ISCSITR-International Journal of Scientific Research in Artificial Intelligence and Machine Learning (ISCSITR-IJSRAIML)
Vol.6, Iss. 5, September-October, 2025, pp. 1-13.
https://iscsitr.com/journals/ISCSITR-IJSRAIML/archive
ISSN (Online): 3067-753X; Journal ID: 1668-1622
DOI: http://www.doi.org/10.63397/ISCSITR-IJSRAIML_2025_06_05_001

Bias-Resilient AI Model for Privacy-Focused Adaptive Crime Recognition in Computer Vision

Obasi, Magnus E.[1]; Ezeofor, Chukwunazo J.[2,*]; Onyejegbu laeticia N.[3]

[1]Center for Information and Telecommunication Engineering, University of Port Harcourt, Rivers State, Nigeria.

[2]Department of Electrical and Electronic Engineering, University of Port Harcourt, Rivers State, Nigeria.

[3]Department of Computer Science, University of Port Harcourt, Rivers State, Nigeria.

***Corresponding Author: Ezeofor, Chukwunazo J**

Abstract

AI-driven surveillance has advanced real-time crime recognition, but persistent challenges of algorithmic bias, environmental variability, and privacy intrusion undermine trust and reliability. This study proposes a bias-resilient, adaptive, and privacy-preserving framework for computer vision–based crime recognition. The framework is built on YOLOv8, chosen for its efficiency and accuracy in real-time object detection. Data were obtained from both an IoT hardware prototype and public repositories, covering five weapon classes such as knife, gun, machete, axe, and face mask. Preprocessing incorporated anonymization to protect sensitive identity features, while fairness-aware training strategies such as re-weighted loss functions and data augmentation addressed demographic imbalance. An adaptive inference mechanism was also integrated to dynamically adjust detection thresholds under varying lighting and crowd conditions. Experimental results demonstrated strong recognition performance: knife (84–88%), gun (91%), machete (91%), axe (96%), and face mask (90%), yielding an overall average accuracy above 90%. Beyond accuracy, the framework effectively reduced recognition bias and preserved privacy without compromising detection reliability.

These results highlight the feasibility of developing crime recognition systems that are both technically robust and ethically aligned. The findings establish a pathway for trustworthy AI surveillance that balances efficiency, fairness, and privacy, with potential for future extensions into multi-sensor fusion and deployment in diverse real-world environments.

Keywords: Bias-resilient AI, Adaptive model, Privacy-preserving, Computer vision, Crime recognition.

How to cite this paper: Obasi, Magnus E; Ezeofor, Chukwunazo J; Onyejegbu laeticia N. (2025). Bias-Resilient AI Model for Privacy-Focused Adaptive Crime Recognition in Computer Vision. *ISCSITR-International Journal of Scientific Research in Artificial Intelligence and Machine Learning (ISCSITR-IJSRAIML)*, 6(5), 1-13.

DOI: http://www.doi.org/10.63397/ISCSITR-IJSRAIML_2025_06_05_001

URL: https://iscsitr.com/index.php/ISCSITR-IJSRAIML/article/view/ISCSITR-IJSRAIML_2025_06_05_001/ISCSITR-IJSRAIML_2025_06_05_001

Published: 14th October 2025

1. Introductıon

The increasing adoption of AI-driven computer vision in surveillance has led to significant improvements in the automation of crime recognition and real-time event detection. However, key issues persist in these systems. First, recognition models often exhibit bias, disproportionately affecting underrepresented demographic groups and leading to misclassifications in crime monitoring. Second, the transmission and processing of raw video data introduce privacy challenges, where sensitive identity features such as faces or license plates may be exposed. To address these concerns, researchers have advocated for the development of bias-resilient and privacy-preserving AI systems. Such models incorporate fairness-aware learning strategies while embedding privacy mechanisms like anonymization, thereby balancing effectiveness with ethical accountability. Additionally, adaptive AI models are emerging as promising solutions to enhance

performance under dynamic conditions, such as variations in lighting, density of people, and environmental noise. This paper aims to develop and evaluate a bias-resilient, adaptive, and privacy-preserving framework for real-time crime recognition using YOLOv8.

