Intelligence Surveillance System for Bank Security against Robbery

Aniket Kamble[1], Lalit Deore[1], Neha Bangar[1], Usha Verma[2] and Lakhan Dev Sharma[3]

[1,2]E&TC Engineering, MIT Academy of Engineering, Pune, India
[3]VIT AP University, AP, India
[2]uyverma@mitaoe.ac.in

Abstract. Bank robbery is a criminal act involving the theft of money and valuable items from a bank, often accompanied by the use of force, violence, and threats. This poses a significant problem for banks. To address this issue, the proposed work aims to design an Intelligence Surveillance System for Bank Security against robbery. The system automatically detects suspicious activities, such as the presence of guns, using the YOLO architecture of Convolutional NeuralNetworks (CNN). The methodology involves training the CNN model on a comprehensive dataset of suspicious activities in banks and fine-tuning the model using transfer learning techniques. Once the system detects such suspicious activity, the intelligent surveillance system promptly triggers an automated alert system. The system notifies the nearby police station by generating an immediate SMS, alerting them to the ongoing robbery. Furthermore, the system captures and transmits relevant images of the robbery to provide visual evidence to law enforcement. The performance of the intelligent surveillance system is evaluated based on key metrics such as accuracy, precision, recall, and F1 score. The algorithm is tested on different training testing ratios such as 90-10, 80-20, 70-30, and 60-40. The highest accuracy of 91 %, precision of 95 %, recall of 98 %, and F1-score of 95% are achieved with an 80:20 ratios. The resulting system aims to provide an effective and reliable solution for enhancing bank safetyand reducing the impact of bank robberies on both financial institutions and public safety.

Keywords: Intelligent Surveillance System, Convolutional Neural Networks, Bank Security Enhancement

1 Introduction

Global security concerns have taken center stage in the contemporary landscape, driven by the compelling need to safeguard a spectrum of valuable and sensitive assets. These encompass not only individuals, residences, and communities but extend to the security of entire nations. Amid these complex security imperatives, financial institutions, epitomized by banks, standas pillars of economic stability. However, their vulnerability to threats like robbery presents a compelling case for the integration of advanced security measures. Traditional security methods, though partially effective, often lack the real-time monitoring and adaptive response capabilities required to counter modern criminal strategies. Nearly Rs 180 crore was lost in 2,632 cases of robbery, theft, dacoity, and burglaries at India's 51 banks in the last three years. In 2021, the capital union territory of Delhi had the highest rate of robbery,with over 11 reported cases per 100,000 people. There were over 29 thousand robbery cases in the country that year [1].

Human behavior detection within video surveillance systems presents an intelligent and automated approach to identifying suspicious activities. Convolutional NeuralNetworks (CNN) based algorithms are developed to automatically detect human behavior in public spaces such as malls, airports, railway stations, banks, offices, and examination halls [26]. The integration of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL), enables computers to emulate human cognition and learn from training data to predict activities [25]. The current landscape benefits from robust Graphics Processing Unit (GPU) processors and vast datasets, propelling Convolutional Neural Networks to the forefront. This technology facilitates the automatic extraction of intricate patterns from data, especially in the emerging domain of video surveillance applications. By harnessing AI, ML, and DL, computers adeptly mimic human behavior detection and prediction, bolstered by the adoption of Convolutional Neural network approaches catalyzed by the availability of GPUs and expansive datasets, consequently augmenting the effectiveness and precision of automated surveillance systems [2], [5].

The integration of computer vision with video surveillance offers significant potential to enhance public safety and security. Computer vision encompasses vital stages from environment modeling to behavior analysis, involving preprocessing for feature extraction. This approach utilizes both supervised and unsupervised classification techniques, with support from the prowess of Convolutional Neural Networks (CNNs) for direct visual pattern capture. The proposed system utilizes CCTV camera footage to oversee human behavior on bank premises, promptly alerting to suspicious incidents, with event detection and behavioral pattern recognition as key components. The system leverages CNN algorithms for precise navigation of video footage of bank premises, amplifying event detection and behaviorrecognition for enhanced safety and security.

