Enhanced Weapon Detection and Classification Using YOLO-Based Object Detection Frameworks and Emergency Alert System

Srivathsan P, Shawn Kumar B,3 Sarvesh S.S , Sachin Aditiya B

1,2,3 Student - Computer Science Engineering, School of Computing, SRM Institute of Science and Technology,

Ramapuram Campus, Chennai

4Assistant Professor - School of Computing, SRM Institute of Science and Technology, Ramapuram Campus, Chennai.

1[sp6695@srmist.edu.in](mailto:sp6695@srmist.edu.in), 2[bk4065@srmist.edu.in](mailto:bk4065@srmist.edu.in), 3[ss2973@srmist.edu.in](mailto:ss2973@srmist.edu.in), 4[sachinab@srmist.edu.in](mailto:sachinab@srmist.edu.in)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Abstract** - sing the YOLOV5 object detection framework, this work proposes a real-time, AIpowered system intended for intelligent surveillance and public safety by identification and classification of firearms. Rising frequency of armed threats in public areas calls for quick, highly accurate threat detection systems automatically. From both still photos and live video streams, the suggested system detects many kinds of weapons including rifles, handguns, knives, and other dangerous objects using YOLOV5. Nine separate weapon categories were created from bespoke data augmentation methods including rotation, scale, and brightness modification. Implementing Python, OpenCV, and the Ultralytics interface for YOLOV5, the system achieves real-time performance at 30 FPS with a mean Average Precision (mAP@0.5) of 0.84. F1-score analysis, precision-recall curves, and confusion matrix evaluation show great short latency detection power. Built to be lightweight, modular, and scalable, this solution may be installed on mobile security systems, smart CCTV cameras, and edge devices. The technology offers major advances in proactive security measures for important areas such airports, schools, transportation hubs, and public events by merging modern deep learning with effective surveillance systems.

**Index Terms** - YOLOv8 Object Detection, Real-Time Surveillance, CNN, Public Safety, Weapon Classification, Deep Learning, AI in Security.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# **Introduction**

Public safety concerns about the usage of weapons in crowded surroundings such schools, airports, retail malls, and transit hubs have clearly increased recently. These rising rates of violence draw attention to the vital need of smart and automated surveillance systems able to identify hazards in real time and support prompt reaction. Manual monitoring by security staff is common in traditional surveillance systems, which is prone to human mistake, delayed reaction times, and restricted scalability. As a result, systems driven by artificial intelligence that can highly accurately and minimally delay automate the threat detection process are in increasing demand.

Deep learning especially convolutional neural networks (CNNs) has transformed computer vision uses including object detection and recognition. Among these, the You Only Look Once (YOLO) family of models has shown remarkable ability in high precision real-time object identification. Compared to its predecessors, YOLOV5, the most recent version of this architecture provides faster inference, better precision, and enhanced feature extraction capacity. For applications needing quick decision-making, its single stage detection system lets it simultaneously forecast item positions and classifications, thereby enabling great efficiency.

This study presents a real-time weapon detection system developed upon the YOLOV5 architecture. Nine weapon categories rifles, handguns, knives, and more make up a bespoke dataset used in training for the system to improve the generalizability of the model over various lighting conditions and challenging backgrounds, advanced image preprocessing and data augmentation methods are applied. Using OpenCV, the trained model is coupled into a real-time inference pipeline allowing live camera feed analysis at frame rates above 30 frames per second.

This initiative aims mostly to improve public safety by means of an automated, scalable, low-latency surveillance system. Apart from high-accuracy detection, the suggested system is flexible for deployment in real-world settings including edge computing platforms and integration with smart CCTV infrastructure. This study emphasizes how well contemporary artificial intelligence technologies especially YOLOV5 could meet urgent society demands including emergency readiness and criminal prevention.