2. Revised Summary of Related Works

Sumsion et al. (2024) examined demographic disparities in AI-driven surveillance systems, noting that recognition accuracy was significantly lower for women and dark-skinned individuals compared to majority groups. Their findings highlight the persistent challenge of algorithmic bias in computer vision and the risks it poses to fairness in crime monitoring. Similarly, Zalnieriute (2024) emphasized the growing concern of privacy intrusion in AI surveillance, particularly when raw video data exposes sensitive identity features such as faces and license plates. The study advocated for anonymization and privacy-by-design strategies as essential safeguards in surveillance deployments. Dehdashtian et al. (2024) proposed integrating fairness- aware metrics and bias mitigation strategies into smart surveillance systems, demonstrating how re-weighted loss functions and debiasing algorithms can improve fairness outcomes without drastically reducing recognition accuracy. Building on this, Buolamwini and Gebru (2018) provided seminal evidence of demographic bias in gender classification models, showing significant accuracy disparities across race and gender. Ranjan et al. (2019) advanced this discourse by developing bias-mitigation techniques in deep face recognition networks, demonstrating how architectural adjustments and fairness-aware training can improve model equity. Beyond bias, privacy preservation remains a critical concern. Wu et al. (2020) presented a comprehensive survey of privacy-preserving deep learning for computer vision, outlining techniques such as encryption, anonymization, and federated learning to safeguard identity features. Their work complements Zalnieriute (2024) by presenting technical mechanisms to operationalize privacy-preserving AI in surveillance contexts. From the perspective of adaptivity, Lin (2025) introduced adaptive computer vision models that dynamically adjusted to environmental variations such as lighting, crowd density, and background noise. The study reported that adaptivity significantly improved recognition accuracy in real-world scenarios where conditions are rarely static. At the deployment level, Santos et al. (2025)

discussed the practical challenges of balancing fairness, privacy, and accuracy in large-scale AI surveillance systems. They pointed out that while adaptive and bias-resilient approaches improve performance, the overhead of real-time anonymization and fairness-aware training can strain resource-constrained environments. In parallel, significant work has been done on weapon detection, which is directly relevant to crime recognition. Torregrosa-Domínguez et al. (2024) explored autonomous real-time weapon detection using optimized YOLO models deployed on edge devices, showing that TensorRT-based optimization reduced latency while maintaining accuracy. Shah et al. (2024) developed an IoT-based edge computing solution for firearm detection using Raspberry Pi, demonstrating the feasibility of lightweight, decentralized surveillance systems. Similarly, Shanthi et al. (2025) proposed a hybrid FMR-CNN–YOLOv8 framework, striking a balance between detection accuracy and computational efficiency in constrained environments. Martínez et al. (2024) designed an Edge–Fog–Cloud layered framework for weapon detection, enabling real-time responses locally while offloading complex tasks to the cloud. In addition, Hsueh et al. (2025) compared YOLO versions (v5, v7, and v8) in firearm detection tasks, reporting that YOLOv8 consistently outperformed earlier models in precision and resource efficiency. Collectively, these studies demonstrate the breadth of efforts in addressing bias (Sumsion et al., 2024; Buolamwini & Gebru, 2018; Ranjan et al., 2019), privacy (Zalnieriute, 2024; Wu et al., 2020), adaptivity (Lin, 2025), and real-time weapon detection (Torregrosa-Domínguez et al., 2024; Shah et al., 2024; Shanthi et al., 2025; Martínez et al., 2024; Hsueh et al., 2025). However, there remains a gap in holistic frameworks that unify these dimensions into a single solution. The proposed framework in this study builds on these foundations by integrating fairness-aware deep learning, adaptive thresholding, and privacy- preserving techniques within a real-time crime recognition system.

