Automatic weapon detection using Deep Learning

1st Poloju Akshaya
Computer Science and Engineering
B V Raju Institute of Technology
Narsapur, India
21211a05n8@bvrit.ac.in

2nd Pamulaparthy Bharath Reddy
Computer Science and Engineering
B V Raju Institute of Technology
Narsapur, India
21211a05l4@gmail.com

3rd Pratheek Panuganti
Computer Science and Engineering
B V Raju Institute of Technology
Narsapur, India
21211a05l8@bvrit.ac.in

4th P.Gurusai
Computer Science and Engineering
B V Raju Institute of Technology
Narsapur, India
21211a05l2@bvrit.ac.in

5th Abdus Subhahan
Computer Science and Engineering
B V Raju Institute of Technology
Narsapur, India
abdus777@gmail.com

2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE) | 979-8-3503-0570-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/RMKMATE59243.2023.10369889

***Abstract*—Safety and security is important in today's world scenario. For a country to be strong in terms of economy it must have secure and safe environment. CCTV surveillance is majorly used to monitor activities such as robbers and threats in public places, but these cameras need constant human observation. The fast and accurate automatic weapon detection system is useful to avoid these types of risks in public places such as schools, hospitals, museum and traffic etc. In many places the crimes are caused by pistols, guns and knives are very more, mainly in the countries and cities where there is no gun control laws. This work, focuses on automatic weapon detection in CCTV footage by making use of deep learning algorithms. In our work we are going to use YOLOv8 model to detect weapons as it is better in terms of accuracy and speed when compared to other models. It uses larger data set which consists of various images of weapons. Data set includes pictures from youtube robbery videos, CCTV footages, GitHub repositories and Internet movies firearms data base. Sliding window technique is used to extract features from objects and classify weapons. YOLOv8 has shown improvement in detection with greater speed and high accuracy.**

***Index Terms*—Weapon Detection, Deep Learning, Object detection, CCTV Surveillance, YOLOv8 algorithm, Sliding window technique, Alert generation**

I. INTRODUCTION

Today, closed-circuit television (CCTV) systems are crucial for preventing security risks, where early detection is of utmost importance. .Nowadays, all public and private places, including banks, museums, train stations, airports, and government offices, have their own CCTV surveillance systems. These forms of technology are great for helping security personnel control crowds and for post-event investigations. The major drawback of these kinds of surveillance systems is the need for continuous human operator who has to monitor everytime. As the number of places monitored by video surveillance cameras increases, attention eventually goes, making these systems ineffective. Related study in this field shows the importance of early identification of threats and dangers in order to reduce the damage that is caused. Examples of this type of risk situations involves firearms include handgun attacks and shootings, shootings on school property, and terrorist attacks. The widespread use of handy weapons during violent behaviour is an important component in increasing rates of crime around the world. The law and order system must be under control for the country develop and move forward. All of them need a secure and peaceful environment, whether we want to bring in the investors for investment in our country or produce revenue from the tourism business. In many regions of the world, the crime rate due to firearms is quite troublesome. It is primarily more in nations where it is acceptable to keep a gun.By going through several CCTV captured robbery videos on YouTube and other websites it is found that almost 95 percentage of these cases have pistol or revolver as the weapon. In the context of public safety and security, automatic weapon detection through surveillance cameras is a crucial application of computer vision technology. Systems that can quickly and effectively find weapons in public spaces are becoming more and more important due to the rising anxiety over gun violence and terrorism. One such system uses the deep learning model for CCTV monitoring to automatically detect weapons. Modern object detection models use a single neural network to detect and classify things in real time. Due to its high speed and accuracy, it has gained popularity and is an acceptable choice for many types of computer vision applications, like automatic weapon detection. The main idea behind the automatic weapon detection id using deep learning model to train the model on a large set of data images which are labelled as weapons with their names. The model learns to detect various types of weapons such as handguns, pistols, and shot guns, and classify them based on data set. The model is then deployed to CCTV surveillance cameras to identify weapons in real-time application. The process of weapon detection using deep learning models involves several steps to be followed. First, the surveillance footage is captured by cameras and fed into this model. The model processes the footage and detects any weapons that are present in the footage. If a weapon is detected, an alert is sent to the security authority or law enforcement agencies, so that they can take appropriate action. One of the main advantages of using this type of model for automatic weapon detection is its real-time processing performance. The model can process CCTV footage at a high speed, making it suitable for real-time surveillance applications. And the model can be worked on specific data sets to improve its accuracy for detecting specific types of weapons.