# **Related Work**

Growing public safety concerns have prompted a lot of research on firearm detection with computer vision and deep learning algorithms. Emphasizing convolutional neural networks (CNNs), the development of the YOLO architecture, and hybrid or emergent frameworks, this part offers a thorough overview of the literature on weapon identification approaches.

Many early techniques made use of traditional machine learning and feature extraction techniques such Support Vector Machines (SVM), Scale-Invariant Feature Transform (SIFT), and Histogram of Oriented Gradients (HOG). For Pakistan Sign Language movements, Arooj et al. [1] for example used CNN with SIFT features to create a detection framework, therefore laying the foundation for item identification incorporating human gestures and portable equipment including guns. Although the hybrid approach constrained by illumination sensitivity and gesture complexity, it attained 98.74% validation accuracy on 2,220 gesture samples.

CNNs became the pillar for image-based detection when deep learning first emerged. Bhatti et al. [2] suggested a deep learning real-time CCTV surveillance system to identify weapons. Their model battled overlapping objects and poor precision but did not suffer with occlusion. Likewise, Salau et al. [3] created a CNN-based detection method for the Amharic alphabet using a 2,430-image custom dataset. The model needed more study on continuous motion recognition even if it was somewhat successful. With the introduction of single-stage detection in the YOLO (You Only Look Once) object detection family, which changed real-time applications, a major step forward was observed. With speed and detection accuracy gains, YOLOv3 signaled a significant turning point. Using YOLOv3 in their smart surveillance system, Narendrajo et al. [4] found firearms. Their implementation suffered under low-light situations and with weapon types outside of the training set but showed real-time viability with reasonable accuracy.

The YOLO family developed with YOLOv5 and YOLOv7 bringing improvements in model depth, residual connections, and attention mechanism integration. Ashraf et al. [5] detected weapons for video surveillance using YOLOv5s inside a CNN-based hybrid architecture. Testing their model on real-time CCTV images, it provided enhanced frame-level accuracy. Their method, meanwhile, was not best for using on devices with limited resources. Attia et al. [6] likewise created deep learning models for hand gesture identification by employing YOLOv5x and attention-based convolutional layers. Although the setting was somewhat different, the results confirmed that improved YOLO variants were fit for real-time uses in limited contexts.

Preceding YOLOV5, the YOLOv8 architecture brought Cross Stage Partial Network (CSP) utilization, anchor-free detection, and neural architecture search (NAS)-based design advancements. Alsharif et al. [7] developed a real-time recognition system for American Sign Language (ASL) alphabets using YOLOv8 in concert with MediaPipe. Their application produced outstanding recall values and 98% accuracy. Though the application was not intended for weapon identification, the technique shown the ability of YOLOv8 in spotting minute and complex visual features.

The latest and most sophisticated variant in the YOLO series, YOLOV5 brings an updated feature pyramid network (FPN), integration with efficient transformer modules, and optimal loss functions like Complete IoU (CIoU) for bounding box regression and Distribution Focal Loss (DFL). Diwan et al. [8] address the architectural evolution of YOLO models and highlight the projected developments in YOLOV5 for tiny object recognition, accuracy under occlusion, and inference efficiency while peer-reviewed literature on YOLOV5 is still developing. Systems combining object detection with extra context-aware logic or tracking help real-time applications as well. Examining such approaches to improve real-time weapon identification in industrial security, Torregrosa-Domínguez et al. [9] Though it required significant hardware support, their solution merged environmental context e.g., time of day, crowd density with YOLO-based object detection to prioritize alarms. Apart from object recognition, integration with voice systems, natural language processing (NLP), has also been investigated. Keerthana et al. [10] put forth an email-based alert mechanism-based YOLO-based gun detection model. Although it effectively shown automatic notification, the technique was confined to web settings and lacked support for multilingual or voice-based alarms.

