International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

ODSAC: An Innovative Approach for Detection of Suspicious Human Activity and Crime Prediction

Ms. A. M. Bhugul-Rajurkar[1], Dr. V. S. Gulhane[2]

Submitted: 27/12/2023 **Revised**: 03/02/2024 **Accepted**: 11/02/2024.

Abstract: The escalating crime rate has spurred a surge of research into real-time object detection for video surveillance. An automatic video detection system is required since it is challenging to continuously watch camera footage taken in public areas in order to identify any unusual activity. Real-time systems find it challenging to identify suspicious or small moving objects against a blurry background. In order to identify suspicious human activity in real-time video, this research introduces a novel algorithm dubbed ODSAC (Object Detection And Suspicious Activity Classification). Darknet53 is used by the system as an object-detecting tool. We have designed a custom classifier that accurately distinguishes whether activity is suspicious or not. The classifier is trained on a self-created dataset to ensure robust performance across various environmental conditions and scenarios. Extensive experimentation on the self-created dataset validates the effectiveness of the proposed approach. The evaluation metrics, including precision, recall, and F1 score, showcase the system's high accuracy in detecting individuals with weapons. The achieved detection accuracy of 99.31% underscores the reliability and efficiency of the architecture with the custom classifier. We have also done a comparative analysis with some existing Methods studied during the literature study. This research paper will be useful for applications such as Criminal Justice to identify suspicious criminal behaviors, Healthcare, Law Enforcement, etc. Integrating this technology with the current monitoring infrastructure might have a significant impact, especially in important locations like airports, schools, and public areas.

Keywords: *Machine Learning, Suspicious Activity, Classifier, Object Detection, Neural Network*

1. Introduction

Video surveillance systems are now widely available to guarantee human safety. The principal aim of putting these systems into place is to efficiently keep an eye on the regions that have been assigned, which will allow for the detection of any possible dangers or security risks. [1]. Nowadays, most systems depend on human operators to monitor these images. Every operator is also in charge of concurrently managing the video streams from many security cameras. Due to this reality and established human characteristics like fatigue or a lack of focus over time, people may fail to notice important events or respond too late, which could cause catastrophic injury and render current video surveillance systems inefficient or useless. [2,3]. Given the rise in violent crimes, particularly in public places like shopping centers, airports, banks, and ATMs, enhanced security measures are absolutely essential. By tracking and identifying possible security threats, closed-circuit television (CCTV) surveillance has become a widely used tool to improve public safety. The development of automated systems has garnered significant attention as a result of the labor-intensive and error-prone nature of manual surveillance. These technologies are a potential way to improve security monitoring accuracy and efficiency by looking for weapons in real-time CCTV footage. A recent terror attack in J&K, the Rajouri Terror Attack in January 2023, claimed seven lives and injured a few more. A man killed six people in a mass shooting in Mississippi in February 2023 [30]. In Punjab, India, thieves stole Rs. 22 lakhs from a bank in 2023 while holding a gun to their victims[29]. In January 2023, a gun-wielding man in California murdered eleven innocent people and injured nine more [31].

1 Department of Computer Sci. and Engineering, Sipna College of Engineering &Technology, Amravati-444701, India

ORCID ID : 0000-0001-5142-

2 Professor, Department of Information Technology, Sipna College of Engineering and Technology, Amravati - 444701, India

** Corresponding Author Email: ashwinibhugul@gmail.com/ashwinirajurkar593@gmail.com*

Regrettably, there is an increase in the frequency of dangerous and firearm-related incidents. These scenarios include terrorist attacks, shootings on school property, handgun attacks, and mass shootings [4,5]. Many of these incidents may be avoided or greatly decreased if potentially dangerous objects or suspicious activities were quickly identified in real-time CCTV footage.

Algorithms utilizing machine learning techniques can detect forbidden objects in luggage or discern questionable conduct in congested airport areas. Models for machine learning can learn continuously and are flexible. Several state-of-the-art algorithms have been developed for complex object detection such as YOLO (You Only Look Once)[9], Faster R-CNN (Region-based Convolutional Neural Network)[10], SSD (Single Shot Multibox Detector)[9], Mask R-CNN[10], RetinaNet[9], Cascade R-CNN[10], etc. When it comes to accurately and efficiently identifying objects in images, these algorithms are most widely used.

The area of automatic object detection has been thoroughly researched recently, and a number of methods have been put out in published works. As Deep Learning (DL) gains traction, previous methods based on traditional computer vision[9] techniques are gradually being replaced by automatic object detection approaches based on such paradigms, which outperform them in terms of speed and reliability. While many algorithms have been developed to identify different human motions including running[34], walking[34], boxing, kicking, punching, crawling, jumping, and so on, the problem of real-time weapon recognition in video still has to be solved. Small moving objects in movies are hard for many state-of-the-art approaches to detect, especially when the background is blurry. Even with the most advanced techniques, it is still difficult to identify small moving objects in videos, and unclear backgrounds can cause many algorithms to malfunction. While there are many methods that perform well on standard datasets, real-time datasets are not used.

Multiple object detection in a single frame is one of the main issues in object as a weapon detection[27]. To detect several firearms in a single frame of a static image or video, multiple weapons must be located and identified. Occlusions, fluctuating lighting, weapon size, and the requirement to distinguish weapons from other things make this task extremely challenging. This paper is based on implementing a new approach for suspicious human activity such as a human with a weapon such as a gun or a human wearing a helmet from one of the categories of humans with objects. We have focused on the implementation of multiple weapon(Guns) detection in a single frame along with the implementation of a custom classifier. The custom classifier classifies the activity in two categories. If the classifier classified it correctly an alert call is going to be sent to the authorized person along with the crime location. We have designed an Android application for the same. As of now for our prototype, we have taken objects such as guns of different shapes and sizes and a person wearing a helmet. We can change the dataset and train the model for any of the suspicious human activity categories mentioned above. We have implemented the architecture with Darknet-53[9] and Dense net Architecture to improve the training process and to avoid over fitting.

1.1 Darknet -53 for object detection

The neural network architecture known as Darknet-53[9] was created with object identification, classification, and localization in mind. It is a component of the open-source Darknet framework, which is a C and CUDA neural network framework[ ]. It refers to an open-source neural network framework that has been very popular recently because of its real-time object identification speed and accuracy. The prominent real-time object recognition system YOLO (You Only Look Once) version 3 [9] is known to leverage Darknet-53[9].With no residual connections,

In addition, the framework supports many Deep learning architectures that cover different neural network models, such as recurrent and Convolutional neural networks, which enable the development of a wide range of applications beyond object detection. Increasing the number of Convolutional layers in neural networks is a common technique to enhance their ability to identify finer details or smaller objects in images. To achieve this, the network's depth may be increased, the number of filters in each layer can be increased, or skip connections can be added to make it easier for data to flow between levels. By making these changes, the network's ability to extract intricate information from pictures will be improved, leading to better object identification performance. To facilitate the later phases of object detection[9], the architecture is in charge of obtaining and extracting hierarchical features from the input image. A feature pyramid network (FPN) is incorporated into Darknet-53 in order to capture features at many scales.