2 Related Work:

Automatic detection of suspicious activities is very important from a security point of view. Many researchers are working in this field. Mehmet Tevfik Ağdaş et. al presented an approach that harnesses transfer learning techniques to enhance the accuracy of gun and knife image classification [1]. This method capitalizes on pre-trained neural network models to effectively adapt to the unique visual characteristics of weapons, demonstrating its potential to address challenges posed by limited data availability and intricate image features in the domain of weapon classification. Abdullah Alshammari et. al contributed to the field with their research presenting an integrated approach tailored for enhancing urban security througha multi-camera surveillance system [2]. In the work, Erssa Ayed et. al focused on the development of a specialized model for identifying firearms within surveillance footage [3]. Juan Du's exploration highlighted the significance of object detection methodologies based on CNN architectures - Yolo [4].

Shraddha Dubey presented a study underscoring the importance of harnessing deep learning techniquesfor creating effective gun detection systems [5]. Mai Kamal El den Mohamed et. al delved into the realm of automatic gun detection for video surveillance [6]. Cherine Fathy et. al research showcases a multidimensional integration of cutting-edge technologies to enhance weapon detection capabilities [7]. Maryam Qasim Gandapur et. al worked with an integrated deep-learning model designed to detect and prevent criminal behaviors within surveillance footage [8]. The study by Ayush Goyal et. al proposed border security through the application of machine learning techniques within remote video surveillance [9].

Shehzad Khalid et. al research contributes to enhancing security measures by creating a specialized system to identify weapons in surveillance settings[10]. Mrunal Malekar's work underscores the significance of deep learning techniques in deciphering criminal activities from surveillance footage [11]. VP Manikandan et. al study adds sophistication to existing techniques by combining neural networks and customized tuning to enhancegun detection accuracy [12]. Natan Santos Moura et. al focused on the detection of Weapon Possession and Fire in Public Safety Surveillance Cameras [13]. Mahadevan Narayanan et. al contributed with their research on advancing real-time detection of threats within public environments using advanced computer vision techniques [14]. Dammalapati Neelima et. al proposed an Intelligent Suspicious Activity Detection Framework (ISADF) for Video Surveillance Systems to enhance security measures [15].

The work has been done using ML, DL, and AI for the surveillance of different locations to detect suspicious activities. In the proposed work, the research embarks on the mission to address the security challenges banks face by developing an intelligent surveillance system. The key objective is to create a system that can proactively identify and prevent potential robbery attempts by indicating the potential danger, like the presence of an individual openly carrying a firearm in public, and serve as critical warning signs.

3 Methodology

The proposed work involves utilizing CCTV camera footage to oversee the bank robbery, promptly notifying relevant authorities of any questionable incidents through automated alerts, and enhancing bank safety and security.

3.1 System Architecture

The proposed bank robbery detection system comprises multiple components that collectively work to detect potential threats within a given environment. The architecture of the system is shown in figure 1. The system uses computer vision techniques for object detection, with an emphasis on identifying weapons such as guns, rifles, and knives. The primary modules of the system include Video capture, pre-processing, Object detection and alert generation.