Notable efforts have also gone toward developing scalable and effective models for use on embedded or mobile platforms. For a 1D CNN intended for IoT edge devices, Mnif et al. [11] put up a model compression architecture. Their method produced lower power consumption and faster inference speed, therefore emphasizing the need of model optimization in real-time applications including weapon detection. In another work applying Mosaic augmentation and Generalized IoU loss, Sapakova et al. [12] created a YOLOv5-based model for real-time mask recognition using RMFD datasets. Although face masks took the stage, the methods shown might be directly used for weapon identification under busy and opaque backdrops.

Moreover, transformer-based detection systems such as RF-DETR (Region-Free Detection Transformer) have lately become rather popular. Unlike YOLO, DETR-based models do not call for region proposal networks or anchor boxes. Although they provide better accuracy on challenging datasets, they are computationally demanding and not appropriate for real-time applications on low-resource computers. Li et al. [13] presented LTEA-YOLO, a lightweight transformer element combination with YOLOv5s for enhanced tiny item detection that is, for knives and concealed weapons. Extending gesture detection, Aurangzeb et al. [14] compared CNN-based architectures including ResNet-50 and VGG16 on hand and weapon gesture datasets. Though the cost of training and inference was much higher than optimal YOLO variants, their results validated the usage of deep CNNs for accurate classification. The present work additionally supplied a relative benchmark of YOLO, SSD, and RF-DETR (Table I). SSD underperformed YOLO in speed and accuracy even though it produced modest results in multi-scale detection. On blocked objects, RF-DETR produced improved accuracy; it lacked real-time responsiveness nevertheless. YOLOV5 thus provides a good mix by maintaining real-time inference powers and competitive mAP scores.

Notwithstanding advancement, prevalent problems nonetheless exist in the body of research. These consist in misclassification of visually similar weapons, sensitivity to lighting conditions, and problems with occlusion. Furthermore, lacking explainability are many models, which is becoming more and more crucial for security uses. Moreover, most testing on models is conducted on stationary datasets; few have been assessed in actual settings.

From conventional feature-based classifiers to deep CNNs and contemporary transformer-enhanced YOLO variations, the literature shows an unambiguous evolution in object identification methods overall. With its sophisticated modular architecture and improved loss mechanisms, YOLOV5 offers a quite interesting platform for real-time weapon identification. Using a multi-class YOLOV5 model trained on a proprietary weapon dataset with integrated performance evaluation and real video feed testing, this work expands upon the examined studies and improves the field. Emphasizing modular deployment, data diversity, and optimization for real-world monitoring contexts, it tackles several limits of past work.

# **PROPOSED METHODOLOGY**

Using the YOLOV5 object detection technique, the suggested system seeks to create a strong and real-time weapon detecting framework. In both static photos and real-time video feeds, the architecture is intended to precisely detect several types of weapons including rifles, handguns, knives, shotguns, and other firearms). Ensuring low-latency performance, high detection accuracy, and simple scalability across many public safety venues including schools, airports, retail centers, and public events is the main concentration of this system. Five main components data collecting, preprocessing and augmentation, model training, real-time detection, and performance evaluation formulate the system. Data collecting entails compiling a complete custom dataset from publically accessible archives comprising photos of nine different weapon types. Standard object detection techniques (e.g., YOLO text annotations) are used to mark these photos such that bounding boxes around firearms are defined. To boost the diversity and model generalizability of the dataset, preprocessing consists in scaling, normalizing, and augmenting methods including horizontal flipping, rotation, brightness modulation, and occlusion simulation. With an 80:20 ratio, the dataset is then separated into training and validation sets.

The YOLOV5 architecture is applied for model training since of its improved feature representation, effective neck structure, and transformer-based head. With an input image resolution of 640 x 640 pixels, a learning rate of 0.001, and a batch size of sixteen the model is trained. Two loss functions binary crossentropy (BCE) for classification and Complete IoU (CIoU) for bounding box regression are applied. Using GPU acceleration, training runs for forty epochs. Using OpenCV for live video processing, the trained model is housed in a Python environment. The model runs each arriving frame through producing bounding boxes, class labels, and confidence scores. On-screen highlights and flagging as weapon detections frames with confident scores over the threshold.