This makes it possible for the model to identify different-sized items in the input image. In a single forward pass, Darknet-53 and detecting heads can anticipate bounding boxes, object classes, and confidence scores for many objects. YOLOv3[9] expands the architecture to increase detection accuracy and uses Darknet-53 as its backbone. There are grid divisions in the input image by YOLOv3[9], which then forecasts bounding boxes and class probabilities for each grid cell. Figure 1 shows the network architecture of Darknet-53. It is the most widely used model for real-time object detection. Darknet-53, the main network, is made up of 53 Convolutional layers. YOLOv3[9] also has a detecting head that adds 53 more layers, making it a fully Convolutional neural network with 106 layers overall. This architecture effectively processes the full image in a single forward network pass, enabling real-time object detection. *Modifying* Darknet layers allows for the creation of customized network architectures that can improve the performance of the

network on specific tasks or datasets. The flexibility and versatility of Darknet make it a powerful tool for developing deep learning applications.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig 1: Network Architecture of Darknet-53[9]

1.2 DensenNet for Network Training

The neural network architecture is known as DenseNet[25], which stands for Densely Connected Convolutional Networks. DenseNet, in contrast to conventional Convolutional neural networks (CNNs)[25], promotes dense connections among layers to increase feature reuse and parameter economy. Dense connections between layers are introduced by DenseNet. When a block is dense, every layer in the block gets input from every layer before it in addition to the previous layer. Reusing features, cutting down on parameters, and assisting in the mitigation of the vanishing gradient issue are all facilitated by this pattern of connection. DenseNet has demonstrated performance in several computer vision[7] applications, including object detection and image classification. Transition layers and dense blocks are included in the DenseNet[25] architectural framework[25]. There is a distinct pattern of connectivity within each dense block, with each Convolutional layer being closely connected to all the other layers. Making "shortcut" links between each layer's output and the next layer's input establishes this connectivity. Transition layers are included to aid in the efficient growth of the network by optimizing feature map sizes as they pass through dense blocks.

A wide range of computer vision[9] applications, including semantic segmentation, object recognition[9], and image classification, are covered by Dense Net's capabilities. DenseNet[25] has established itself as a pioneer in attaining state-of-the-art performance across these many applications thanks to its unique ability to effectively leverage on feature reuse while concurrently reducing parameter count. Through the deliberate integration of both DenseNet[25] and Darknet-53[9] into our model, we have increased both accuracy and efficiency. Darknet-53[9], with its 53 Convolutional layers, is a strong backbone known for its efficiency in object detection. Simultaneously, Dense Net's distinct connection enhances feature reuse, improving performance in tasks such as semantic segmentation and picture classification while maintaining optimal parameter efficiency.

1.3 Suspicious Human Activity Categories

As shown in Figure 1, it is interesting to see the variety of categories of suspicious human behavior that scholars are currently concentrating on. A dedication to comprehending, recognizing, and resolving any security issues and anomalies in diverse circumstances is demonstrated by the investigation of these categories. The scope of the research emphasizes the necessity of a thorough strategy for public safety, security, and surveillance. We have done the literature study from which we have put the activities under various categories.

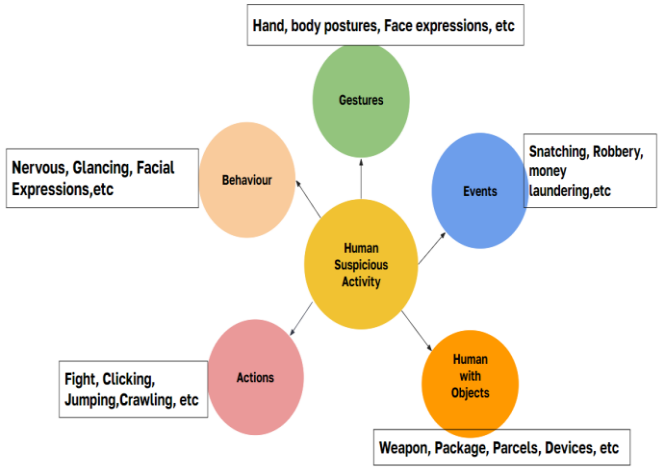

Fig 2: Categories of Suspicious Human Activities

As we have implemented a Prototype, We have taken humans with objects, in which we work on humans with weapons such as guns. We found it is difficult to locate multiple weapons in the video frame in a single frame and also found the research gap that works on small moving objects needs accuracy, we chose to implement a strong real time system to detect multiple weapons in a single frame as object detection. We have concentrated on identifying people in our prototype implementation who are carrying guns. Finding several firearms accurately in a single video frame is one significant obstacle we faced. We have identified a research gap based on our observations, especially in the domain of small moving objects, where obtaining high precision is a major difficulty.

2. Literature Review

Traditional machine-learning techniques have been used to video surveillance pistol recognition using ordinary RGB images. In their study, Tiwari and Verma [7] proposed an identification pipeline for automatic handgun detection. They suggested a pipeline that utilized Harris interest point detectors and Fast Retina Key point (FREAK) descriptors [7]. The objective of this method was to improve the precision and effectiveness of gun location within a specific environment. Not only can CNN-based classifiers produce outstanding results, but other deep-learning architectures have also been designed for gun detection. Sliding windows are used to extract tens of thousands of sections from the source image, and each region is then classified by a CNN according to the first approaches [8]. This family of methods' main flaw is its computational complexity, which makes real-time systems difficult to use. Processing several regions in the image necessitates a large amount of computer power, which could lead to delays and limits in real-time performance.

Olmos et al. (2018) investigated the sliding window and region method for weapon detection systems [11]. A quicker R-CNN trained on its dataset and using a Visual Geometry Group 16 CNN as the feature extraction foundation produced the best results. Without the use of pre-made candidate region suggestions, bounding box regression and object classification are performed using one-stage detectors such as YOLO [12], RetinaNet [13], or SSD [14]. Mobile devices are better suited for these architectures since they typically have faster inference times. Recent research [15,16] that is essentially simple implementations of YOLOv3 trained on handgun-specific datasets has employed YOLOv3 for handgun detection. Furthermore, by using a two-stage training approach based on both fake and real images, Salazar Gonzalez et al. [17] proposed a strategy that improves the accuracy of the Faster-RCNN in detecting firearms.

Indeed, the works do not provide any additional background; instead, they just rely on the object of interest's visual appearance, such as weapons. Several techniques also use temperature-based analysis [18]. The temperature is calculated in order to identify metal weaponry. DNN is used for both feature-based analysis classification [20] and DNN is used for object segmentation from a background in medical images [19]. Incorporating body position has been done in previous works using different ways.