II. Literature Survey

Nowadays, all public and private places, including banks, museums, train stations, airports, and government institutions, have their own video surveillance systems. These devices are great for helping security personnel manage crowds and for post-event investigations. The main drawback of these kinds of systems is the need for continuous human operator monitoring. As the number of places monitored by video surveillance cameras increases, attention eventually becomes distracted, making these systems ineffective. Related research in this field shows the importance of early detection of security dangers or risks in order to reduce the damage that is caused.

[3] Closed-circuit television (CCTV) systems are essential CCTV surveillance systems are playing a vital role which are used to prevent threats. In our work, a new method is proposed to combine, in a single architecture, both appearance of weapon and human pose information. Pose key points should be identified in order To take regions of hand and produce binary pose images which is the input given to model inputs. After each input has been separately analysed across multiple sub networks, the final bounding box is generated.Instead of giving multiple suggestions for improvement, these systems apply a single deep neural network to the whole image. Fixed elements of the image are identified, and probabilities and bounding boxes are calculated for each. This method uses a visual heat map representation of both the posture and the weapon location and convolutional layers to generate a final gray scale image that detects probable handgun hot spots on the image.This technique creates a final gray scale image that detects likely handgun hot spots on the image using convolutional layers and a visual heat map representation of the posture and weapon position. Multiple sources are used to create the data set for detection, including YouTube videos, public weapon data sets, and even artificially generated images from video games.By using the open pose frame work, it is possible to define the location and pose of a human. There are two main branches to the pose combination method, which goes by the abbreviation HRC+P (Hand Region Classifier + Pose data). The hand region classifier (HRC), which is the first, is used. A different specialised sub network, however, is responsible for processing the pose image. False positives that might come up in different parts of the image can be eliminated.

[7] CCTV surveillance system needs human supervision. In today's world, an automatic weapon detection system is necessary. Although we use a number of techniques to get there, our main objective is to increase accuracy. You Only Look Once (YOLO) and the Efficient Net are the algorithms used in this article. Both high-level and low-level objects can be detected with YOLO. But its effectiveness is poor. A more accurate algorithm is the efficient net. A weapon hidden behind loose garments can also be found using an effective net algorithm. We will select data sets for the Efficient Net Algorithm from CCTV camera recordings. The data is pre-processed into frames and these frames are labelled. Applied to labelled frames is an algorithm. Although CNN is frequently used to identify objects, it has several limitations. The accuracy of the efficient net algorithm is good.

[1] There are many algorithms in deep learning for automatic weapon detection. We need to choose a proper model to implement our idea to reduce false results. This is an existing problem which already has a solution, but our idea is to improve the accuracy and speed in detecting the weapons automatically. The two algorithms used are here FASTER R-CNN, SINGLE SHOT DETECTOR(SSD). FASTER RCNN is one of the object detection algorithm. An image is given as an input to find out an object from the given image. First, region proposal network find out the areas where there is a possibility of object can be found. And labels areas as Foreground class and background class. Anchor boxes are set of predefined boxes used to detect objects. They can be of different sizes. Because we have to detect different kinds of objects. The next step is Calculate IOU(Intersection over union), in an image if the overlapped area is more than 50 percent then object will be detected. Else object will not be detected. The main aim of RPN is to predict anchor boxes with Foreground class. The second step is "Region of Interest". The data which is saved in ROI is the output of Region Proposal network. The output of RPN is anchor boxes where object can be captured. The next step is Classifier and Regressor. Classifier is used to find out if the object such as weapon is identified in the image or not. Regressor is used to draw a box with bounding if object is found. In terms of speed SSD is better choice for implementation. In terms of accuracy FASTER R-CNN is best choice for implementation.

[6] They have created and presented a real-time system for arms identification that outperforms previously suggested methods in terms of interference time and mean average precision score(mAP). For achieving great precision, the YOLOv4 algorithm is used. By employing this approach and pre-processing procedures, the accuracy compared to the existing model was enhanced. They also looked into the

model submission on various computing devices in order to increase throughput, decrease delay and improve security. For weapon detection in real-time CCTV surveillance movies, better performance will be obtained when the improved system is implemented on Jetson Nano.