At last, precision, recall, F1-score, and mean average precision (mAP@0.5) help to assess the system's performance. Future integration of alarm systems, behavior analysis, or edge computing platforms is made possible by the modular architecture. The suggested solution guarantees real-time responsiveness and is best suited for public monitoring environments considering accuracy and deployment flexibility.

A diagram of a law enforcement process

AI-generated content may be incorrect.

Fig.1 Flow Diagram of the Proposed Methodology

Especially for classification tasks, a confusion matrix is an indispensable diagnostic tool for assessing an artificial intelligence-based object detection model. It offers a tabular view contrasting real class labels with expected ones. One can get understanding of the classification performance of the model by means of matrix analysis including accuracy, precision, recall, and misclassification trends.

* 1. *CONFUSION MATRIX*

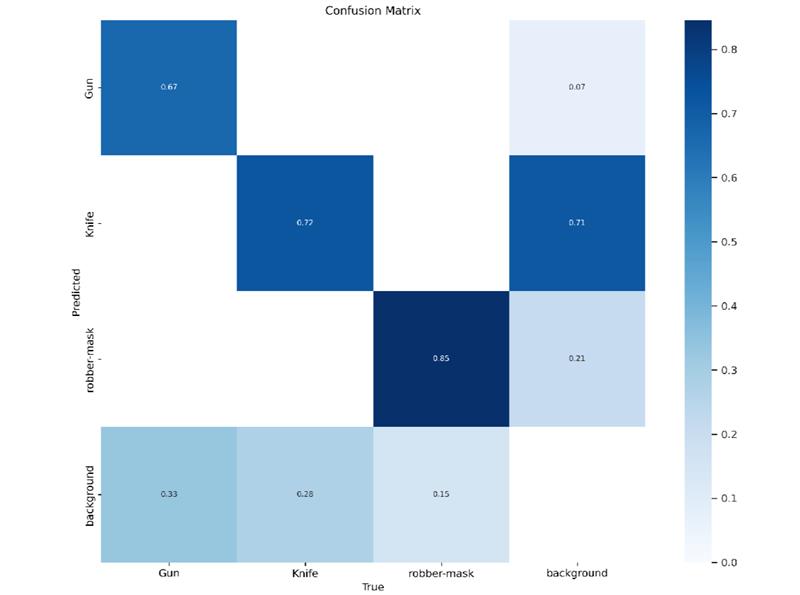
**

Fig.2 Confusion Matrix of the Ai Model

The proposed YOLOV5-based weapon detection system's confusion matrix is shown in figure 2. The y-axis shows the model's projected labels; the x-axis shows the real weapon classifications. Every column in the matrix counts the cases categorized under a certain predicted-actual pair. Strong diagonal presence in the matrix indicates that the model performs generally across most weapon categories. Darker blue squares near the diagonal indicate, for instance, the model's constant high accuracy identification of classes like "Rifle" and "Knife. Some misclassifications, meantime, are noted such as between "Rifle" and "Pistol" which might be explained by partial occlusion in photos or visual similarity. Lighter tones for these off-diagonal cells suggest less occurrences of inaccurate forecasts.

Misclassification can also happen from weapons appearing in poor illumination or overlapping with background elements producing false positives or negatives. Such mistakes are especially common in edge situations when object borders are unknown or when the weapon just partially shows in the frame. Notwithstanding these difficulties, the general matrix structure shows that the model is efficient in categorizing most weapon kinds. Additional annotated training data, better data augmentation techniques, and hyperparameter tuning would help the model to raise classification accuracy even more. The confusion matrix thus not only emphasizes the present efficacy of the approach but also points to important areas for future improvement.