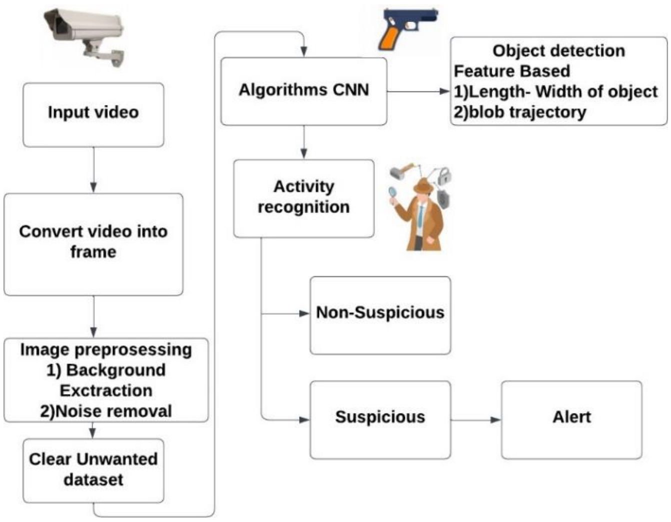


Fig. 1. Proposed System Architecture for bank security

3.2 Video capture

The initial phase of a video surveillance system involves the capture of video content. This is typicallyachieved through the installation of Closed-Circuit Television (CCTV) cameras strategically positioned to monitor the target area. These cameras capture diverse types of video footage, providing comprehensive coverage of the surveillance area. In the proposed context of implementation, video processing is performed at the frame level, necessitating the conversion of video content into individual frames. Consequently, the captured videos are transformed into a series of frames to facilitate subsequent processing.

3.3 Preprocessing

Background extraction is a vital process in computer vision, isolating subjects by removing static backgrounds. The code snippet uses OpenCV to create a background subtractor with cv2.bgsegm.createBackgroundSubtractorMOG(), employing the Mixture of Gaussians algorithm for dynamic background modeling. Additionally, the snippet hints at noise reduction through a Gaussian filter, likely implemented with functions like cv2.GaussianBlur(). This step enhances background-subtracted image quality by smoothing and reducing noise, contributing to more precise object segmentation.

3.4 Object Detection

Object detection, a vital element of computer vision, has greatly advanced with deep learning, with YOLO serving as a pre-trained object detector utilizing a Convolutional Neural Network (CNN) model. The architecture of YOLO is shown in Figure 2 [23]. In CNNs, learnable weights and biases are employed to discern objects by extracting high-level features, like edges, through convolutional layers. These layers use a *k* x *k* filter, known as a kernel, to produce activation maps that represent the detected features. In contrast to traditional methods involving hand-engineered filters, CNNs learn these filters through iterations, reducing the need for extensive preprocessing. The Max-Pooling layer further reduces spatial dimensions, thereby lowering computational requirements. Rectified Linear Unit (ReLU) activation defined by eq. (1), eliminates undesirable values, and fully connected layers transform data into a 1-dimensional array, enabling the network to learn nonlinear combinations of high-level features through backpropagation in a feedforward neural network.

$$\text{ReLU} : f(x) = \max(0, x) \tag{1}$$

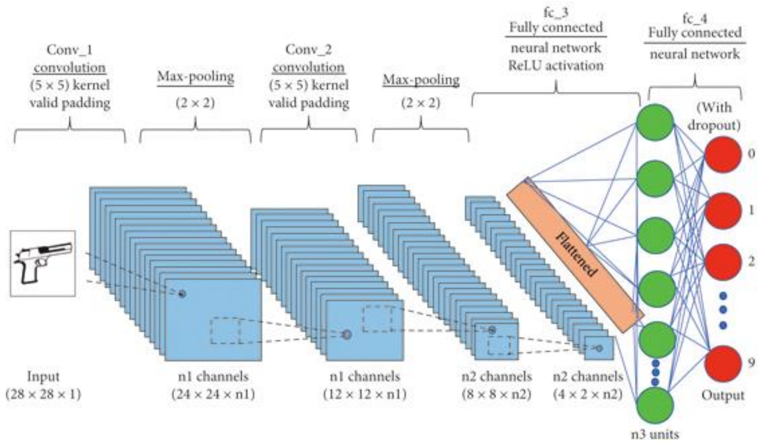


Fig. 2. YOLO architecture

3.5 Alert Generation

When a potential threat is detected, the system initiates an alert mechanism to promptly notify security personnel. The alert mechanism involves the following steps:

Frame Saving. To assist in further analysis, the system saves the frame containing the detected object. The captured frame is stored in the “saved frame” directoryas a JPEG image.