[4] The object localization and feature extraction methods were merged into a one block in YOLOv3. Their YOLO (You Only Look Once) single-stage design provides extremely rapid inference time. The coordinates of bounding boxes and probabilities of class for these boxes are calculated by taking the complete image to consideration in single or one instance. The primary advantage of adopting YOLO is its incredible speed. It can process 45 frames per second and is extremely quick. In YOLO, the entire image is saved into a convolutional neural networks, which makes an output prediction in single pass.The implementation uses two distinct dataset kinds. One dataset had already labelled photos, whereas the other is a collection of images. The YOLOv3 algorithm is being trained on this dataset to classify weapons.

[2] A Single Shot Multi Box Detector was designed to identify guns in surveillance footage. It says to recognize a variety of weapons, including shotguns, automatic rifles, sniper rifles, and handguns. The intersection over union (IoU) value of 0.50 and 0.75 gives the system precision.Grayscale conversion from RGB is used for picture processing. The altered images were sent into the Single Shot Multi-Box Detector (SSD) and MobileNet low weight deep neural networks-based SSD MobileNet model. The system's neural network underwent two stages of training: 1.a weapons data set containing 1218 machine gun images which was extracted from the COCO data set. 2.The weapons detector was trained by using fine-tuning the pre-trained model.

[5] The YOLOv3 algorithm is used in the proposed system to detect weapons. Images of knives, heavy weapons, and handguns compose the data set. This data set was trained using the You Only Look Once version 3 algorithm for the purpose of verifying the presence of weapons. The device can detect weapons after training and issue an alarm to the authorities. The data sets images are divided into SS cells. A cell is in charge of detecting the presence of an object if it's center falls in that cell. Each cell predicts where the bounding boxes will be, along with a level of confidence and a probability that item is class conditioned on the presence of a weapon.

[8] Architecture of a convolutional neural network.There are two fully connected (FC) brains in MSD-CNN, and each of them is in charge of a distinct categorization task.There are two distinct branches; the first identifies the abnormal subclass pictures, like guns, while the second classifies the usual subclass images, such activities (like walking and cycling) and job (like office work and household chores). The MSD-CNN Methodology will be used to locate the weapon using photos. It is possible to build various MSD-CNN instances because the MSD-CNN network is comparatively simple. to implement threading and increase computing efficiency

[9] The faster RCNN algorithms and SSD-based convolution neural network (CNN) are used in this study to detect firearms. Learning algorithms are used in object detection to recognise instances of various classes of items. This emphasises accurate gun detection. From the input video, frames are extracted from it. Before the detection of the object, a bounding box is built using the frame differentiating algorithmic program. The method takes into account machine learning models Region Convolutional Neural Network (RCNN) and single shot detection (SSD) With 0.736 s/frame, the SSD formula offers a faster speed. While RCNN is faster, its 1.606s/frame speed is still subpar compared to SSD algorithm. Faster RCNN has a greater accuracy of 84.6 percent in terms of precision. SSD offers associate degree accuracy of 73.8 percent in terms of contrast.

[10] The model's training time and calculating complexity are decreased via the HIPSO method.To detect weapons successfully, the HIPSO-SVM model is used. The SVM classifier's computational complexity and training time can be reduced by the HIPSO algorithm's selection of discriminative feature vectors. The suggested HIPSO-SVM model includes two vital steps, including the classification of weapons and non-weapons and weapon detection. The feature extraction is carried out using AlexNet, ResNet 18, and SIFT models for obtaining feature vectors from the segmented images after the weapon was identified in the video sequences using the GMM technique. The overall description states that the surveillance footage is captured by cameras and fed into model. The model process the footage and detects if any weapon present in the footage. If the weapon was detected in CCTV, an alert message is sent to the security authorities or law enforcement agencies. Many deep learning models are used to detect weapon but every model has its own advantages and disadvantages.
In conclusion, there are several object detection algorithms that have been used for weapon detection, including YOLOv4, Faster R-CNN, and SSD. Each algorithm has its own strengths and weaknesses, and the choice of algorithm depends on the specific application and environment. False positives and false negatives are always a concern when using object detection algorithms, and it's important to use a model in conjunction with other security measures, such as physical security barriers or human security personnel.

