Grenze International Journal of Engineering and Technology, June Issue

AI Powered Proactive Surveillance: Design and Implementation of Weapon Detection using YOLOv8 and Deep Learning Techniques

Bhavya[1], Chaya N.D[2], Jazim Muhsil[3], Abhijai Ranjith T[4] and Yogeesha C.B[5]
[1-4]Department of Computer Science and Engineering (AI and ML)
Email: bhavyanaik063@gmail.com, chayaprakash2004@gmail.com, jasimmuhsil010@gmail.com, abhijairanjith123@gmail.com
[5]Department of Artificial Intelligence and Machine Learning, Mangalore Institute of Technology and Engineering, Moodbidri, Mangalore, Karnataka – 574 225, INDIA
Email: yogeeshacb@mite.ac.in

***Abstract*— For real-time, low-false-alarm surveillance, the system integrates YOLOv8 weapon detection with Convolutional Neural Network-based behavioral analysis. It attains 84.1% accuracy for behavior analysis, 87% mAP@50 for weapon identification, and 88% mAP with 32 ms inference time. It gives real-time notifications with location and timestamps by processing CCTV (Closed-circuit television) feeds on edge devices. Because it balances accuracy, speed, and modular integration, it performs better than YOLOv5(You Only Look Once version 5) and YOLOv7(You Only Look Once version) and may be deployed in public areas.**

***Index Terms*— YOLOv8(You Only Look Once version) weapon detection,Convolutional Neural Network-based behavioral analysis, Real-time surveillance, mAP accuracy, Edge device processing.**

I. Introduction

Due to an increase in occurrences involving firearms, prohibited goods, and tension among students, campus safety has become a serious concern. Closed-circuit television (CCTV) and faculty monitoring are examples of traditional surveillance methods that are constrained by human weariness and slow reaction times. Thus, early threat identification requires automated, real-time intelligent systems.

This work suggests an integrated system that combines behavioral and emotional analysis with You Only Look Once version 8 (YOLOv8) weapon identification. While the behavioral module detects fear and anxiety, You Only Look Once version 8 (YOLOv8) uses photos, videos, and live feeds to identify weapons like guns and knives. When combined, they facilitate quicker reactions and lessen manual monitoring.

The system, which was created with Python, Ultralytics, Flask, Simple Mail Transfer Protocol (SMTP), and DeepFace, has an easy-to-use interface and automatically creates warnings with timestamps and thorough threat descriptions. It provides educational institutions with a dependable, proactive safety solution that reduces human error and provides early alerts while operating in real time.

Grenze ID: 01.GIJET.12.2.269_21

A. Main Contributions

- Dual-Function System: This system combines weapon detection and behavioral analysis on a single platform to ensure campus safety.
- Real-Time Processing: Using learned models and a user-friendly web interface, the Flask client-server model enables quick image processing.
- You Only Look Once (YOLOv8) Implementation: Reliably and precisely identifies weapons, including handguns, baseball bats, knives, and firearms.
- Accessible Design: Designed to be used by non-technical personnel, such as administrators and teachers, without requiring a high level of machine learning expertise.

II. Relative Study on Existing Systems

Real-time weapon identification has significantly improved thanks to recent developments in edge computing, deep learning, and multimodal sensing. According to [1] Berardini et al., Edge Artificial
Intelligence (Edge-AI) with super-resolution improves weapon visibility in low-resolution Closed-circuit television (CCTV) footage [13]. Partially hidden firearms can also be found with the aid of context-aware techniques that use human-pose estimation [2].

The subject is still dominated by deep learning. While [4] You Only Look Once–Multiscale (YOLO-MS) enhances the detection of small or ambiguous guns, Convolutional Neural Network (CNN)-based handgun detection achieves real-time performance [3]. [12]Low-light weapon identification is strengthened by multispectral imaging, which includes thermal, infrared, and Red Green Blue (RGB) [5].

Strong Convolutional Neural Network (CNN) feature extraction, enhanced detectors and tracking techniques [6], and Convolutional Neural Network (CNN)–Transformer hybrids are examples of architectural advancements [7]. Research on multimodal fusion investigates millimeter-wave radar[14], thermal imaging for concealed weapons, and anchor-free X-ray imaging [10].

