand, in RNNs, the outputs from the past are given as inputs to the present state. RNN's hidden layers can remember knowledge. The output created in the past state was used to upgrade the hidden state. RNNs can be used to predict time series because they have a memory that allows them to remember prior inputs. The LSTM is one type of RNN [41]. In our work, we used BiLSTM for anomaly identification purposes, where BiLSTM contains two LSTMs, one attempting to take input in one direction and the other in the opposite direction (forward and backward), while for feature extraction purposes, we used ResNet50, where ResNet50 is a 50-layer deep convolutional neural network with 23.5 million trainable variables [1].

4.Experimental results and discussions

The computer codes for the proposed work were implemented in the FFMPEG platform for the first part of the work which includes the down-sampling of each video and in the MATLAB software environment (version 2021a) for the second part of the work which contains deep features extraction and anomaly detection, on Windows 10, 1 TB SSD hard drive, Intel Core i7 processor, 16 GB RAM, 64-bit operating system, and NVIDIA GeForce MX450 graphics processing unit.

4.1Input dataset UCF-crime

The experimentation for this work was done on the UCF-Crime dataset for 13 anomaly classes in addition to the normal class. In this work, the anomaly identification is done at the video level that is mean the length of the video doesn't affect the anomaly detection operation so we took the video at a length equal to or less than 2 minutes. Based on this condition, we had: 1116 videos for the training stage and 208 videos for testing. The best partition which gives the best performance during the training stage was 90% for training and 10% for validation.

4.2High-efficiency video coding (HEVC) -H265

The size of the original dataset was 97.6 GB, the frame rate was 30 frames/second, and the size of the frame was 240×320. After applying HEVC, the size became 3.79 GB, the frame rate was 30 frames/second, and the size of the frame became 120x160. Despite the success of the proposed work in reducing the required storage space, it affected the process of separating the normal event from the abnormal event. Although the scale of AUC has increased compared to the existing work in this field when using HEVC, in contrast with the results got from using the original data, the results were not as expected, this may have happened due to the compression ratio used.

4.3Center crop

The resize method used in this work was bi-cubic, which uses 16 (4×4 neighborhood) to determine the output. This step is considered a pre-processing step for the UCF-Crime database, the method chosen in this work for resizing the video’s frame show improvement in the performance of the classifier model compared with the existing model in terms of

AUC score for the UCF-Crime database unlike down-sampling the UCF-Crime database, this may have happened due to the method used for resizing.

4.4Deep learning (DL)

In this work, a pre-trained ResNet50 for feature extraction purposes was used, this neural network takes each frame from the video as input and returns 1000 features for each frame [1]. The extracted features are taken out from the fc1000 layer, then before feeding the features to the classifier model, we combine the features of fifteen consecutive frames by taking the summation of their values and generating new vectors [10], which feed into the classifier model BiLSTM for anomaly identification purposes. The BiLSTM classifier model has been applied to the UCF-Crime database with the parameters shown in *Table 1*, the optimizer used in the whole work was Adam. The model parameters have been chosen using trial and error and the value of dropout and L2Regularization have been chosen to reduce the overfitting of the model. The loss function evaluation during the training stage is demonstrated in *Figure 4.*

Table 1 The parameter values of the classifier model

Classifier parameters	Methods		
	SUM15 [10]	Original	H265
Minimum Batch-Size	16	8	16
Hidden layer nodes No.	450	80	450
Dropout	0.7	0.7	0.7
Initial Learning Rate	1e-5	1e-4	1e-5
Maximum epochs	100	60	100
L2Regularization	0.5	0.5	0.5