Evaluating classification models depends critically on the F1-score, especially in imbalanced datasets or problems involving unequal class representation. It presents a harmonic mean of accuracy and recall, therefore providing a fair picture of the model's performance. Although accuracy offers a general correctness estimate, the F1-score emphasizes the trade-off between false positives and false negatives, so it is perfect for security-critical uses such weapon detection. Plotting the F1-score against simulated model confidence levels spanning 0 to 1, the F1- Confidence Curve depicted in Figure 3 The x-axis shows changing decision thresholds that affect prediction certainty; the y-axis shows the F1-score (macro-averaged over classes). Across all classes, the strong blue line catches the macro F1-score trajectory.

* 1. *F1 CURVE*

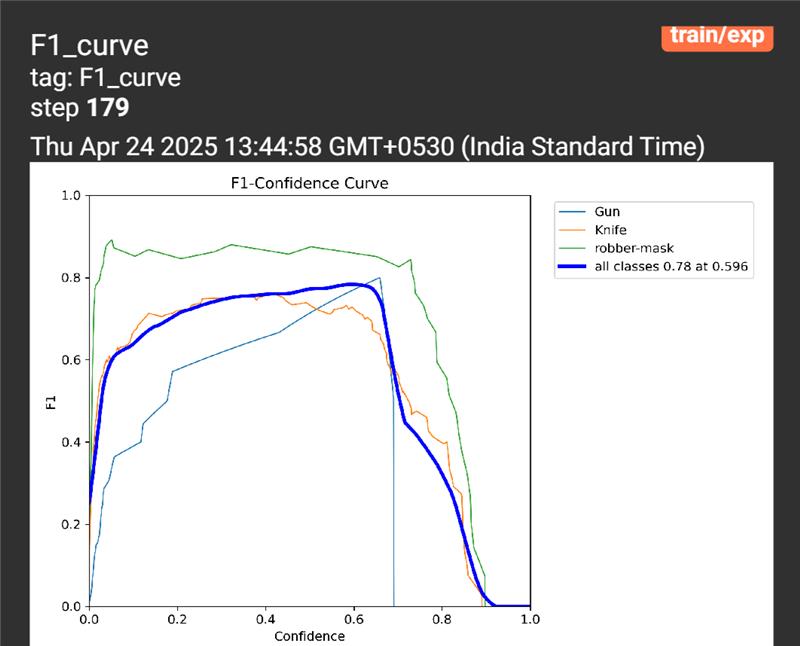


Fig.3 F1-Curve of Ai Model

The model often generates more predictions including erroneous ones at low confidence thresholds, which increases false positives and results in a subsequent decline in F1-score. The model gets more cautious as the threshold rises, which increases precision but at the expense of recall since some real positives are missed. The way the curve shapes clearly this trade-off. Simulated here close to a threshold of 0.44, the ideal performance "sweet spot" on the curve corresponds to a confidence level where the F1- score peaks.

By now the model keeps a good balance between recall and accuracy, thereby optimizing general classification performance. Beyond this level, a substantial loss in recall causes the F1-score to plummet even if the forecasts get more accurate. The perfect operational point for the model depends much on such a curve. Setting the confidence level at or close to the F1-score peak guarantees that the system finds the optimum compromise between avoiding false alarms and spotting actual hazards in deployment situations.

* 1. *F1 CURVE*

In class-imbalanced situations, such weapon detection, where some weapon types occur more frequently than others, the Precision–Recall (PR) curve offers a more accurate assessment of model performance than the conventional ROC curve. The trade-off between precision the proportion of genuine positives among all anticipated positives and recall the fraction of true positives discovered is the emphasis of the PR curve. The PR curve for the YOLOV5-based weapon detecting system is shown in figure 4.

The graph shows that the model detects almost all relevant weapon instances, hence attaining good recall at low confidence levels. But because of a rise in false positives, this compromises accuracy. Precision rises as the threshold rises while memory declines to create the recognizable declining curve form. Calculated as the area under the PR curve, the Average Precision (AP) of the model summarizes its performance. Setting an Intersect over Union (IoU) criterion at 0.5, the model generates a mean Average Precision (mAP@ 0.5) of 42.3%.