The effects of altering the training datasets by incorporating body position overlays into the source images were examined by Salido et al. [21]. The findings demonstrated that the metrics for a number of detectors improved. Here, the location data is utilized to improve the input images during pre-processing; however, only visual components are employed in training. The combination architecture proposed by Velasco-Mata et al. [22] combines the output mask of a YOLOv3 base handgun detector with a body posture detection heatmap. As a result, the model generates grayscale images that suggest possible handgun positions. For each dataset, this method primarily uses a threshold for the grayscale images that are manually chosen to obtain the ideal trade-off between FPs and FNs. Moreover, Ruiz-Santaquiteria et al. [23] presented a technique for identifying handguns based on body position data. This method employs a Calculator for 2D body positions to extract bounding boxes for the hand area and posture critical points. Next, two distinct CNNs are built, one to process images of the body position and the other to process hand regions. After that, a detection judgment is produced by combining the output from the two branches[23]. This method is similar to the one we suggested, with the exception that it exclusively uses CNN-based architectures to extract location and visual data.

A deep learning method was used by Amrutha CV, Jyotsna et al. to detect suspicious activities in surveillance footage. The Visual Geometry Group, or VGG-16, has been used to identify students who are fighting or fainting in school as well as those who are using mobile phones while on campus[32]. This human activity category is based on humans with Actions. Multi-object identification and Tracking Using Machine Learning was implemented by Ayush Sahay et al. Using the COCO[34] and KITTI[34] datasets, they used YOLOV3 for multiple object identification and the CNN model for single object detection in the image[34].

A unique and reliable algorithm developed by Nuha and Hadeel [35] is intended to monitor and identify objects in real-time situations that are recorded by cameras. To strategically combine the advantages of principal component analysis (PCA) and deep learning networks, their suggested method creates an intelligent detection and tracking system. To maximize performance, this novel algorithm functions adaptively and smoothly and combines the advantages of PCA and deep learning.

Using artificial intelligence, M. Baranitharan et al.[37] Created automatic Human Detection in Surveillance Cameras to Avoid Theft behaviors in the ATM Center, which proved to detect suspicious behaviors in the ATM Center. To prevent suspicious actions, their suggested method effectively utilizes vector graphics and outlines an algorithm that takes use of backdrop modeling, subtraction, object recognition, and tracking.

Suspicious human activity recognition: a review, provided by Rajesh Kumar Tripathi et al. [38], examines six anomalous actions, including the detection of abandoned objects, theft, falls, accidents, unlawful parking on public property, violent activity, and fire. In the literature, they have detailed every step involved in identifying human activity from surveillance films, including feature extraction, activity analysis, recognition, and foreground object extraction. Object detection can be based on tracking or non-tracking algorithms.

In the Crime Prediction & Monitoring Framework Based on Spatial Analysis developed by Hitesh Kumar et al. [39], different machine learning algorithms and visualization approaches are used to forecast crime. To extract knowledge from datasets and analyze crime, they have tried to identify trends in crime as well as employ certain built-in machine learning techniques.

We have also done the literature study of Machine Learning built-in classifiers. Ismail H. et al[40] have done a study on, 'A Comparative Analysis of Machine Learning Classifiers for Twitter Sentiment Analysis study'. The study shows how the Naive Bayes classifier is better in sentiment mining of Twitter data rather than other classifiers. Laxmi Shanker Maurya et al[41]. They have proposed a few supervised machine learning classifiers that may be utilized to forecast a student's placement in the IT industry depending on their academic standing. A confusion matrix[41], heatmap[41], accuracy score[41], percentage accuracy score, and classification report are among the characteristics used to compare and evaluate the outcomes of various constructed classifiers. The parameters precision[41], recall[41], f1-score[41], and support make up the classification report that is produced by the constructed classifiers. To create the classifiers, the following techniques are used: Support Vector Machine[41], Gaussian Naive Bayes[41], K-Nearest Neighbor[41], Random Forest[41], Decision Tree[41], Stochastic Gradient Descent[41], Logistic Regression[41], and Neural Network.[41].

Sanjeev T et al[42]., Their research provides a fully-integrated analog neural network classifier architecture in order to decrease memory access, for low-resolution image categorization. In order to achieve high accuracy, they have used a hardware-software co-design process to develop custom activation functions utilizing single-stage common-source amplifiers. This information has been used for the training phase. The prototype IC manages a mean classification accuracy of 81.3%[42].

2.1 Research Gap Identified

Our Observations from the literature study are:

- The body of research to date emphasizes the need for an intelligent real-time system that can identify several objects in a single frame and is specifically tailored to identify suspicious human activity.
- There is a notable gap in research concerning real-time datasets, suggesting a need for more studies in this domain.
- Detection systems frequently have restrictions, mainly with regard to how well they can handle particular object types and sizes. Closing this gap will improve detection algorithms' adaptability
- Research on the identification of small moving objects in video sequences is clearly needed, suggesting a potential gap in the current methods
- For object detection, researchers mostly use well-known machine learning (ML) and deep learning models, such as CNN, RNN, faster RCNN, YOLO, etc. This dependence on pre-built models raises the possibility of creativity and the investigation of different approaches.
- Use of built-in ML classifiers such as Decision Tree, SVM, KNN, Naive Bayes, etc to test classification results.
- Although academics have employed customized object detection techniques, using pre-made datasets is still common, suggesting room for advancement and creativity.
- Only concentrating on classification frequently leads to imprecise estimation. Examining self-designed classifiers could produce outcomes that are more accurate and trustworthy.
- The potential for enhancing the accuracy of classification is evident, suggesting opportunities for refining existing methodologies and exploring novel approaches.

3. **Methods**

However, we acknowledge the flexibility to modify the dataset and redefine the category of suspicious activity. Alternatives mentioned in Figure 2 provide options for exploring other facets of suspicious behavior. Adapting the dataset to focus on specific categories outlined in Figure 2 could further refine the model's ability to identify and classify various types of suspicious activities beyond those involving weapons. This adaptability allows us to tailor our approach based on evolving requirements and insights gathered from ongoing research and practical considerations.