Alert Notification. The system triggers the alert notification process, which includes sending an email notification to a predefined recipient. The email containsrelevant information about the detected object, including the class label and confidence score.

Alert Script Execution. The alert notification is facilitated by executing an external Python script (”mail.py”). This script is responsible for composing and sending the email alert.

4 Implementation and Performance Metrics

The implementation is conducted using Python programming language, OpenCV library, numpy on a laptop with an Intel i7 processor, and 16GBRAM. Performance assessment is done based on Accuracy, precision, recall, and F1 score.

4.1 Dataset Description

The gun detection system's training dataset, obtained from Kaggle [24], is a product of collaborative efforts, featuring annotated images of robbers emphasizing locations. Rigorous processing and human-guided annotations guarantee high-quality data, while stringent privacy measures and ethical attribution standards are maintained. The size of the dataset is 1300 including instances of handguns, rifles, and knives, with specific images of 500 for handguns, 400 for rifles, and 400 for knives. This dataset is pivotal for training optimized convolutional neural network models, continuously evaluated for precision and performance enhancement.

4.2 GUI (Graphical User Interface)

The system features a login window enabling user access with a username and password (figure 3), supplemented by a registration option for new accounts. The registration form facilitates the entry of user details. Successful login directs users to the main dashboard displaying live surveillance camera feeds, selectable via a dropdown. A customizable settings tab enhances usability, allowing adjustments to camerasettings and enabling alerts for suspicious activity. In summary, this system offers efficient and precise surveillance camera detection, comprising login, registration, and camera customization functionalities,including a firearm detection component.

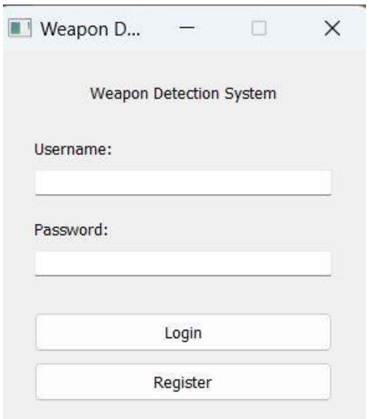


Fig. 3. GUI for registration of user

4.3 Training and Testing

The study comprehensively evaluated model performance across various training and testing ratios (90:10, 80:20, 70:30, 60:40) using key metrics such as accuracy, precision, recall, and F1-score. The 90:10 split demonstrated an 85% accuracy with balanced precision (92.9%), recall (80%), and F1-score (85%). Whereas, the 80:20 split has shown improved accuracy of 91.8%, with precision (95.8%), recall (98.5%), and F1-score (95.6%), indicating enhanced positive identification. The 70:30 split showed an accuracy of 87.8%, revealing a precision-recall trade-off (90%, 77.3%). The 60:40 split exhibited challenges with 84.8% accuracy, highlighting a potential issue in larger testing sets. Overall, the 80:20 split emerged as optimal, emphasizing the crucial role of training-to-testing ratios in model development for enhanced machine learning outcomes.

4.4 Performance Matrix

To evaluate the performance of a weapon detection system, four key metrics such as accuracy, precision, recall, and F1 score are used which play a pivotal role in providing a comprehensive evaluation of the classification model's effectiveness. These metrics are calculated from the confusion matrix as shown in figure 4. They enable stakeholders to gauge the model's ability to correctly identify positive cases, its capacity to detect actual weapons, the overall accuracy of its predictions, and the balance between precision and recall.