For quick and precise identification, the You Only Look Once (YOLO) family is still frequently utilized [9]. With its improved backbone[8], decoupled detection heads, and anchor-free predictions, You Only Look Once version 8 (YOLOv8) is appropriate for edge deployment [17]. In complex circumstances, transformer-based models such as Vision Transformer (ViT) perform well[18], and hybrid techniques that
][combine You Only Look Once and Single Shot Detector (SSD) reduce false alarms [16].

In general, contemporary research focuses on multimodal, edge-ready, and extremely accurate weapon detection; You Only Look Once version 8 (YOLOv8) excels in real-time flexibility, speed, and precision.

III. System Design

The system employs a client-server architecture in which the server handles all calculations and the client communicates via a web interface.[11] Models may be upgraded or retrained without affecting input processing or the user interface thanks to its modular design.

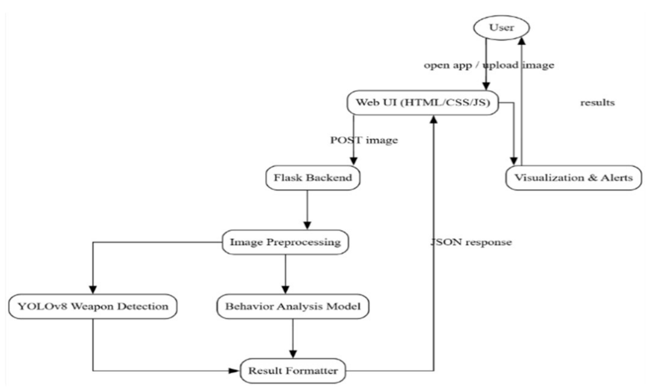

Figure 1. System Architecture Diagram

A web browser interface created with JavaScript, Cascading Style Sheets (CSS), and HyperText Markup Language (HTML) allows users to access the system. The Flask backend server receives an image or video that users upload. Two trained models are loaded by the server: a behavior analysis model for emotional recognition and You Only Look Once version 8 (YOLOv8) for weapon detection. Following picture preprocessing, the DeepFace model examines facial behavior, while YOLOv8(You Only Look Once version 8) uses bounding boxes to identify firearms. The Simple Mail Transfer Protocol (SMTP) is used to automatically send an alarm to the registered email address if a weapon is found.

The frontend receives the processed results from the backend and presents them to the user in an easily comprehensible manner.A weighted combination of context, edges, color, texture, and form is represented by X in Equation [1]. Higher weights are assigned to the shape (S) and edge (E) attributes within X. A Cross-Stage Partial with 2f modules (C2f) Convolutional Neural Network (CNN) backbone is used to extract spatial data [15]. A weighted total of all the features is created by combining the (Convolutional Neural Network) CNN-extracted data with manually supplied contextual features, including position or motion cues. Using CNN(Convolutional Neural Network) layers to extract shape and texture data.

$$X=W1S+W2T+W3C+W4E+W5Ct \tag{1}$$

The model classifies detected objects as weapons. While deeper layers detect texture elements like metal surfaces, shape features capture outlines and geometric details. Convolutional filters identify sharp edges that are essential for weapon outlines, and RGB inputs aid in identifying common weapon colors. Environmental context is also taken into account, such as the location of the holster or hand posture. The object is designated as a weapon and an alert is triggered if the score X X, which is derived from the weighting of each attribute (W1 to W5), exceeds a certain threshold. Figure.3 displays the use case diagram that depicts actor interactions and system capabilities, while Figure.2 shows the object-oriented design of the system with classes, properties, and methods.