3.1 ODSAC Architecture:

We have implemented Object Detection and Suspicious Activity Classification (ODSAC) architecture by combining Darknet-53 with DenseNet Architecture. Our Proposed system creates video frames by processing and converting incoming images from a real-time video camera. We created our custom dataset with a range of gun kinds and sizes to train the model. Next, a feature extraction technique is applied to the input image to locate the target item using equations 1 and 2. To obtain F_ψ^i (k, c, d), the picture (a, b) is put through a feature extraction procedure.

$$F_{\varphi}(k_0,c,d) = {}^{1}/_{\sqrt{MN}} \sum_{x=0}^{M-1} \sum^{N-1} image(a,b)\varphi_{j=0}(a,b) \quad (1)$$

$$F^{i}_{\varphi}(k,a,b) = {}^{1}/_{\sqrt{MN}} \sum_{x=0}^{M-1} \blacksquare \sum_{y=0}^{N-1} image(a,b)\varphi^{i}j_{k,c,d}(a,b)\psi_{i} \quad (2)$$

Here, k_0 is scaling factor of $F\varphi(k_0,c,d)$ coefficients define an approx. of image(a,b) at scale $k_0 F^{i}_{\varphi}(k,c,d)$ parameters for the scales $k>=k_0$. Given the F_{φ} and F^{i}_{φ},image(a,b).

To increase the network's performance, we have added residual blocks of sizes 1x64 and 2x128 to the original Darknet-53[9] layers. The output of this step is sent to a Dense Architecture, in which each layer is interconnected with all the other layers in a dense block to facilitate feature reuse and help the network acquire more resilient and discriminating features. To enhance information flow throughout the network, the outputs from each layer are concatenated and transferred to the following dense block. ReLU activation functions and batch normalization improve the training process and guard against over-fitting. Finding and locating objects in an input image falls under the purview of the front detection layer (FDL). Typically, a Convolutional layer, it searches for regions in the input image that might contain the items of interest by examining them at various scales and locations. Following is the proposed algorithm for multiple object detection in a single frame

.

Algorithm: ODSAC

1. **Modified Darknet-53 Layers:**

- let $X \in R^{H \times W \times C}$ *be the input frame*
- Apply darknet_output =Darknet_53(X,residua l_blocks)

  with residual blocks of size 1×64 and 2×128 for

  improved performance.

2. **Dense Blocks:**

Initialize an empty list dense_blocks_output=[]

For each layer in the dense architecture:

Di=Concatenate(darknet_output,dense_blocks_output[0], …,dense_blocks_output[$i-1$])

Di=BatchNorm(ReLU(Di))

Di=DenseLayer(Di)

dense_blocks_output.*append*(Di)

Here, Di represents the output of the i-th layer in the dense architecture

3. **Concatenate and Pass to Next Dense Block:**

dense_block_input=Concatenate(dense_blocks_output)

darknet_output=DenseBlock(dense_block_input)

4. **FDL Layers:**

- FDL_output13=ConvolutionalLayer(darknet_output,size =13×13×21)
- FDL_output26=ConvolutionalLayer(darknet_output,size =26×26×21)
- FDL_output52=ConvolutionalLayer(darknet_output,size =52×52×21)

The front detection layer (FDL) uses predefined anchor boxes or prior boxes, which are small rectangles with predefined sizes and aspect ratios, to detect objects in the input image. For each anchor box, the FDL calculates scores representing the probability of an object being present in that box, and bounding box offsets indicate the difference between the anchor box and the actual object's bounding box. The architecture is optimized for detecting small objects, such as a gun in a video that appears small and far away from the camera. Up-sampling layers are used to increase the feature map's resolution and improve detection accuracy. This enables accurate detection of small moving objects in the video.

Front detection layer (FDL) Calculates:

1. **Anchor Box**: Let A_i represents an anchor box for i^{th} feature map.

$$A_i=(x_i, y_i, w_i, h_i) \quad (3)$$

where x_i and y_i are the center coordinates, w_i is the width, and h_i is the height of the anchor box.

2. **Confidence Scores**: The front detection layer calculates confidence scores Pi for each anchor box representing the probability of an object being present.

$$P_i=Probability\ of\ object\ presence\ in\ A_i \quad (4)$$

3. **Bounding box offsets:** predicted adjustments Δxi, Δyi, Δwi, Δhi indicate the difference between the predicted bounding box and the anchor box.

$$Predicted\ Bounding\ Box=A_i+(\Delta x_i, \Delta y_i, \Delta w_i, \Delta h_i) \quad (5)$$

4. The final prediction for Ai is given by combining the confidence score and the adjusted bounding box.

$$Final\ Prediction\ for\ A_i= (P_i A_i+(\Delta x_i,\Delta y_i,\Delta w_i,\Delta h_i)) \quad (6)$$

Our network is working in depth and we have modified the size of the residual block in the last layer to increase

the performance. A residual group consisting of different residual blocks, such as 1x, 2x, 4x, and 8x, is created after each Convolutional layer[9]. Before each residual group, stridden convolution with a stride of 2 is applied to downsample the spatial dimension of the feature maps. As a result, low-level characteristics were preserved and positional data important for object detection was encoded. Moreover, because each convolution contains parameters, the downsampling is not as completely non-parametric as max-pooling. It enhanced the capacity to identify smaller objects. The three residual groups' output is taken out and supplied into the detector as a feature vector at three different scales. The three feature vectors that result, assuming that the network's input is 416 x 416, are 52 x 52, 26 x 26, and 13 by 13[9]. These vectors are in charge of identifying tiny, medium, and large items, respectively. The model can learn more relevant semantic information from the upsampled later-layer feature maps and more fine-grained information from the earlier feature maps by upsampling and concatenating features of different sizes.

We have trained our ODSAC with 8914 gun images and 9951 helmet images of different types and sizes. This architecture is fed into the proposed algorithm given below. We have created a custom classifier that will categorize activity into 2 categories: suspicious and non-suspicious. If an activity is found to be suspicious an alert call and SMS will be sent to the authorized person with a crime location.

3.2 Custom Classifier for Suspicious Activity Classification

We have implemented a Custom classifier that relies on a set of rules and heuristics. The classification is based on the presence of an object in the captured frames. Following is the procedure for a custom classifier.

Procedure: Custom Classifier

Procedure custom_classifier (od: object detection, Ft: extracted frame at time t, ODSAC: Proposed Architecture, output R: class);

1. Begin camera_status←' initialized'

od ← 0

2. *Ft←frame(50ms)*
3. *Ft<-ODSAC*
4. *{ initialize rules };*

If object=helmet":

R<- Warning "Remove Helmet"

If object<- weapon detected :

od<- od+1

else :

od<-0

If od > 3:

R ← Suspicious Activity

5. *Do*

Ft←frame(50ms)

6. end

The algorithm begins by initializing the camera, which will capture the video stream to be analyzed. We are working on real-time videos. The od (object detection) variable is also initialized so the system extracts a frame from the camera. If a helmet is detected in the frame, the system issues a warning message asking the person wearing the helmet to remove it. The warning aims to ensure that the face of the person is visible to the camera, which can be important for identification purposes.