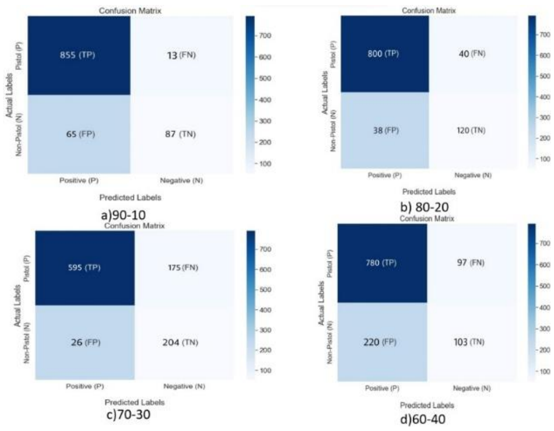
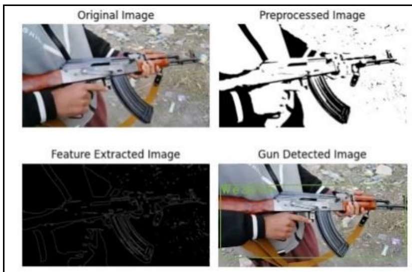


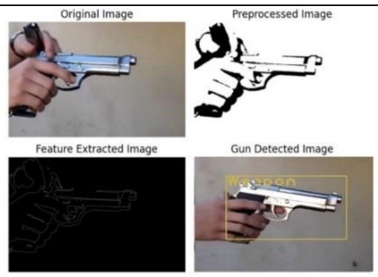
Fig. 4. Confusion matrix for different Training: Testing ratio

5 Results and Discussion

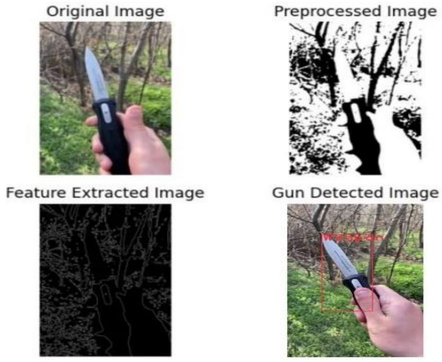
The proposed algorithm/model is trained and tested on different types of weapons that can be used during an act of robbery. The case-wise results are presented in Figure 5 along with the performance analysis in Table 1. AK-47 detection employs machine learning for rifle identification in images or videos. Preprocessing involves resizing, grayscale conversion, and noise removal. Relevant features like shape, size, and color are extracted. A model is trained on a dataset with AK-47 images to differentiate it. Similarly, Case 2: Gun and Case 3: Knife is detected by the algorithm successfully. Edge detection identifies handgun contours. CNN learning extracts features like the height of the object and key points' characteristics. Detection involves pinpointing regions of interest through proposalalgorithms or trained classifiers, ensuring accurate identification of handgun presence across diverse image scenarios.



(a) Case 1: AK47



(b) Case 2: Gun



(c) Case 3: Knife

Fig. 5. Case results with different weapons

Table 1. Performance parameters for different training:testing ratios

Metric (%)	Different Training: Testing ratio			
	90:10 Split	80:20 Split	70:30 Split	60:40 Split
Accuracy	0.85	0.918	0.878	0.848
Precision	0.929	0.958	0.90	0.780
Recall	0.80	0.985	0.773	0.889
F1-Score	0.85	0.956	0.857	0.831

The algorithm's performance was evaluated using varying Training: Testing ratios, as presented in Table 1. Through systematic experimentation, each ratio was meticulously tested to illuminate its impact on the algorithm's performance metrics. By systematically altering these ratios and observing corresponding outcomes, the study aimed to uncover patterns and trends that shed light on the algorithm's strengths and potential limitations. This meticulous examination not only enriches our understanding of the algorithm's behavior but also contributes to its refinement, fostering its utility across a broader spectrum of real-world scenarios. The graphical representation of this analysis is presented in figure 6. It can be seen the best results are achieved with 80:20 ratio.