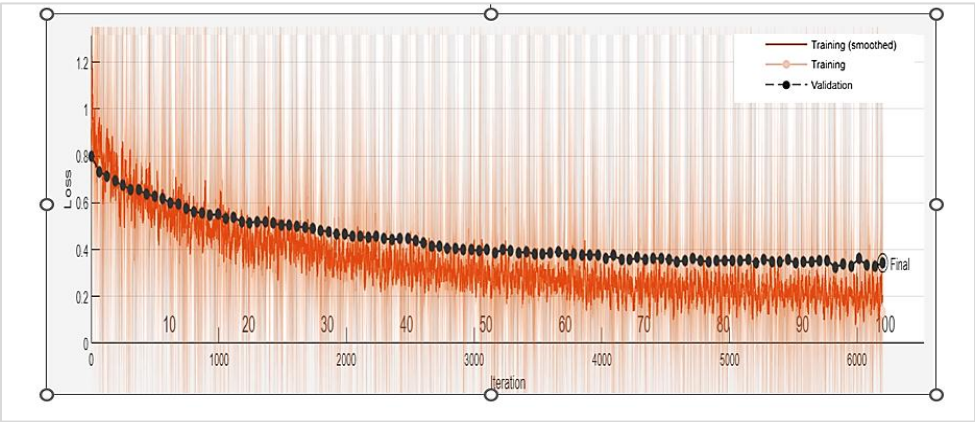

Figure 4 The loss function evaluation during the training stage

Following the previous work, the receiver operating characteristics (ROC) and the AUC were utilized as performance indicators to evaluate the performance of the proposed work and a compression method. The confusion matrix of our classifier model can be seen in *Figure 5*. The false negative (FN) which means anomaly classified as normal was equal to 8, false positive (FP) which means normal classified as an anomaly was equal to 34, and true positive (TP) which means anomaly classify as anomaly was equal to 85 and true negative (TN) which mean normal classify as normal was equal to 81, as shown in the confusion matrix of the model. The ROC Curve of our classifier model is shown in *Figure 6*. Furthermore, we calculated the detection accuracy of our classifier, and it was 79.81%.

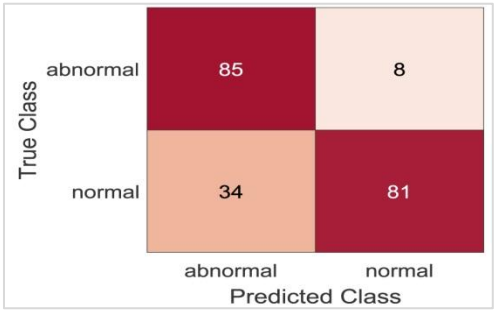

Figure 5 The Confusion matrix of our classifier model, Detection accuracy = 79.81%

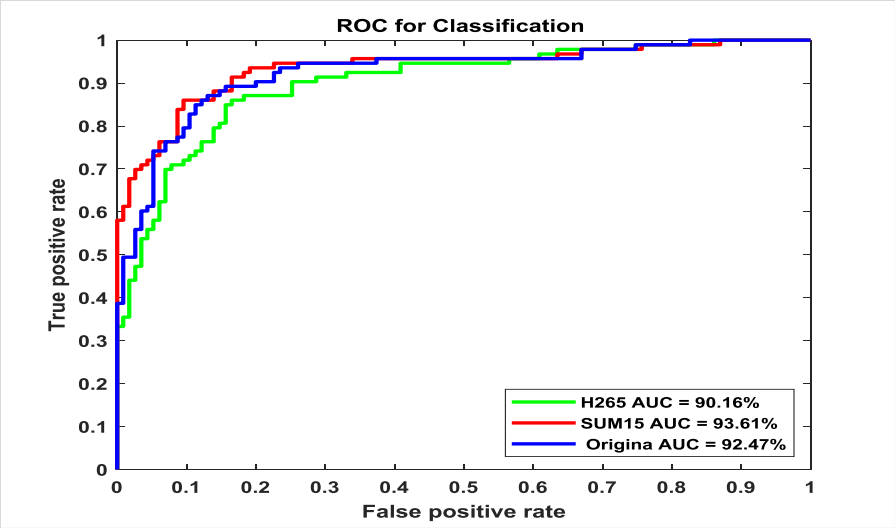

Figure 6 The ROC curve of our classifier model

Table 2 compares the AUC scores with existing methods, and it is clear that our approach reached the best AUC of 90.16 %, a 2.51% increase over other existing methods, whenever the value of AUC is close to 1, this means the model has good separability between normal and anomaly. *Table 3* compares the AUC scores and detection accuracy for all methods used in this work. Despite the success of the proposed work in reducing the required storage space, it affected the process of separating the normal event from the abnormal event. Although the scale of AUC has increased compared to the existing work in this field when using HEVC, in contrast with the results got from using the original data, the results were not satisfactory.