To Detect multiple objects(as of now our object is weapon(gun)): If a weapon is detected in the frame, the od variable is incremented by one. This indicates that a weapon has been detected in the current frame. If no weapon is detected in the frame, the od variable is reset to zero. This ensures that the od variable keeps track of consecutive frames in which an object has been detected. If the value of the od variable is greater than three, it indicates that a weapon has been detected in at least three consecutive frames. The custom classifies this activity as a suspicious human activity, and the system issues a warning message alerting the security personnel to investigate the situation. The system goes back to step 2 and repeats The process, capturing frames, detecting multiple guns or a person wearing a helmet in a single frame, and issuing warning messages as necessary. This process continues until the surveillance system is stopped.

3.3 Extracting Frames

The technique for extracting frames from a video involves reading each frame of the video file and saving it as an image file. Here is a step-by-step process to extract frames from a video:

Steps for Extracting Frames:

1. *Initialization:*

   Input:video_file_path

2. *Import necessary libraries, including OpenCV:*

   video=cv2.VideoCapture(video_file_path)

3. *Check if the video is opened successfully. If not, display an error message and exit.*
4. *Frame Processing Loop:*
5. *Initialize variables: frame_count=0*
6. *While video has more frames:*

   Read the next frame using success,frame=video.read()

7. *If success is True:*

   Increment frame_count frame_count by 1.

8. *Save the frame as an image file using:*

   cv2.imwrite(output_path_format.format(frame_count, frame)

9. *end while*
10. *Release the video file using the release() method of the video Object:*

    Release video video.release()

The camera that will record the visual stream for analysis is first initialized by the algorithm. Feature extraction is a method for identifying and removing elements from an image that can be used to present the image in a more simplified format. Here, four distinct methods for feature extraction have been applied:

Color-based features: Color histograms, color moments, or color correlograms can be retrieved to explain the distribution of colors in an image.

Texture-based features: Utilizing features with a texture, including gray-level co-occurrence matrices, local binary patterns, or Gabor filters, to extract texture information from an image.

Edge-based features: Using methods like Canny edge detection, Sobel Operator, etc., we utilize this to extract and characterize edges or boundaries in images or other visual data.

Shape-based features: Using geometric moments, corner points, and contours to extract information about an object's shape from an image.

3.4 Dataset

Our other contribution to this research is to create our own video and image dataset. To train and test our model we have used these videos. Following are some video screenshots from the video dataset. Figure 3 shows few samples of our video dataset.

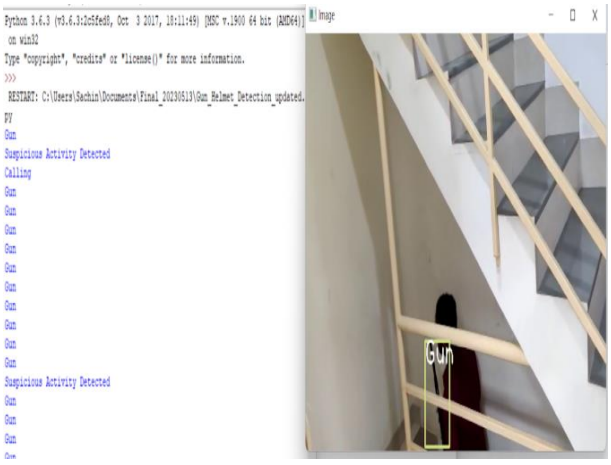

Fig 3: Screenshots from video dataset

To train the model for extracted frames from these videos, we have created a dataset. As this is a prototype we have used toy guns of different shapes, colors and sizes. A total of 18,865 images have been used for the training of ODSAC. Figure 4 shows few images from the image dataset.

Fig 4: Screenshots from Image Dataset

4. Results

4.1 Results of ODSAC

We have tested our Proposed method on some videos from our dataset and also tested it on real-time video from a live camera. Here are the screenshots of results:

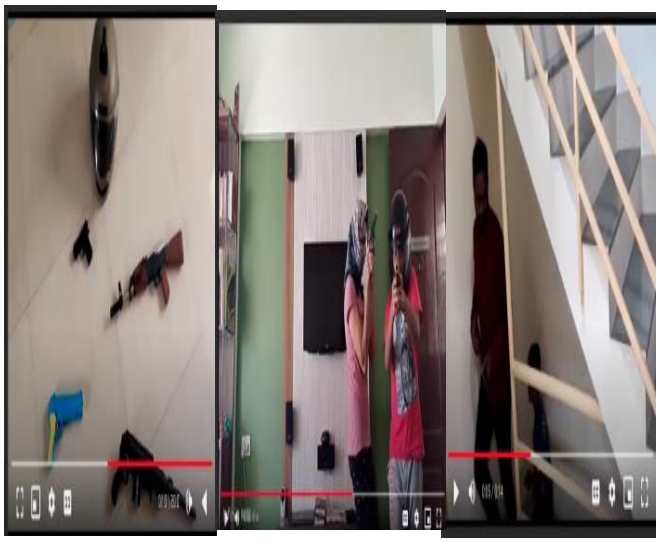

Fig 5: Result of ODSAC and Custom classifier along with output on Python console. Even though the camera is away. Our system is able to detect a moving object from video.

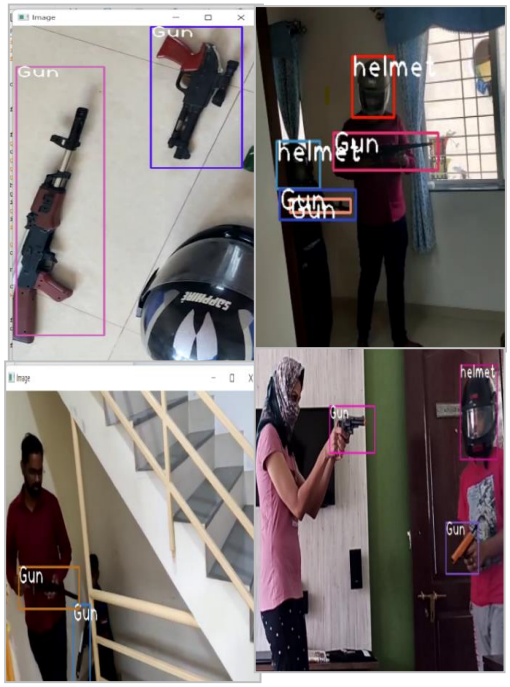

Fig 6: A few more sample results from ODSAC

Figures 5 and 6 show the end outcome of the proposed system architecture. It shows the results of multiple gun detection from real-time video in a Single Frame. Even though the background is black or blurred our ODSAC is able to detect multiple guns in a single frame. It is able to detect small, medium, and large moving objects in the video frame.