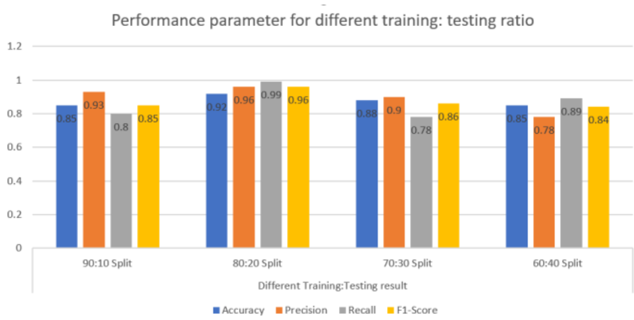


Fig. 6. Performance analysis for different Training:Testing ratio

It can be seen that the algorithm is able to detect the different types of weapons. This all has been done withan 80:20 algorithm of the dataset as the highest accuracy occurred with it. Additionally, exploring edge computing solutions could further optimize real-time processing and response.

Further the performance of proposed algorithm is evaluated by comparing the performance metrics with the results of other authors. It is presented in table 2.

Table 2. Comparative analysis of performance of proposed work

Metric (%)	Maryam Qasim Gandapur [8]	Mrunal Malekar [11]	Shankargoud Patil at.el [16]	Proposed
Accuracy	88	85	84	91.8
Precision	90	82	81	95.8
Recall	84	87	78	98.5
F1-Score	82	81	80	95.6

The Table 2 presents that the performance metrics values achieved with proposed work are higher than the other's result.

6 Conclusion

The proposed model employs Yolo architecture of the CNN algorithm to detect guns in real-time video surveillance, enhancing security and crime prevention for bank premises. Comparative analysis of algorithms considered factors like time complexity,space complexity, and response time, concluding CNNas the optimal choice. Evaluating the CNN model across different training-to-testing ratios, it is found that 80:20 to offer the best accuracy, precision, recall, and F1-score metrics. CNN-based gun detection system displays promise for real-time surveillance.

The future scope includes expanding the system's capabilities to detect other dangerous objects along with face masks which robbers usually wear and integrating a multi-camera.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ağdaş, M. T., Türkoğlu, M. and Gülseçen, S.: Deep neural networks based on transfer learning approaches to classification of gun and knife images. *Sakarya University journal ofcomputer and information sciences*, 4(1):131–141 (2021).

2. Alshammari, A. and Rawat, D. B.: Intelligent multi-camera video surveillance system for smart city applications. In *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0317–0323 (2019).
3. Ayed, E., Shahzad, S. K., Mustafa, R., Jaffar, M. A. and WaseemIqbal, M.: Deep neural networks for gun detectionin public surveillance. *Intelligent Automation & SoftComputing*, 32(2): ----- (2022).
4. Du, J.: Understanding of object detection based on CNN family and yolo. In *Journal of Physics:Conference Series*, volume 1004, page 012029. IOPPublishing (2018).
5. Dubey, S.: Building a gun detection model using deep learning. *Program Chair & Proceedings Editor: M. Afzal Upal, PhD Chair of Computing& Information Science Department Mercyhurst University*, 501 (2019).
6. Mohamed, M. K. E., Taha, A. and Zayed, H. H.: Automatic gun detection approach for video surveillance. *International Journal of Sociotechnology and Knowledge Development (IJSKD)*, 12(1):49–66 (2020).
7. Fathy, C. and Saleh, S. N.: Integrating deep learning-based IoT and fog computing with software-defined networking for detecting weaponsin video surveillance systems. *Sensors*, 22(14):5075 (2022).
8. Gandapur, M. Q.: E2evsdl: End-to-endvideo surveillance-based deep learning model to detect and prevent criminal activities. *Image and Vision Computing*, 123:104467 (2022).
9. Goyal, A., Anandamurthy, S. B., Dash, P., Acharya, S., Bathla, D., Hicks, D., Bhan, A. and Ranjan, P.: Automatic border surveillance using machine learning in remote video surveillance systems. In *Emerging Trends in Electrical, Communications, and Information Technologies: Proceedings of ICECIT-2018*, pages 751–760 (2020).
10. Khalid, S., Waqar, A., Tahir, H. U. A., Edo, O. C. and Tenebe, I. T.: Weapon detection system for surveillance and security. In *2023 International Conference on IT Innovation and Knowledge Discovery (ITIKD)*, pp. 1–7. Location (2023).
11. Mrunal Malekar. Detecting criminal activities of surveillance videos using deep learning. *IJSCSEIT* (2021).
12. Manikandan, V. P. and Rahamathunnisa, U.: A neural network aided attuned scheme for gun detection in video surveillance images. *Image and VisionComputing*, 120:104406 (2022).
13. Moura, N. S., Gondim, J. M., Claro, D. B., Souza, M. and Figueiredo, R. D. C.: Detection of weapon possession and fire in public safety surveillance cameras. In *Anais do XVIII Encontro Nacionalde Inteligência Artificial e Computacional*, pages 290–301 (2021).
14. Narayanan, M., Jaju, S., Nair, A., Mhatre, A., Mahalingam, A. and Khade, A.: Real-time video surveillance system for detecting malicious actions and weapons in public spaces. In *Computer Networks and Inventive Communication Technologies: Proceedings of Third ICCNCT 2020*, pp. 153–166, Location (2021).
15. Dammalapati, N., Gera, J. and Gera, I. P.: An intelligent suspicious activity detection framework (isadf) for video surveillance systems. *International Journal of Computer Applications*, 84(14): ----- (2019).
16. Patil, S. and Prabhushetty, K. S.: Detection of abnormal activity to alert the nearby persons via m-dnn based surveillance system. *WSEAS Transactions on Systems and Control*, 16:668–685 (2021).
17. Phyo, C. N., Zin, T. T. and Tin, P.: Deep learning for recognizing human activities using motions of skeletal joints. *IEEE Transactions on Consumer Electronics*, 65(2):243–252 (2019).
18. Qi, D., Tan, W., Liu, Z., Yao, Q. and Liu, J.: A dataset and system for real-time gun detection in surveillance video using deep learning. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 667– 672. IEEE, Location (2021).