III. Problem Statement

- Automatic weapon detection using Deep learning.
- Current methodologies uses various deep learning models to detect weapons, but the accuracy and speed is less. so, here we are using YOLOv8(You only look once version) model.
- This proposed system will be more accurate than previously existing system and can be used for vast number of applications.

IV. Objectives

1) The main objective is to enhance security and safety measures by detecting weapons in public places like airports , museums , malls, traffic signals and schools to prevent threats and violent incidents.
2) An Automatic weapon detection system to detect weapons using YOLOv8 model in real-time applications.
3) To reduce false positives and negatives in detecting the weapon and improve accuracy of the system.

V. Existing Work

Automatic weapon detection through CCTV surveillance is an area of ongoing research and development with a focus on improving detection accuracy, speed and real time performance. Many deep learning based approaches and techniques such as convolutional neural network(CNN's), YOLOv3, Efficient-net and SSD are used to detect weapons in CCTV footage. These approaches use larger data sets containing of various pre-labelled images of weapons. These data sets are used to train the model to detect different types of weapons. One of the major disadvantage of CNNs is that they are computationally expensive, especially when used for high-resolution images. CNNs may struggle with detecting weapons in low-light or heavily occluded environments.One of the limitations of YOLO is that it may struggle with detecting small or distant objects, which can be a problem for weapon detection in surveillance videos. YOLO may produce false positives in complex scenes, such as those with overlapping objects or occlusions. Faster R-CNN model can be slow compared to other object detection algorithms, especially when used for high-resolution images or videos. Efficient net algorithm is efficient in terms of accuracy but the speed is less compared to other deep learning models. Single shot multibox detection(SSD) is slow to detect smaller objectives and it may produce false positives. The speed of SSD is more but efficiency is less when compared to others. The Deep learning model used in weapon detection should be efficient in term of both speed and accuracy

VI. Proposed Work

Automatic weapon detection through CCTV surveillance is an area of ongoing research and development with a focus on improving detection accuracy, speed and real-time performance. Many researchers have proposed various techniques for detecting weapons ranging from deep learning based approaches to feature based approach. Many deep learning based approaches and techniques such as convolutional neural network(CNN's), YOLOv3, Efficient-net and SSD are used to detect weapons in CCTV footage. These approaches uses large data sets to train the model to detect different types of weapons.

YOLOv8 is one of the recently released popular object detection model used for object detection. It uses fully convolutional neural networks which helps to detect weapons more accurately and it is a pretrained model. Dataset for this model this done by collecting various images from internet and youtube videos and labelled using bounding boxes. And the model is trained using the images and videos then this trained system is integrated with automatic weapon detection system. when the system is deployed, it analyzes input images and videos in real time using YOLOv8 model.when weapon is detected it sends an alert to security personnels. YOLOv8 shows promising results due to its high speed and real time performance.

VII. Methods

A. Data Collection

We have collected a data set from Roboflow named as Weapon detection.It consists various images divided into 3 sets or sections as train , valid and test. The train set consists of 80 percent images, valid set has 15 percent images and test set has 5 percent images. These images are labelled using bounding boxes. This data set has various images of Knife , gun, sword, revolver, rifle,ax,shotgun etc. The images of weapons in this data set are from different angles and of different sizes which helps the model to detect efficiently. Fig. 1. is the image of dataset consisting of various weapon images.

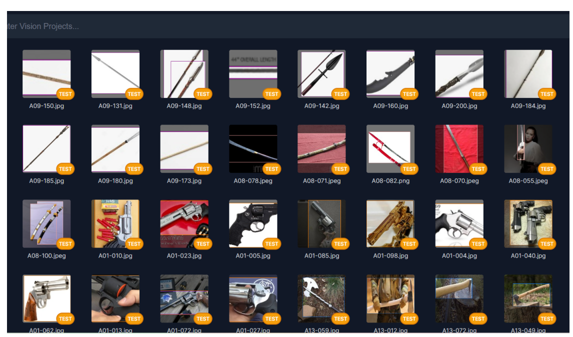

Fig. 1. Weapon images in dataset.