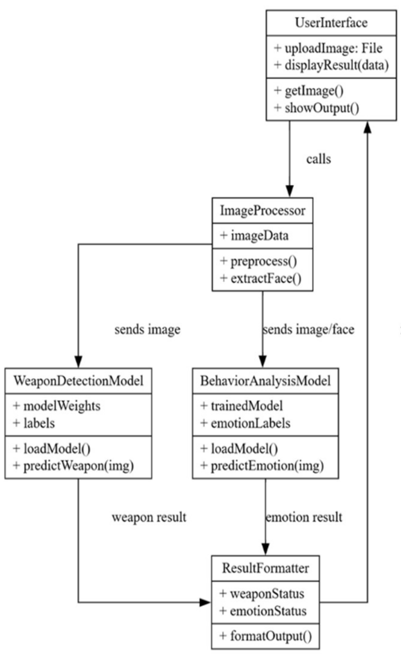

Figure 2. Class Diagram

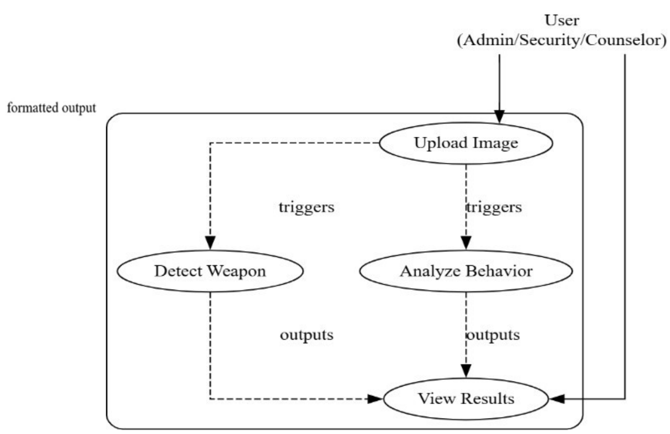

Figure 3. Use Case Diagram

IV. Experimental Setup

A contemporary object detection pipeline's sequential computing steps are shown in Figure 4, which is referred to as the Data Flow Diagram.

Frame differencing allows for pixel-wise subtraction between successive frames, which helps identify motion areas and can be used to effectively assess change regions. After that, the YOLOv8(You Only Look Once version 8) model receives each frame (or region of interest) for the Detection of Objects stage[14]. There, the neural network generates bounding boxes around potential weapons, predicts item placements, and produces class labels and confidence ratings. The Draw Bounding Box Around step, which is crucial for alert auditing and user interfaces, is used to represent these predictions on frames.

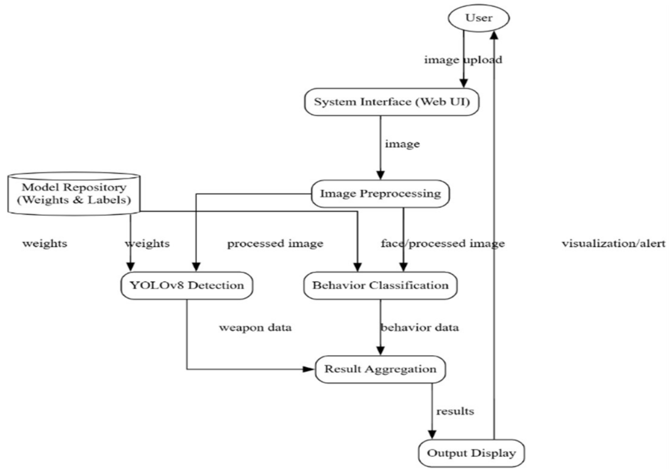

Figure 4. Data Flow Diagram

In Object Classification, as mentioned in equation (1) each detected object is further identified. Yolov8(You Only Look Once version 8) CNN (Convolutional Neural Network) already handles a lot of this, several deployments and studies use additional classifiers or feature extraction pipelines to verify or classify objects as knives, weapons, or other items.

Model training and testing were performed on a system with the following specifications:

- CPU: 11th Gen Intel Core i5-11400H
- GPU: NVIDIA GeForce RTX 2050, 4 GB VRAM
- RAM: 16 GB DDR4
- OS: Windows 11 (64-bit)
- IDEs: Google colab, Visual Studio Code

V. METHODOLOGY

A controlled simulation environment, deep learning models, and a proprietary dataset were used to assess the system's performance. Preprocessing, model training, backend Flask integration, frontend output visualization, and dataset preparation were all part of the workflow.

A. Dataset Preparation and Preprocessing

A proprietary collection of more than 3,000 photos of guns, knives, and bats is used by the weapon detection module. In Figure 5 Pixel values are normalized and images are scaled to the model's input dimensions. Rotation, flipping, and brightness adjustment are examples of data augmentation techniques that increase dataset diversity and decrease overfitting. To guarantee dependable model performance, the dataset is divided into training, validation, and testing sets. No additional dataset is needed for behavioral analysis since face emotion recognition is handled by the DeepFace package.