A complete list of abbreviations is shown in *Appendix I.*

Table 2 Comparison of AUC score of proposed work with the state-of-the-art techniques

Method	**AUC % on UCF-Crime database**
Sultani et al. , 2018 [14]	75.41
Anala et al. 2019 [18]	85
Liu and Ma 2019 [19]	82
Zhang et al. 2019 [20]	78.66
Zhu and Newsam 2019 [21]	79
Zhong et al. [22]	82.12
Shreyas et al. 2020 [16]	79.8
Hao et al. 2020 [23]	81.22
Zaheer et al. 2020 [26]	83.03%
Dubey et al. 2021 [27]	81.91
Ullah et al. 2021 [3]	78.43
Ullah et al. 2021 [1]	85.53
Zaheer et al. 2021 [32]	78.27
Majhi et al. 2021 [33]	82.12
Wu et al. 2018 [34]	87.65
Tian et al. 2021 [35]	84.30%
Cao et al. 2022 [36]	83.14%
Ours Proposed Method	90.16

Table 3 Compares the AUC scores and detection accuracy for all methods used in this work

Method	**AUC %**	**Detection accuracy %**
Center crop + SUM15 [10]	93.61	86.06
Center crop	92.47	83.17
H265 + Center crop + SUM15	90.16	79.81

5.Conclusions and future work

Based on current anomaly datasets, we provide an effective model for real-world outlier detection in a surveillance system with state-of-the-art accuracy. First, each video was coded using HEVC; secondly, using pre-trained Resnet50, the feature vector was extracted for each video; finally, the feature vector was fed into BiLSTM for normal and abnormal class identification. Despite the success of the proposed work in reducing the required storage space, it affected the process of separating the normal event from the abnormal event. Although the scale of AUC has increased compared to the existing work in this field when using HEVC, for the UCF-Crime dataset after down-sampling, the experimental findings show an increase in the AUC value of up to 90.16%, an increase of 2.51% compared to other existing methods. Furthermore, we calculated the detection accuracy of our classifier, and it was 79.81%. And this illustrates the success of our proposal in improving the accuracy of anomaly event detection, but in contrast with the results got from using the original data in the same steps, the results were not satisfactory. In the future, we will work to increase the accuracy indicator by using another compression ratio for HEVC. Other methods related to resizing, feature selection algorithms and dimensionality reduction algorithms can be combined with the suggested framework in the near future.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

Author's contribution statement

Zainab K. Abbas: Methodology, software design, analysis and interpretation of results, paper draft writing, review and editing. **Ayad A. Al-Ani:** Analysis and interpretation of results, review and supervision.

References

[1] Ullah W, Ullah A, Haq IU, Muhammad K, Sajjad M, Baik SW. CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. Multimedia Tools and Applications. 2021; 80(11):16979-95.

[2] Öztürk Hİ, Can AB. ADNet: temporal anomaly detection in surveillance videos. In international conference on pattern recognition 2021 (pp. 88-101). Springer, Cham.

[3] Ullah W, Ullah A, Hussain T, Khan ZA, Baik SW. An efficient anomaly recognition framework using an attention residual LSTM in surveillance videos. Sensors. 2021; 21(8):1-17.

[4] Morais R, Le V, Tran T, Saha B, Mansour M, Venkatesh S. Learning regularity in skeleton trajectories for anomaly detection in videos. In proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019 (pp. 11996-2004).

[5] Dos SFP, Ribeiro LS, Ponti MA. Generalization of feature embeddings transferred from different video anomaly detection domains. Journal of Visual Communication and Image Representation. 2019; 60:407-16.

[6] Fan Y, Wen G, Li D, Qiu S, Levine MD, Xiao F. Video anomaly detection and localization via gaussian mixture fully convolutional variational autoencoder. Computer Vision and Image Understanding. 2020.

[7] Ren J, Xia F, Liu Y, Lee I. Deep video anomaly detection: opportunities and challenges. In international conference on data mining workshops 2021 (pp. 959-66). IEEE.

[8] Abbas ZK, Al-ani AA. A Comprehensive review for video anomaly detection on videos. In international conference on computer science and software engineering 2022 (pp. 1-1). IEEE.

[9] Nouripayam M, Sheikhipoor N. HEVC (H. 265) Intra-Frame prediction implementation using MATLAB. 2014.

[10] Khaire P, Kumar P. 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. 2022.

[11] Sharfuddin AA, Tihami MN, Islam MS. A deep recurrent neural network with bilstm model for sentiment classification. In international conference on Bangla speech and language processing 2018 (pp. 1-4). IEEE.

[12] Chaudhary S, Khan MA, Bhatnagar C. Multiple anomalous activity detection in videos. Procedia Computer Science. 2018; 125:336-45.