B. Training the model

The model is trained using the above data set for automatic weapon detection. The algorithm YOLOv8 is a pretrained model on coco data set. But it may no detect weapons and classify them efficiently without training the model with data set.Training the data with thousands of images increases accuracy. But it should not be over-fitted.The model takes more time for training the model. It is one of the longest process in detection.

C. Input data and pre-processing

The video from the CCTV surveillance is taken as the input for weapon detection.The input video is converted into frames and frames are converted into pixels. Frame Conversion involves some different stages. They are Extraction of frame where input video is divided into individual frames and Resizing and Adjusting the frames so algorithm can work properly. It means adjusting the frames such that weapon is only focussed more. The YOLOv8 algorithm uses single convolutional neural network and sliding window technique for pre-processing the data. This will reduce the noise, brightness and making it suitable for model to work it efficiently. Pre-processing gives an assurance that the given input is in the correct format, size so that weapons can be accurately detected and performance can be improved. This module also has some components such as resizing of objects to make them same size per required dimensions of algorithm, normalization refers to the process of adjusting or data to a common scale or range and color space converting for converting one color to other.

D. Weapon detection

We train the model using the collected data set by importing modules called Ultralytics and basic code from GitHub, after the training the model stores the results and ready to detect the weapons of any kind from input video.Firstly, the input image of different sizes and pixels is given to the backbone of YOLOv8 ,feature extraction is done there inside the convolution layers. The output of this backbone is sent to the neck where the concatenation of the all the features extracted undergoes followed by the head where the detection of the weapon is done. Fig. 2. describes that the weapon is detected. It is bounded using bounding boxes and confidence score.

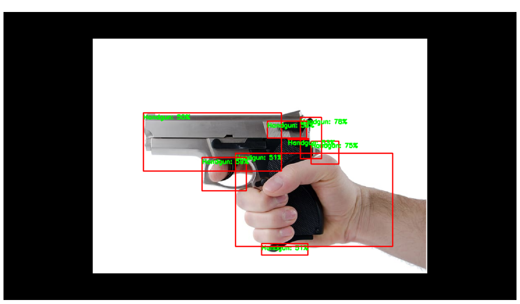

Fig. 2. Detected short gun.

E. Weapon classification

Features are extracted from the detected weapons and very important features are selected from them . These are used to classify the images. The selected features are compared with the trained images , as the trained images are priorly labelled as gun, shot gun, knife, etc the will classify

Fig. 3. Detected pistol.

the images. Fig. 3. is the image where the weapon is detected in the image and is classified as pistol.

F. Alert generation

Once the Weapon is detected and classified the system sends an alert message to security authorities regarding weapon detection and name of the weapon to take necessary action. Fig. 4. describes the architecture diagram of our automatic weapon detection model . It consists three parts. Algorithm training, Weapon detection and alert generation.

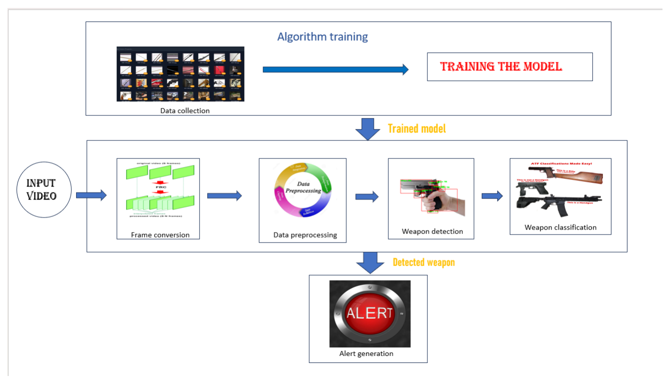

Fig. 4. Architecture diagram.