4.2 Performance of Custom Classifier

To measure the custom classifier's Performance parameters, We have used a confusion matrix. A confusion matrix is frequently used in the computation of performance characteristics including accuracy, precision, recall, and F1 score. Figure 7 summarizes the performance of a classification model. As our suspicious human activity category right now is humans with objects such as guns and helmets, we have tested the classifiers' performance for the same. Equation (5.1), Equation (5.2), Equation (5.3), and Equation (5.4) are used to compute accuracy, precision, recall, and F1 score, respectively, to determine the performance parameters [24]. Equation (11.3) is used to generate the confusion matrix.

$$Accuracy = \frac{TN+TP}{TN+FP+TP+FN} \quad (5.1)$$

$$Precsion = \frac{TP}{TP+FP} \quad (5.2)$$

$$Recall = \frac{TP}{TP+FN} \quad (5.3)$$

$$\text{F1 Score} = 2 * \frac{Precision * Recall}{Precision+Recall} \quad (5.4)$$

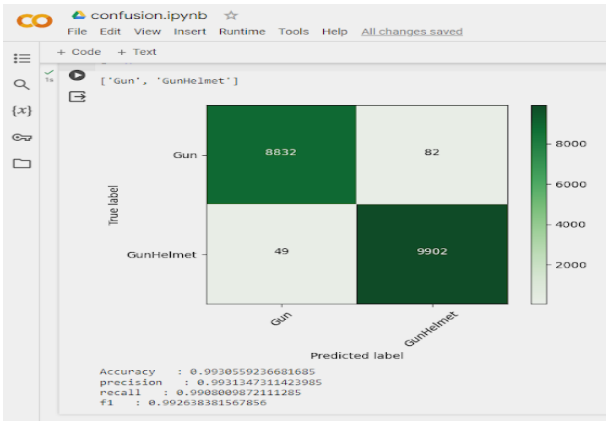

Fig 7: Prediction results for the identification and categorization of Suspicious Activity

Figure 7 shows the screenshot from our Google Colab Python output of the Confusion matrix with all the performance parameters mentioned above. In Table 1, the performance metrics are listed. Using ODSAC, the accuracy for detecting multiple objects(guns, helmet) is 99.31%.

Parameters	Accuracy	Precision	Recall	F1-score
Gun	99.30%	99.31%	99.08%	99.26%
Helmet	99.30%	99.31%	99.08%	99.26%

Table 1: Performance parameters for detection of objects using ODSAC

We conducted a comprehensive comparison between our proposed methodology's results and those of numerous other well-established techniques. Using a wide range of machine learning-based models for object detection and classification, we conducted a thorough evaluation of our proposed method. We specifically examined the performance of ODSAC Architecture with object detection using FasterRCNN, YOLO V3, VGG16, Extended CNN from our literature study shown in figure 8. Along with this we have also examined our custom classifier with Multi class SVM Classifier, Decision Tree Classifier, KNN Classifier and SVM Classifier shown in figure 9.

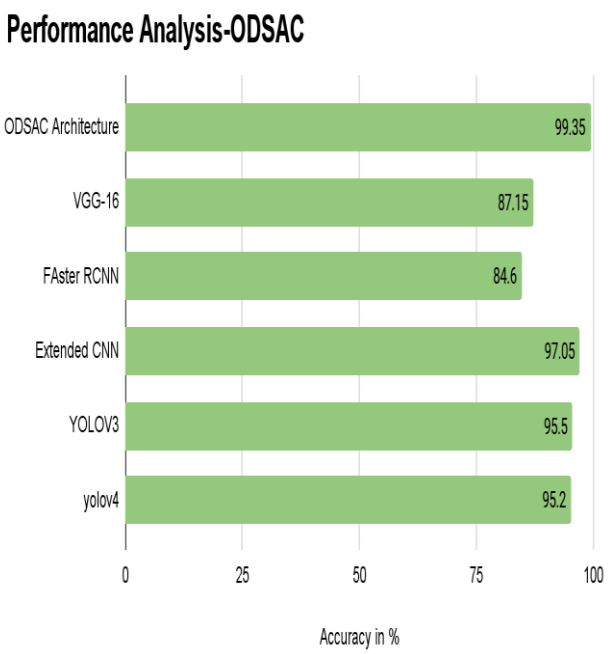

Fig 8: Comparative Analysis of ODSAC with some existing Methods

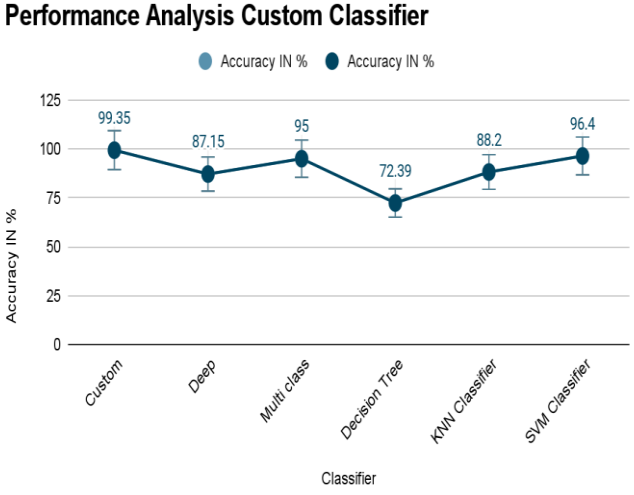

Fig 9: Comparative Analysis of Custom Classifier with some existing Methods

The proposed method is also able to overcome some limitations of machine learning built-in methods such as the detection of small objects from video, working on large data sets, and detection of objects correctly from blurred or noisy video. Our findings demonstrate significant improvements over the limitations encountered in existing Methodologies**.**

4.3 Implementation of Alert Call Module

We have created an Android application that, in the event that the activity classifier detects suspicious behavior, will send an alert call to a designated recipient. Figure 10 shows a screenshot of the Alert call Module.

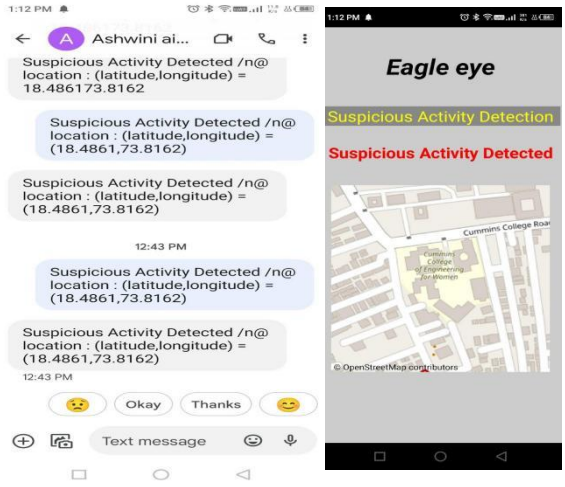

Fig 10 results of Alert call Module

After the system makes a crime prediction, the module is further equipped to send real-time crime location data. This feature comes in quite handy when it comes to being proactive about stopping illegal activity or quickly seeking assistance.