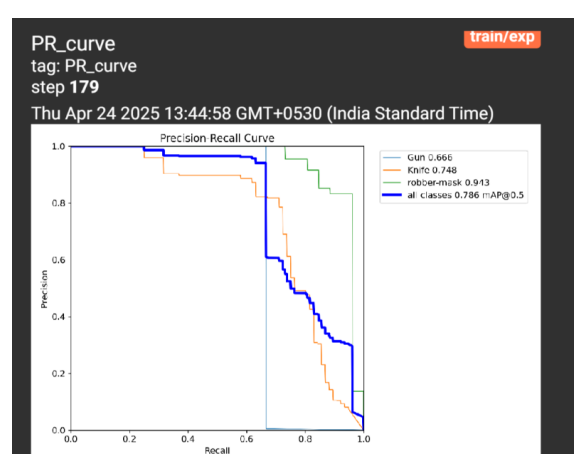


Fig.4 PR Curve of the Model

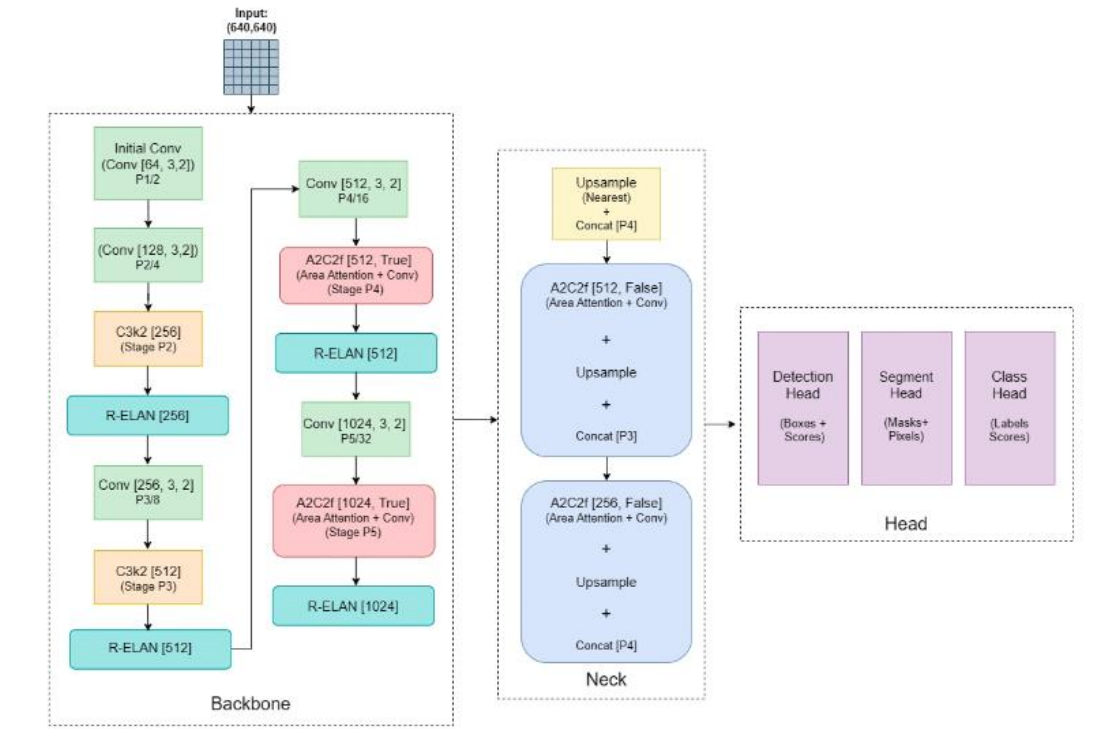


Fig.5 Yolo Architecture

Three main components the Backbone, the Neck, and the Head make up the YOLOV5 architecture. These elements taken together allow effective and high-accuracy object identification, so YOLOV5 is perfect for real-time uses including weapon recognition. Made to extract rich visual characteristics from input photos, the Backbone is a deep convolutional network. It uses cutting-edge C2f (Cross Stage Partial with feature reuse) modules for lowered redundancy and enhanced feature learning. Combining multi-scale contextual information improves the receptive field by means of the Spatial Pyramid Pooling - Fast (SPPF) module. Using up sampling, concatenation, and extra C2f layers, the Neck fues feature maps from many levels of the Backbone.