VIII. Result

We have detected various types of weapons in real time by CCTV surveillance by using You only look once version 8 algorithm. The performance measure of the YOLOv8 algorithm is measured in terms of different metrics like F1-score, Recall, precision and frames per second. F1 score is calculated by the ratio of precision to recall functions. Precision is the ratio of True positives and sum of false positives and true positives. Recall is the ratio given by True positives and sum of false negatives and true positives. Description about dataset: This dataset is taken from the roboflow, it is annotated data. we performed training and detection with this datasets and achieved good results. Comparision and Analysis: When compared to previous models such as SSD, faster RCNN , YOLOv3 ,this YOLOv8 algorithm has good accuracy and results.We have also minimised false positives and false negatives by providing larger dataset for training purpose. Single Pass Detection: It is the excellent features of

YOLOv8 algorithm which is present in YOLO algorithms. It can perform object detection in a single pass which analyses the whole image at once. So the weapons in the frames can be detected quickly. Analysing the image in single pass has reduced the time required for processing of object detection by reducing complexity. YOLOv8 can show promising results as it is leading and latest algorithm of object detection and it is a pretrained model which is trained over 80 different types of objects using coco dataset. Comparison of detected accuracy with various other algorithms: The model is better able to accurately identify various types of weapons even in complex situations, by improving performance.

IX. Discussions

We have detected weapon in the images by uploading images from google. Our model detected various types of images as our data set has various kinds of images of weapons.Detection using YOLOv8 algorithm has given good results compared to other existing model. It is a pre-trained model which is trained with nearly 80 different types of objects. In our model we are detecting only weapons so, we have gave a data set consisting of weapon images to train our model. And the results we got in our model are given below. Fig. 5. and Fig. 6. are the images of detection of various kinds of weapons. These are detected in the input image and a bounding boxes are bounded around the image in the input for identification purpose.

Fig. 5. Detection of weapon.

The below are the graphs Fig. 7. obtaine the model. The graphs obtained are train/boxloss g train/clsloss graph ,train/dflloss graph, metrics/precisi graph , metrics/recall(B) graph, val/boxloss graph, val/c graph, va/dflloss graph, metrics/mAP50(B) graph metrics/mAP50-95(B) graph.

The below figure Fig. 8. is the confusion r obtained by detecting weapons by our model. This summa the results and performance of the model on the given data set.This can also be used to calculate different parameters of the model.

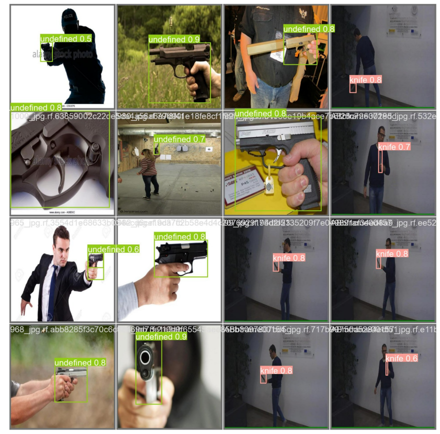

Fig. 6. Detection of various kinds of weapon.

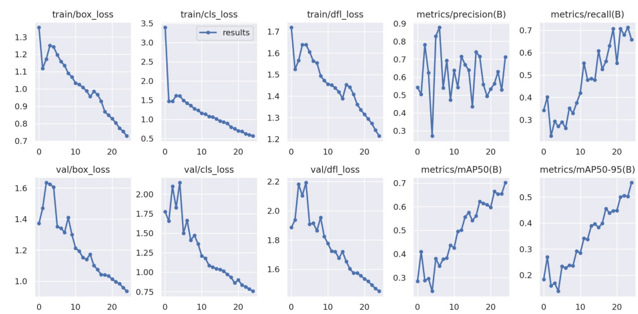

Fig. 7. Graphs from the model.