[13] Bhagyalakshmi P, Indhumathi P, Bhavadharini LR. Real time video surveillance for automated weapon detection. International Journal of Trend in Scientific Research and Development. 2019; 3(3).

[14] Sultani W, Chen C, Shah M. Real-world anomaly detection in surveillance videos. In proceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 6479-88).

[15] Shine L, CV J. Automated detection of helmet on motorcyclists from traffic surveillance videos: a comparative analysis using hand-crafted features and CNN. Multimedia Tools and Applications. 2020; 79(19):14179-99.

[16] Shreyas DG, Raksha S, Prasad BG. Implementation of an anomalous human activity recognition system. SN Computer Science. 2020; 1(3):1-10.

[17] Ramchandran A, Sangaiah AK. Unsupervised deep learning system for local anomaly event detection in crowded scenes. Multimedia Tools and Applications. 2020; 79(47):35275-95.

[18] Anala MR, Makker M, Ashok A. Anomaly detection in surveillance videos. In 26th international

conference on high performance computing, data and analytics workshop (HiPCW) 2019 (pp. 93-8). IEEE.
[19] Liu K, Ma H. Exploring background-bias for anomaly detection in surveillance videos. In proceedings of the 27th ACM international conference on multimedia 2019 (pp. 1490-9).
[20] Zhang J, Qing L, Miao J. Temporal convolutional network with complementary inner bag loss for weakly supervised anomaly detection. In international conference on image processing 2019 (pp. 4030-4). IEEE.
[21] Zhu Y, Newsam S. Motion-aware feature for improved video anomaly detection. arXiv preprint arXiv:1907.10211. 2019.
[22] Zhong JX, Li N, Kong W, Liu S, Li TH, Li G. Graph convolutional label noise cleaner: train a plug-and-play action classifier for anomaly detection. In proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019 (pp. 1237-46).
[23] Hao W, Zhang R, Li S, Li J, Li F, Zhao S, et al. Anomaly event detection in security surveillance using two-stream based model. Security and Communication Networks. 2020.
[24] Venkatesh SV, Anand AP, Gokul SS, Ramakrishnan A, Vijayaraghavan V. Real-time surveillance based crime detection for edge devices. In VISIGRAPP (4: VISAPP) 2020 (pp. 801-9).
[25] Cheng M, Cai K, Li M. RWF-2000: an open large scale video database for violence detection. In 25th international conference on pattern recognition 2021 (pp. 4183-90). IEEE.
[26] Zaheer MZ, Mahmood A, Astrid M, Lee SI. Claws: clustering assisted weakly supervised learning with normalcy suppression for anomalous event detection. In European conference on computer vision 2020 (pp. 358-76). Springer, Cham.
[27] Dubey S, Boragule A, Gwak J, Jeon M. Anomalous event recognition in videos based on joint learning of motion and appearance with multiple ranking measures. Applied Sciences. 2021; 11(3):1-21.
[28] Wu G, Guo Z, Wang M, Li L, Wang C. Video abnormal event detection based on CNN and multiple instance learning. In twelfth international conference on signal processing systems 2021 (pp. 134-9). SPIE.
[29] Aziz Z, Bhatti N, Mahmood H, Zia M. Video anomaly detection and localization based on appearance and motion models. Multimedia Tools and Applications. 2021; 80(17):25875-95.
[30] Boekhoudt K, Matei A, Aghaei M, Talavera E. HR-crime: human-related anomaly detection in surveillance videos. In international conference on computer analysis of images and patterns 2021 (pp. 164-74). Springer, Cham.
[31] Wan B, Jiang W, Fang Y, Luo Z, Ding G. Anomaly detection in video sequences: a benchmark and computational model. IET Image Processing. 2021; 15(14):3454-65.
[32] Zaheer MZ, Lee JH, Astrid M, Mahmood A, Lee SI. Cleaning label noise with clusters for minimally supervised anomaly detection. arXiv preprint arXiv:2104.14770. 2021.
[33] Majhi S, Das S, Brémond F, Dash R, Sa PK. Weakly-supervised joint anomaly detection and classification. In IEEE international conference on automatic face and gesture recognition 2021 (pp. 1-7). IEEE.
[34] Wu J, Zhang W, Li G, Wu W, Tan X, Li Y, et al. Weakly-supervised spatio-temporal anomaly detection in surveillance video. arXiv preprint arXiv:2108.03825. 2021.
[35] Tian Y, Pang G, Chen Y, Singh R, Verjans JW, Carneiro G. Weakly-supervised video anomaly detection with robust temporal feature magnitude learning. In proceedings of the IEEE/CVF international conference on computer vision 2021 (pp. 4975-86).
[36] Cao C, Zhang X, Zhang S, Wang P, Zhang Y. Adaptive graph convolutional networks for weakly supervised anomaly detection in videos. arXiv preprint arXiv:2202.06503. 2022.
[37] Maqsood R, Bajwa UI, Saleem G, Raza RH, Anwar MW. Anomaly recognition from surveillance videos using 3D convolution neural network. Multimedia Tools and Applications. 2021; 80(12):18693-716.
[38] https://www.dropbox.com/sh/75v5ehq4cdg5g5g/AABvnJSwZI7zXb8_myBA0CLHa?dl=0. Accessed 15 April 2022.
[39] Kumari P, Bedi AK, Saini M. Multimedia datasets for anomaly detection: a review. arXiv preprint arXiv:2112.05410. 2021.
[40] Hassan KH, Butt SA. Motion estimation in HEVC/H. 265: metaheuristic approach to improve the efficiency. Engineering Proceedings. 2021; 12(1):1-4.
[41] Mathew A, Amudha P, Sivakumari S. Deep learning techniques: an overview. In international conference on advanced machine learning technologies and applications 2020 (pp. 599-608). Springer, Singapore.
[42] Zhang A, Lipton ZC, Li M, Smola AJ. Dive into deep learning. arXiv preprint arXiv:2106.11342. 2021.