5. Discussions

This Research work is based on a self-implemented dataset which overcomes the research gap of missing use of own dataset instead of a ready dataset. The ODSAC methods is giving good results comparatively with existing methods. We can change the dataset to any suspicious human Activity for crime prediction. A deep learning method was used by Amrutha CV, Jyotsna et al. to detect suspicious activities in surveillance footage. The Visual Geometry Group, or VGG-16, has been used to identify students who are fighting or fainting in school as well as those who are using mobile phones while on campus[32]. Our model can detect all types of Activities with improved accuracy. The Modified Architecture after combining Darknet-53 and DenseNet gives improved performance. We have also overcome some gaps such as the detection of a single object, a specific object with a specific size, and limitations. Our ODSAC can detect multiple suspicious activities in a single frame. Also, as we have taken a prototype of Weapon detection here, results show that multiple weapons in a single frame have been detected and are of different shapes and sizes.

Our Primary focus was to overcome limitations such as limited sizes, Small Moving Objects in a video, and detection in a dark background.

Based on our findings, custom classifiers can give more accurate results. In our study, Laxmi Shanker Maurya et al[41]. They have proposed a few supervised machine learning classifiers that may be utilized to forecast a student's placement in the IT industry depending on their academic standing[41]. The author has used built-in machine learning classifiers for classification. Our findings are results giving 99.31 accuracy for crime prediction. This will help to prevent crime in public places, Analyze suspicious behavior of a person, etc. Our Comparative study shows that Custom classifiers are more accurate than some existing methods.

We have overcome the challenge of processing video in real time. The position of this novel technique is established by demonstrating its comparative advantages, stretching the limits of current studies in this area.

6. Conclusions and Future Work

Our study focused on overcoming the disadvantages of some existing methods along with improved efficiency and performance for their effectiveness in detecting suspicious activity and predicting crimes. ODSAC has proven to be quite accurate and has successfully addressed the shortcomings of the machine learning methods that are currently in use. The paradigm is beneficial for predicting crime in society and has significant advantages in the field of criminal justice to detect suspicious behavior patterns of criminals. It is also useful in the military to predict weapon attacks in healthcare to detect suspicious diseases in patients. The effectiveness of the algorithm may vary based on factors such as environmental conditions, camera quality, and specific use cases. In the future, we can implement this algorithm on IoT-based Devices such as Raspberry Pi to make a real-time wearable suspicious activity detection and crime prediction system. The proposed system can be used to detect concealed weapons by integrating with some sensors or Hardware.

Acknowledgments: We would like to thank all our study participants

Authors Contributions: Ashwini Bhugul-software, Conceptualization, Methodology, Documentation. Dr. Vijay Gulhane- Drafting, Methodology, Software**.**

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] F. Enrique, L.M. Soria, J.A. Álvarez-García, et al., Vision and crowdsensing technology for an optimal response in physical-security, Int. Conf. on Computational Science, Springer, 2019, pp. 15–26.

[2] T. Ainsworth, Buyer beware, Security Oz 19 (2002) 18–26.

[3] S.A. Velastin, et al., A motion-based image processing system for detecting potentially dangerous situations in underground railway stations, Transp. Res. Part C: Emerging Tech. 14 (2) (2006) 96–113.

[4] Everytown for Gun Safety, Gunfire on School Grounds in the United States, 2020.

[5] R.A. Tessler, S.J. Mooney, C.E. Witt, et al., Use of firearms in terrorist attacks: differences between the USA, Canada, Europe, Australia, and New Zealand, JAMA Intern. Med. 177 (12) (2017) 1865–1868.

[6] G. Flitton, T.P. Breckon, N. Megherbi, A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery, Pattern Recognition, 46 (9) (2013) 2420–2436.

[7] R.K. Tiwari, G.K. Verma, A computer vision based framework for visual gun detection using Harris interest point detector, Procedia Computer. Sci. 54 (2015) 703–712.

[8] F. Gelana, A. Yadav, Firearm Detection from Surveillance Cameras Using Image Processing and Machine Learning Techniques, in: Smart Innovations in Communication & Computation. Sci., Springer, 2019, pp. 25–34.

[9] Joseph Redmon, Ali Farhadi, 'YOLOv3: An Incremental Improvement University of Washington', Computer Vision and Pattern Recognition, Published in arXiv.org 8 April 2018.

[10] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, Adv. Neural. Inf. Process System. 28 (2015).R. Olmos, S. Tabik, F. Herrera, Automatic handgun detection alarm in videos using deep learning, Neurocomputing 275 (2018) 66–72.J. Redmon, A. Farhadi, Yolov3: an incremental improvement, Available online: arXiv preprint arXiv:1804.02767 (2018).

[11] T.-Y. Lin, P. Goyal, R. Girshick, et al., Focal loss for dense object detection, in: Proc. of the IEEE Int. CVPR, 2017, pp. 2980–2988.

[12] W. Liu, D. Anguelov, D. Erhan, et al., SSD: Single shot multibox detector, in: ECCV, Springer, 2016, pp. 21–37.

[13] Warsi, M. Abdullah, M.N. Husen, et al., Gun detection system using yolov3, in: 2019 ICSIMA, IEEE, 2019, pp. 1–4.

[14] R.F. de Azevedo Kanehisa, A. de Almeida, Firearm detection using Convolutional neural networks, in: ICAART (2), 2019, pp. 707–714.

[15] J.L. Salazar Gonzalez, other, Real-time gun detection in CCTV: an open problem, Neural Networks 132 (2020) 297–308.

[16] Adwait A. Borwankar, Ajay S. Ladkat, Manisha R. Mhetre. Thermal Transducers Analysis. National Conference on, Modeling, Optimization and Control, 4th – 6th March 2015, NCMOC – 2015.

[17] Ajay S. Ladkat, Sunil L. Bangare, Vishal Jagota, Sumaya Sanober, Shehab Mohamed Beram, Kantilal Rane, Bhupesh Kumar Singh, "Deep Neural Network-Based Novel Mathematical Model for 3D Brain Tumor Segmentation", Computational Intelligence and Neuroscience, vol. 2022, Article ID 4271711, 8 pages, 2022.

[18] M. Shobana, V. R. Balasraswathi, R. Radhika, Ahmed Kareem Oleiwi, Sushovan Chaudhury, et al, "Classification and Detection of Mesothelioma Cancer Using Feature Selection-Enabled Machine Learning Technique", BioMed Research International, vol. 2022, Article ID 9900668, 6 pages, 2022.

[19] J. Salido, V. Lomas, J. Ruiz-Santaquiteria, O. Deniz, Automatic handgun detection with deep learning in video surveillance images, Appl. Sci. 11 (13), 2021.

[20] A. Velasco-Mata, J. Ruiz-Santaquiteria, N. Vallez, O. Deniz, Using human pose information for handgun detection, Neural Comput. Appl. 33 (2021) 17273–17286.

[21] J. Ruiz-Santaquiteria, et al., Handgun detection using combined human pose and weapon appearance, IEEE Access 9 (2021) 123815–123826.