Figure 5. Model Training Logs (YOLOv8)

B. Model Development and Training

The system employs You Only Look Once version 8 (YOLOv8) for weapon identification and a deep learning model for emotion interpretation. The bespoke weapon dataset with pre-trained weights is used to fine-tune YOLOv8(You Only Look Once version 8), which is selected for its high speed and accurate object localization. Convolutional Neural Network (CNN) models like MobileNet, Visual Geometry Group 16 (VGG16), or Residual Network (ResNet) are utilized for behavioral analysis using transfer learning, where only the classification layers are retrained for emotion categories. Forward propagation, backpropagation, loss computation, and performance monitoring validation are all part of the training process.

C. Backend Integration using Flask

Flask, which handles user requests and makes predictions, is used to incorporate the trained models into the backend. The backend preprocesses uploaded images, applies the behavior analysis and weapon detection models, and outputs the results in JavaScript Object Notation (JSON) format. While the behavior model uses facial expressions to categorize emotions, the weapon model recognizes different kinds of weapons. Easy upgrades are possible without interfering with the workflow because of a modular backend structure that divides model loading, preprocessing, and prediction into separate Python scripts. The frontend, which uses JavaScript, Cascading Style Sheets (CSS), and HyperText Markup Language (HTML), allows users to upload photos asynchronously and shows results with behavior labels and weapon bounding boxes.

D. Dataset Preparation

Bounding box regression loss, as mentioned in the Table.1 objectness loss, and classification loss are the three main components mention in equation (3) of the total loss (L_total) for YOLO(You Only Look Once version 8) models such as YOLOv8(You Only Look Once version 8).

TABLE I: DATASET PREPARATIONS

Parameter	Values
Epochs	100
Batch Size	16
Image Size	640 * 640 pixels
Optimizer	SGD
Split (Train/Val/Test)	70 / 20 / 10

E. Evaluation Metrics

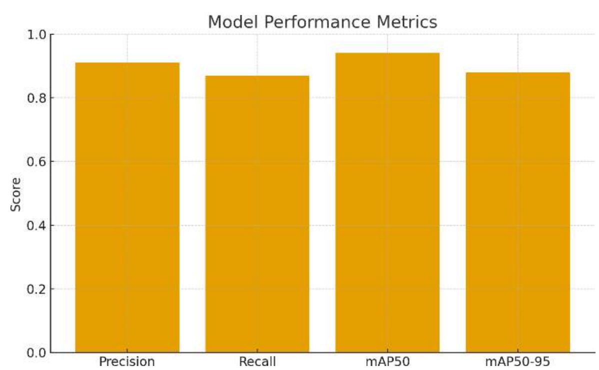

Figure 6. Evaluation Metrics

Standard object-detection criteria, including precision, recall, mAP50, and mAP50–95, were used to assess the model's performance (Figure 6). While recall shows the model's capacity to identify every weapon found in the actual world, precision assesses the accuracy of weapon detections without false alarms. A thorough evaluation of localization and classification robustness is provided by mAP50–95, which averages over IoU thresholds from 0.50 to 0.95. mAP50 assesses detection accuracy at an IoU threshold of 0.50.

VI. RESULT AND DISCUSSION

Even when confronted with difficult situations such as occlusion and poor light, the YOLOv8(You Only Look Once version 8) based weapon identification system exhibits great accuracy in identifying guns. Metrics including precision, recall, and mAP increased steadily throughout training and held up well on test data. Although there is some small misunderstanding in complicated situations, the confusion matrix Fig.9 indicates that the majority of classes, including background and weapons, are correctly identified.

The terminal output from training a YOLO(You Only Look Once version 8) model for weapon detection is shown in Figure 7. Details such as GPU memory utilisation, various loss values (box, class, and DFL), and evaluation metrics from Table 1, such as mAP scores over five epochs are displayed. The accuracy of the training process was about 0.84.