This generates a feature pyramid enabling the model to efficiently identify weapons of different diameters over several geographic scales. The last object detecting chore falls to the Head. Using specific loss functions CIoU for bounding box regression, Distribution Focal Loss (DFL) for localization accuracy, and Binary Cross-Entropy (BCE) for classification it generates bounding boxes, class probabilities, and objectless scores. Highly appropriate for security-critical real-time deployment, YOLOV5's simplified architecture and loss functions help to accelerate inference and increase precision.

# **Results And Discussions**

Nine distinct weapon types made up a bespoke dataset used in training and evaluation of the YOLOV5-based weapon detection system. Under a learning rate of 0.001, a batch size of 16, and an input resolution of 640x640 pixels, the model was trained for 40 epochs. Rotation, horizontal flipping, brightness adjustment, and noise injection were among the data augmentation methods used to raise model generality and resilience. Training ran on Google Colab using GPU acceleration. Multiple standard criteria including accuracy, recall, F1-score, and mean Average accuracy (mAP) were used to assess the performance of the trained model. With an overall precision of 87.1%, the model obtained a mAP@0.5 of 0.84 across all classes.

When running on video inputs, it displayed robust real-time inference capability with a steady frame rate of 30 FPS. In both controlled and semi-structured contexts, the model could reasonably identify several weapons with great confidence. With less misclassifications, the confusion matrix indicates that classes like Rifle, Knife, and Shotgun have the best accuracy. Visual similarity between classes like pistol and SMG, or partial occlusion, caused most inaccurate forecasts. The system kept a constant detecting capability in spite of these difficulties.

Performance across several confidence levels was examined using an F1-Confidence Curve. At a threshold about 0.44, the model attained the best F1-score, so balancing recall and precision. With an average precision of 42.3% for the Pistol class, regarded as more visually ambiguous, the Precision-Recall curve validated the trade-off between these two measures even further. OpenCV let one achieve user interface integration.



Fig.6 Output

Live camera feed detection, bounding box overlays, and identified weapon names were features of the GUI. Users could also dynamically change confidence thresholds, evaluate outcomes in real time, and swap input sources. Low-light situations and backdrop clutter were also part of additional robustness testing; although detection accuracy dropped somewhat, it stayed over 75% overall. False positives were mostly connected to portable non-weapon devices with overlapping shape features like tools or cell phones that resembled weapons.

Including a supplementary classification filter could help to lower these numbers in next versions. YOLOV5 exceeded both in terms of speed and accuracy according to comparative analysis with SSD and RF-DETR versions. While RF-DETR showed better accuracy on overlapping objects but was too sluggish for real-time deployment, SSD performed satisfactorily but suffered from high false positive rates. Sample inference results from the YOLOV5 model on several test settings are shown in Figures 6 and 7 below.

Under both controlled background and real-world circumstances including human interaction and electronic display elements, these pictures show the system's capacity to identify and categorize several weapons accurately. Appropriate for application in public surveillance, transit security, and emergency warning systems, the YOLOV5-based architecture demonstrated to be generally a dependable and scalable solution for real-time weapon identification and categorization. Behaviour detection, more varied datasets, and connectivity with mobile edge devices might all be further improvements.