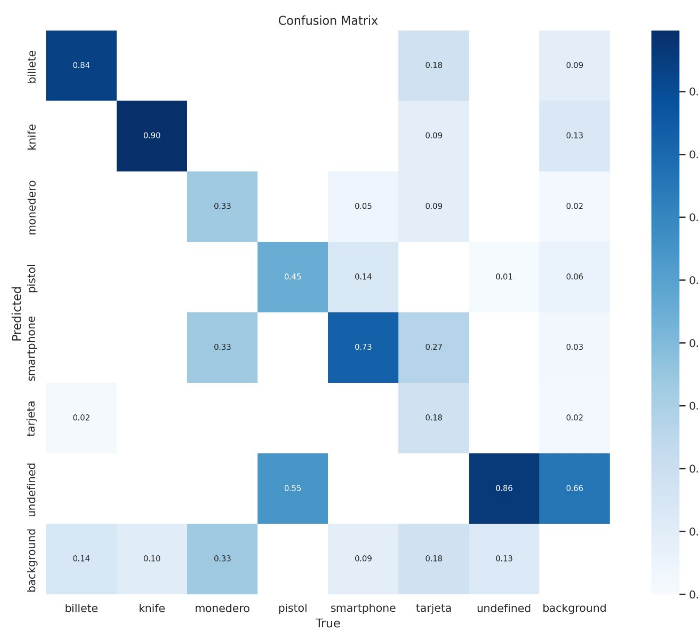

Fig. 8. Confusion Matrix.

X. CONCLUSIONS

Public spaces, including schools, colleges, hospitals, and museums, must be equipped with automatic weapon detecting devices to stop crimes before they happen. Creating a reliable system that produces few false positives and false negatives is a challenging task. In order to create an extremely fine weapon identification model in real time, we used deep learning to CCTV surveillance videos. This will have a beneficial effect on the global economy by luring travellers and investors to make investments in nations with strong security measures. One of the most recent models, the YOLOv8 algorithm, has been used by us to achieve excellent accuracy and efficiency.In this study, we used a weapon detection data set to train our model. This model can both identify weapons on CCTV and notify the administrators in person. One important aspect of weapon detection that must be achieved is speed. The CNN, R-CNN, and faster CNN algorithms are slower than the YOLOv8 algorithm used in the automatic weapon detection system. For surveillance systems to locate an object promptly and notify the authorities, object detection speed and accuracy are crucial factors. In comparison to the earlier models, our model is able to deliver the results more quickly. If new algorithms and methodologies emerge in the future, even greater advancements could be realised.

XI. ACKNOWLEDGEMENT

The dataset was collected from roboflow which has annotated data. Datasets were made accessible by roboflow experts. We have used this data set for our study and achieved good results.

REFERENCES

[1] Harsh Jain, A. Vikram, Mohana, A. Kashyap and A. Jain, ”Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications,” 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Research gate , 2020

[2] Shenghao Xu , Kevin Hung . “Development of an AI-based System for Automatic Detection and Recognition of Weapons in Surveillance Videos.” IEEE xplore , june 06,2020.

[3] Jesus ruiz-santaquiteria, alberto Velascomata,Noelia vallez,Gloriabueno Oscar. deniz,JoanA.Alvarez-garcia,“Handgun Detection Using Combined Human Pose and Weapon Appearance.” IEEE access , September 3, 2021.

[4] Dr. N. Geetha, Akash Kumar. K. S, Akshita. B. P, Arjun. M . “ Weapon Detection in Surveillance System.” IRJET , Vol. 10 Issue 05, May-2021.

[5] Sanam Narejo, Bishwajeet pandey, M.Rizwan Anjum, Doris Esenarro vargas, ciro Rodriguez,“Weapon Detection Using YOLO V3 for Smart Surveillance System.” Hindawi ,12 May 2021.

[6] Soban Ahmed, Muhammad tahir Bhatti,Muhammad gufrankhan, Bennylovstrom,Muhammad shahid. “Development and Optimization of Deep Learning Models for Weapon Detection in Surveillance Videos.” , Applied sciences MDPI , 7 June 2022.

[7] Erssa Arif, Syed Khuram Shahzad , Muhammad Waseem Iqbal, Muhammad Arfan Jaffar,Abdullah S. Alshahrani ,Ahmed Alghamd , “Automatic Detection of Weapons in Surveillance Cameras UsingEfficientNet.” Tech Science press , 08 March 2022.

[8] Palash Yuvraj Ingle , Young gab Kim. “Real-Time Abnormal Object Detection for Video Surveillance in Smart Cities.” , Sensors MDPI , 19 May 2022.

[9] P.Shanmuga Priya, Gurram yugandar reddy, Jammalamadaka mahendra reddy. “Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications.” IRJET , Vol: 09 Issue: 03,Mar 2022.

[10] Kiran kalla, Gogulamanda Jaya suma, “Weapon Detection in Surveillance Videos Using Human Inspired Particle Swarm Optimization Algorithm and Support Vector Machine.” , INASS , February 8 ,2022.