The classification accuracy of the model is shown graphically in Figure 8. The columns represent anticipated labels, and the rows represent actual labels and numerous accurate classifications for "weapon-without-background," there is also some misclassification between the weapon and background, suggesting that there is potential for improvement

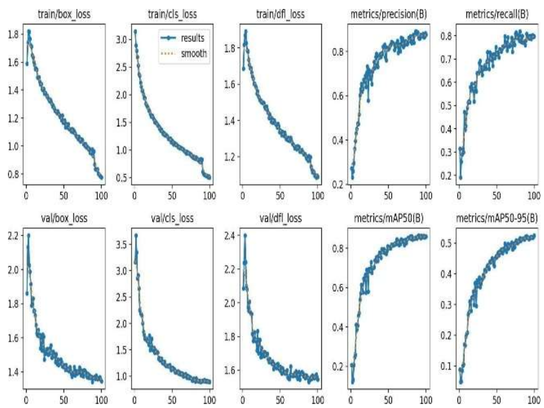

Figure 7. Model Training Logs (YOLO)

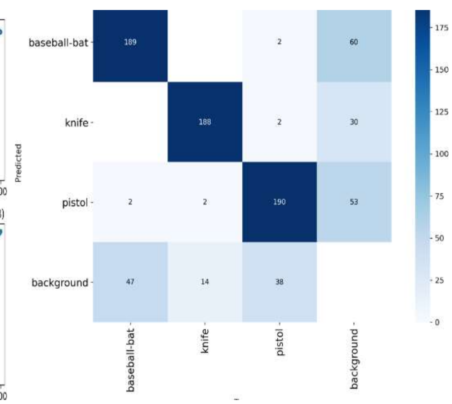

Figure 8. Confusion Matrix for Weapon Detection

Other well-known object identification architectures like(you only look once version) YOLOv5, YOLOv7, and SSD are outperformed by the suggested YOLOv8 model. Out of all the models, YOLOv8 has the greatest Precision (0.876), Recall (0.798), and mAP scores (mAP50: 0.859, mAP50-95: 0.524) with an accuracy of 87%. By contrast, SSD has the lowest accuracy of 68%, while YOLOv7 and YOLOv5 have lower accuracy values of 82% and 79%. These findings unequivocally show that YOLOv8 is the best model for our dataset and application since it provides better detection performance.

Figure 9 shows results of a YOLO(You Only Look Once version 8)-based real-time weapon identification system are displayed in Figure 10. "ALERT! Weapon Detected" is the warning message that appears when it detects a knife with a confidence score of 0.72.

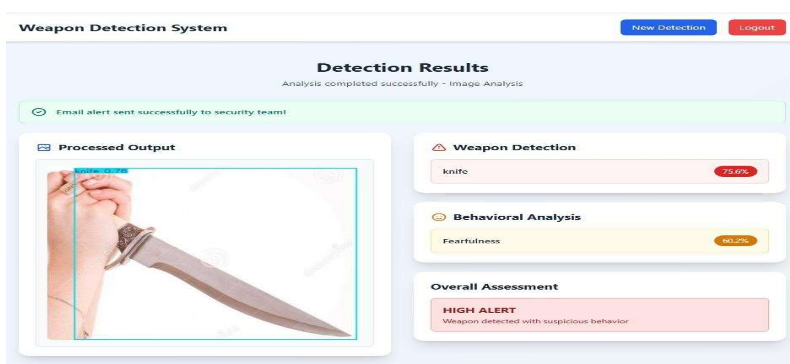

Figure 9. Real-time Weapon Detection Output

VII. CONCLUSION

The model illustrates by comparing to current models such as YOLOv5, YOLOv7, and SSD, the suggested YOLOv8-based weapon detection system achieves higher accuracy and faster detection, making it very effective for real-time surveillance. Even while the technology works effectively, it still has issues with busy areas, dim illumination, and small item visibility. It offers a dependable and scalable solution for public safety applications in spite of these difficulties. Future enhancements will include behavior-based threat identification, improved low-light performance, and edge deployment.

ACKNOWLEDGMENT

The authors would like to thank Mangalore Institute of Technology and Engineering, Mangalore, Karnataka for providing research infrastructure to conduct experiments at MITE Research centre.