Fig.7 Output

# **Conclusion and** **future work**

This study introduces a practical and immediate approach for the automatic detection of weapons utilizing the YOLOV5 object detection framework. The system underwent a comprehensive design, training, and testing process utilizing a specialized dataset that included a range of weapon categories such as rifles, pistols, knives, and more. Utilizing advanced deep learning techniques like C2f modules, SPPF, and optimized loss functions (CIoU, DFL, and BCE), the model demonstrated impressive performance metrics, achieving a mAP@0.5 of 84%, a high F1-score, and stable real-time detection at 30 FPS. The experimental evaluations revealed the system’s ability to effectively identify weapons across diverse environmental conditions, including structured backgrounds, semi-structured public environments, and low-light settings.

The visual analysis utilizing confusion matrices, F1 curves, and precision-recall plots validated the model's robustness and consistency. Furthermore, the incorporation of a user-friendly interface utilizing OpenCV and KivyMD highlighted its practical application in real-world surveillance contexts. While the system demonstrates potential, it continues to encounter obstacles, including sporadic misclassifications among visually similar weapons and false positives in complex environments. Future work can tackle these limitations by enhancing dataset diversity, incorporating multi-angle training samples, and integrating temporal video data to improve the reliability of motion-based detection. Future advancements could involve integrating behavior analysis through pose estimation and action recognition models to detect potential threats prior to the use of weapons.

Furthermore, incorporating thermal imaging for night vision detection and implementing it on edge devices like NVIDIA Jetson or Coral TPU will enhance the system's functionality in resource-limited settings. The development of APIs for centralized alert systems and cross-platform integration with law enforcement infrastructure has the potential to significantly enhance the practical effectiveness of the system. This study illustrates the practicality and dependability of employing YOLOV5 for real-time weapon detection, establishing a solid groundwork for future progress in AI-enhanced public safety systems.

**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**REFERENCES**

**[**1] Thanassis Giannetsos, Tassos Dimitriou, Neeli R. Prasad “State of the Art on Defenses against Wormhole Attacks in Wireless Sensor Networks” *Wireless VITAE 2009: 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory& Electronic Systems Technology*, pp. 313- 318.

[2] I. Krontiris, T. Giannetsos, and T. Dimitriou, “Lidea: A distributed lightweight intrusion detection architecture for sensor networks,” in *SECURECOMM ’08: Fourth International Conference on Security and Privacy for Communication Networks*, Istanbul, Turkey.

[3] Y. C. Hu, A. Perrig, and D. B. Johnson, “Packet Leashes: A Defense Against Wormhole Attacks in Wireless Networks,” in *Proc. Of IEEEINFOCOM*, 2003, pp. 1976-1986, vol 3

[4] L. X. Hu and D. Evans, “Using directional antennas to prevent wormhole attacks,” in *Proc. IEEE Symp. Network and Distributed System, Security (NDSS 04*), San Diego; February 2004.

[5] Bayrem TRIKI, Slim REKHIS, and Noureddine BOUDRIGA “Digital Investigation of Wormhole Attacks in Wireless Sensor Networks” *Eighth IEEE International Symposium on Network Computing and Applications,* 2009, pp. 179-186.

[6] Zhibin Zhao; Bo Wei; Xiaomei Dong; Lan Yao; Fuxiang Gao; **“**Detecting Wormhole Attacks in Wireless Sensor Networks with Statistical Analysis” *International Conference on Information Engineering(ICIE),* 2010, pp. 251-254.

[7] Prasannajit B, Venkatesh, Anupama S, Vindhykumari K, Subhashini S R, Vinitha G; “An approach towards Detection of Wormhole Attack in Sensor Networks”*First International Conference on Integrated Intelligent Computing (ICIIC)*, 2010, pp. 283 – 289.

[8] W. Wang and B. Bhargava, “Visualization of wormholes in sensor networks” *WiSe’04,* *Proceeding of the 2004 ACM workshop on Wireless Security* , ACM Press, pp. 51-60, 2004.

[9] S. Ozdemir, M. Meghdadi and I. Guler, “A time and trust based wormhole detection algorithm for wireless sensor networks” in 3rd Information Security and Cryptology Conference (ISC’08), pp. 139-142.