[22] A. S. Ladkat, S. S. Patankar, and J. V. Kulkarni, "Modified matched filter kernel for classification of hard exudate," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2016, pp. 1-6, Doi: 10.1109/INVENTIVE.2016.7830123.

[23] Gao Huang, Zhuang Liu, Laurens van der Maaten, 'Densely Connected Convolutional Networks', Computer Vision and Pattern Recognition, IEEE, 2017.

[24] Ms. Ashwini M. Bhugul, Dr. Vijay S. Gulhane, "Real Time Video Activity Detection Techniques in Machine Learning", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 23, Issue 6, Ser. I, PP 40-44, December 2021.

[25] Ms. A. M. Bhugul, Dr. V. S. Gulhane, "Novel Deep Neural Network for Suspicious Activity Detection and Classification", IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), February 2023.

[26] Ms. A. M. Bhugul, Dr. V. S. Gulhane, "Detection of Suspicious Human Activities and Prediction of Crime Using Machine Learning Approach", International Research Journal of Science Engineering And Technology, Vol - 10, Issue- 4, Page(s): 83 - 88 (2020).

[27] Yudhvir Rana, Title of subordinate document. In: Armed robbers loot Rs 22 lakh from Punjab National Bank bran, 2023)

[28] Sounak Mukhopadhyay, Title of subordinate document. In: "Mass shooting: 6 people shot dead in Mississippi, the US state with the weakest gun laws", Feb 2023.

[29] The Takeaways, Title of subordinate document. In: "Gun Violence in 2023: Nearly 40 Mass Shootings in 26 Days", 2023.

[30] Amrutha C. and Jyotsna, Amudha J, 'Deep Learning Approach for Suspicious Activity Detection from Surveillance Video', Proceedings of the Second International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2020), IEEE Xplore Part Number: CFP20K58-ART; ISBN: 978-1- 7281-4167-1.

[31] Neelam Dwivedi, Dushyant Kumar, Dharmender Singh Kushwaha, "Weapon Classification using Deep Convolutional Neural Network ", IEEE Conference on Information and Communication Technology (CICT) , IEEE , 2019,

[32] Harsh Jain; Aditya Vikram; Mohana; Ankit Kashyap; Ayush Jain, "Weapon Detection Using AI and Deep Learning for security Applications", IEEE Xplore, International conference on electronics and sustainable communication system, 2020, https://doi.org/10.1109/ICESC48915.2020.9155832

[33] Nuha H. Abdulghafoor, Hadeel N. Abdullah,' A novel real-time multiple objects detection and tracking framework for different challenges', Alexandria Engineering Journal, Volume 61, Issue 12, Pages 9637-9647, 2022.

[34] Sarita Chaudhary, Mohd Aamir Khan, Charul Bhatnagar, "Multiple Anomalous Activity Detection in Videos ", 6th International Conference on Smart Computing and Communications, ICSCC 2017, Procedia Computer Science 125 (2018) 336–345,2018.

[35] M. Baranitharan, R. Nagarajan, G. ChandraPraba ,'Automatic Human Detection in Surveillance Camera to Avoid Theft Activities in ATM Centre using Artificial Intelligence', International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181.

[36] Rajesh Kumar Tripathi1, Anand Singh Jalal1, Subhash Chand Agrawal, 'Suspicious human activity recognition: a review', February 2017, Artificial Intelligence Review 50(3), DOI: 10.1007/s10462-017-9545.

[37] Hitesh Kumar Reddy, Toppi Reddya , Bhavna Sainia, Ginika Mahajan, 'Crime Prediction & Monitoring Framework Based on Spatial Analysis', Elsevier, Procedia Computer Science, Volume 132, 2018, Pages 696-705.

[38] Heba M. Ismail, Saad Harous, Boumediene Belkhouche, 'A Comparative Analysis of Machine Learning Classifiers for Twitter Sentiment Analysis', IConference: 17th International Conference on Intelligent Text Processing and Computational Linguistics - CICLing, 2016.

[39] Laxmi Shanker Maurya, Md Shadab Hussain & Sarita Singh, 'Developing Classifiers through Machine Learning

Algorithms for Student Placement Prediction Based on Academic Performance', Applied Artificial Intelligence, Taylor and francis, Pages 403-420 , Mar 2021.

[40] Sanjeev Tannirkulam Chandrasekaran; Akshay Jayaraj; Vinay Elkoori Ghantala Karnam; Imon Banerjee; Arindam Sanyal, 'Fully Integrated Analog Machine Learning Classifier Using Custom Activation Function for Low Resolution Image Classification', IEEE Transactions on Circuits and Systems I: Regular Papers , Volume: 68, Issue: 3, March 2021.

[41] Savitha Acharya, Vaishnavi M. , Sujith Kumar , Shahid Raza, Halesh R, 'Smart Surveillance Robot for Weapon Detection', International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321-9653, IC Value: 45.98; Impact Factor: 7.177, Volume 7 Issue V, May 2019.

[42] Gaurav Raturi, Priya Rani, Sanjay Madan, Sonia Dosanjh , 'ADoCW: An Automated method for Detection of Concealed Weapon', 2019 Fifth International Conference on Image Information Processing (ICIIP), Shimla, India, Accession Number: 19342607, IEEE, 2019.

[43] Ramzan, M., Abid, A., Khan, H. U., Awan, S. M., Ismail, A., Ahmed, M., Mahmood, A. 'A Review on state-of-the-art Violence Detection Techniques', IEEE Access, 1–1, doi:10.1109/access.2019.2932114, 2019.

[44] Baser M, Mittal M, Samiya D. 'Real Time Foreground Segmentation for Video Sequences with Dynamic Background', IEEE 17th India Council International Conference (INDICON), 2020.

[45] Khawaja MoyeezUllah Ghori, Muhammad Imran, Asad Nawaz, Rabeeh Ayaz Abbasi, Ata Ullah & Laszlo Szathmary, 'Performance analysis of machine learning classifiers for non-technical loss detection', Journal of Ambient Intelligence and Humanized Computing, Springer Link, DOI: 10.1109/CSNT.2017.8418550 , 2020.

[46] Nandini. G, Dr. B. Mathivanan, Nantha Bala. R. S, Poornima. P, 'Suspicious human activity detection', International Journal of Advance Research and Development, 2018, Volume 3, Issue 4.

[47] R. Mahajan and D. Padha, 'Detection of concealed weapons using image processing techniques: A review', First International Conference on Secure Cyber Computing and Communication (ICSCCC), Dec 2018, pp. 375–378.

[48] 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), e-ISSN: 2395-0056, Volume: 04, Issue: 12 , Dec-2017.]

[49] S. Gupta, 'Introducing Custom Classifier — Build Your Own Text Classification Model without Any Training Data'.