REFERENCES

[1] Berardini, D., Migliorelli, L., Galdeli, A., & Marín- Jiménez, M. J. (2025). Edge Artificial Intelligence and Super-Resolution for Enhanced Weapon Detection in Video Surveillance. Engineering Applications of Artificial Intelligence,140, 109684.l https://doi.org/10.1016/j.engappai.2024.109684

[2] Ruiz-Santaquiteria,J.,Velasco-Mata, A., Vallez, N., Bueno, G., Álvarez-García, J. A., & Deniz, O. (2021). Handgun Detection Using Combined Human Pose and Weapon Appearance. IEEE Access, 9, 123815–123827. https://doi.org/10.1109/ACCESS.2021.3110335

[3] Bhatti, M. T., Khan, M. G., Aslam, M., & Fiaz, M. J. (2021). Weapon Detection in Real-Time CCTV Videos Using Deep Learning. IEEE Access, 9, 34366–34376. https://doi.org/10.1007/978-981-16-7389-4_19

[4] Chen, Y., Yuan, X., Wang, J., Wu, R., Li, X., Hou, Q.,& Cheng, M.-M. (2023). YOLO-MS: Rethinking Multi-Scale Representation Learning for Real-Time Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. arXiv:2308.05480. https://doi.org/10.3390/app12125772

[5] Gallagher, J. E., & Oughton, E. J. (2025). Surveying YOLO Multispectral Object Detection: Advancements, Applications, and Challenges. IEEE Access, 13, 1–22. https://doi.org/10.1109/ACCESS.2025.3526458

[6] Abba, S., Bizi, A. M., Lee, J.-A., Bakouri, S., & Crespo, M. L. (2024). Real-Time Object Detection, Tracking, and Monitoring Framework for Security Surveillance Systems. Heliyon, 10,e34922.https://doi.org/10.1051/e3sconf/202339101071

[7] Zhao, Z.-Q., Zheng, P., Xu, S.-T., & Wu, X.(2021). Object Detection with Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems, 32(8), 3212. https://doi.org/10.1016/j.neucom.2021.04.048

[8] Prasad, D. K. (2012). Survey of the Problem of Object Detection in Real Images. International Journal of Image Processing (IJIP), 6(6), 441–456.Huang, Y., Fu, X., & Zeng, Y. (2022).

[9] Anchor-Free Weapon Detection for X-Ray Baggage Security Images. IEEE Access, 10, 32241–32250.https://doi.org/10.1016/j.dsp.2023.103790

[10] Muñoz, J. D., Ruiz-Santaquiteria, J., & Deniz, O. (2025). Concealed Weapon Detection Using Thermal Cameras. Journal of Imaging, 11(1), 42. https://doi.org/10.1016/j.eswa.2022.118698

[11] Gao, X., Liu, H., Roy, S., & Xing, G. (2022). Learning to Detect Open Carry and Concealed Objects with 77 GHz Radar. IEEE Journal of Selected Topics in Signal Processing, 16(3),512–526. https://doi.org/10.1007/s11042-023-16231-x

[12] Wang, G., Ding, H., Duan, M., et al. (2023).Real-Time CCTV Video Analysis: Deep Learning for Weapon Detection. Digital Signal Processing, 132, 103790. https://doi.org/10.1007/s11042-023-16231-x

[13] Yadav, P., Gupta, N., Sharma, P.K. (2023). Automatic Handgun Detection with Deep Learning in Video Surveillance Images. Expert Systems with Applications, Article 118698. https://doi.org/10.1007/s11042-023-16231-x

[14] Gallagher, J.E., & Oughton, E.J. (2022). A Hybrid YOLO and SSD Framework for Efficient Weapon Detection in Public Surveillance IEEE Access. https://doi.org/10.1007/s11042-023-16231-x

[15] Ruiz-Santaquiteria, J., et al. (2023). Deep Learning-Based Gun and Knife Detection for School Safety Applications. IEEE Access, 9, 123815–123827. https://doi.org/10.1007/s11042-023-16231-x

[16] Bhatti, M.T., et al. (2023). Lightweight Deep Neural Networks for Weapon Detection on Mobile Surveillance Devices . IEEE Access, 9, 34366–34376. https://doi.org/10.1007/s11042-023-16231-x

[17] Berardini, D., et al. (2024). Multimodal Fusion for Weapon Detection: Combining Visual, Audio, and Infrared Cues. Engineering Applications of Artificial Intelligence, 140, 109684. https://doi.org/10.1007/s11042-023-16231-x

[18] Kong, L., Wang, J., & Zhao, P. (2022). Vision Transformers for Weapon Detection in Complex Environments. IEEE, Access, 10. https://doi.org/10.1007/s11042-023-16231-x