19. Sumi, L. and Dey, S.: Yolov5-based weapon detection systems with data augmentation. *International Journal of Computers and Applications*, pages 1–9 (2023).
20. Vishwakarma, S. and Agrawal, A.: Asurvey on activity recognition and behavior understanding in video surveillance. *The Visual Computer*, 29:983–1009 (2018).
21. Warsi, A., Abdullah, M., Husen, M. N., Yahya, M., Khan, S. and Jawaid, N.: Gun detection system using YOLOv3. *In Proceedings of the 2019 IEEE International Conference on Smart Instrumentation, Measurement, and Application (ICSIMA)*, pp. 1–4. IEEE, Malaysia (2019).
22. Castillo, A., Tabik, S., Pérez, F., Olmos, R. and Herrera, F.: Brightness-guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning. *Neurocomputing*, Volume 330, pp. 151–161 (2019).
23. Narejo, S., Pandey, B., Vargas, D. E., Rodriguez, C. and Rizwan A. M.: Weapon Detection Using YOLO V3 for Smart Surveillance System. *Mathematical Problems in Engineering*, volume 2021, Article ID 9975700, 9 pages, (2021).
24. Kaggle datasets, “Guns Object Detection”, https://www.kaggle.com/datasets/issaisasank/guns-object-detection, last accessed 2023/01/07
25. Unkule, P., Shinde, C., Sarkar, P., Agarkar, S. and Verma, U.: CNN-based Approach for Sign Recognition in the Indian Sign Language. *In International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, pp. 92-97, IEEE, Trichy, India (2022).
26. Gaidhane, T., Shinde, S., Aurade, N. and Verma, U.: Smart Disease Prevention System. In *Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology* (ICERECT), pp. 1-6. IEEE, Mandya, India (2022).
27. S. Madkar, D. Y. Sakhare, K. A. Phutane, A. P. Haral, K. B. Nikam and S. Tharunyha, "Video Based Forest Fire and Smoke Detection Using YoLo and CNN," *2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, Chennai, India, 2022, pp. 1-5, doi: 10.1109/ICPECTS56089.2022.10046717.