Zainab K. Abbas, Born in Baghdad/Iraq, in 1993. She gained B.Sc. and M.Sc., from Department of Control and System Engineering/ University of Technology, in 2015 and 2018, respectively. She is currently a Ph.D. student in the Department of Information and Communication Engineering/College of Information Engineering/Al-Nahrain University. Her works studies are focused on Digital processing with Artificial Intelligence.
Email: zainabkudair@gmail.com

Ayad A. Al-Ani, Born in Baghdad/Iraq, in 1961. He gained the B.Sc., M.Sc., and Ph.D., from Department of Physics/College of Science/ Baghdad University, in 1983, 1990, and 1995, respectively. Previously, he was University Vice President for Administrative affairs /Al-Nahrain University, Deputy Dean for Post Graduate Studies and Scientific Research/Baghdad University, and Head of Space and Astronomy Department/College of Science/Baghdad University. Since 2006, he is a Professor of Digital Image Processing in Department of Information and Communication Engineering/College of Information Engineering/Al-Nahrain Univ. He published 74 papers and 4 books
Email: ayad.a@nahrainuniv.edu.iq

Appendix I

S. No.	Abbreviation	Description
1	AD	Anomaly Detection
2	ADNet	Anomaly Detection Network
3	AI	Artificial Intelligence
4	ANNs	Artificial Neural Networks
5	AUC	Area Under Curve
6	BiLSTM	Bidirectional Long Short-Term Memory
7	C3D	Convolution 3D
8	CNN	Convolution Neural Network
9	CRF	Conditional Random Field
10	DL	Deep Learning
11	DMRMs	Deep-network with Multiple Ranking Methods
12	FFMPEG	Fast Forward Moving Picture Experts Group
13	FN	False Negative
14	FP	False Positive
15	GCN	Graph Convolutional Network
16	GPU	graphics processing unit
17	HEVC	High-Efficiency Video Coding
18	HR-Crime	Human Related Crime
19	I3D	Inflated Inflated Three Dimensional
20	JCTVC	Joint Collaborative Team on Video Coding
21	LAD	Large-scale Anomaly Detection
22	LSTM	Long Short-Term Memory
23	MIL	Multiple Instance Learning
24	RCNN	Recurrent Convolution Neural Network
25	RGB	Red-Gree-Blue
26	RNN	Recurrent Neural Networks
27	ROC	Receiver Operating Characteristics
28	RTFM	Robust Temporal Feature Magnitude
29	RWF	Real World Fighting
30	ST	Spatio-Temporal
31	TCN	Temporal Convolutional Network
32	TN	True Negative
33	TP	True Positive
34	TSM	Temporal Shift Module
35	UCF	University of Central Florida
36	UHD	Ultra-High-Definition
37	VSS	Video Surveillance Systems