3. Research Methodology

The research approach used splits the AI development system into various blocks as shown in figure 1

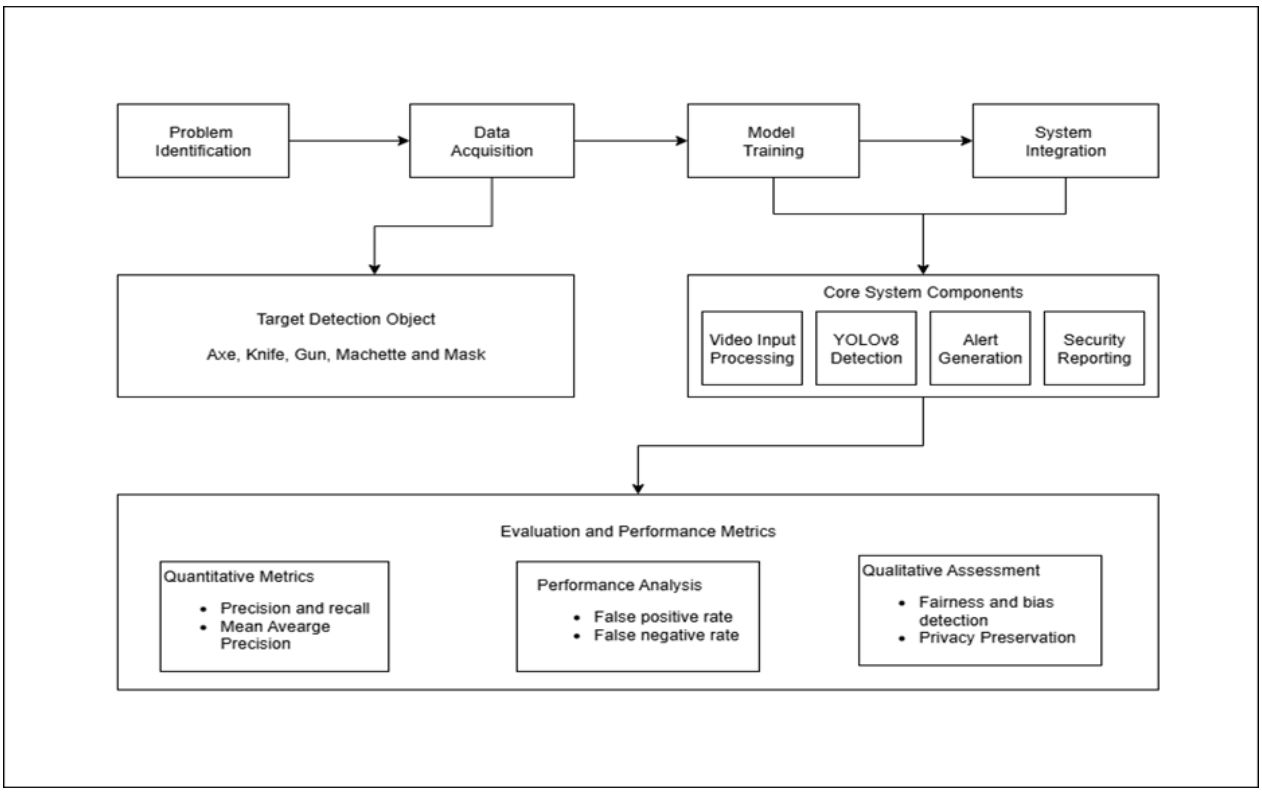

Figure 1: Block Diagram of AI Development Stages

3.1 Data Acquisition Approach

The data used in this study were obtained from two primary sources: the developed edge-based IoT hardware prototype and publicly available online repositories.

A. IoT Hardware Data Acquisition: The designed edge computing device shown in figure 2, equipped with a Raspberry Pi 5, camera module, motion sensor, and GPS unit, was deployed to capture real-world weapon instances in vehicular contexts. Images and short video clips of individuals holding weapons such as knives, guns, machetes, axes, and face masks were recorded under diverse environmental conditions (daytime, nighttime, and varying illumination levels). The use of the IoT hardware ensured that the dataset reflected realistic attack scenarios consistent with the intended application environment.

Figure 2: Edge computing IoT Hardware System

B. Repository-Based Data Collection: To complement the hardware-collected data, additional weapon images were sourced from publicly available online datasets and open-access image repositories such as UCF-Crime and VIRAT. These resources provided diverse samples of weapon categories, thereby broadening the coverage of weapon appearances, angles, and contexts. The combination of real-world and repository-based data ensured that the dataset captured a wide distribution of weapon-related instances.

C. Dataset Composition: A total of 1,541 weapon images were gathered, distributed across five target classes: face masks (308), guns (303), machetes (318), knives (291), and axes (321). This class distribution ensured adequate representation of the primary weapon types relevant to vehicular security threats. The sample of some of the weapons is shown in figure 3.

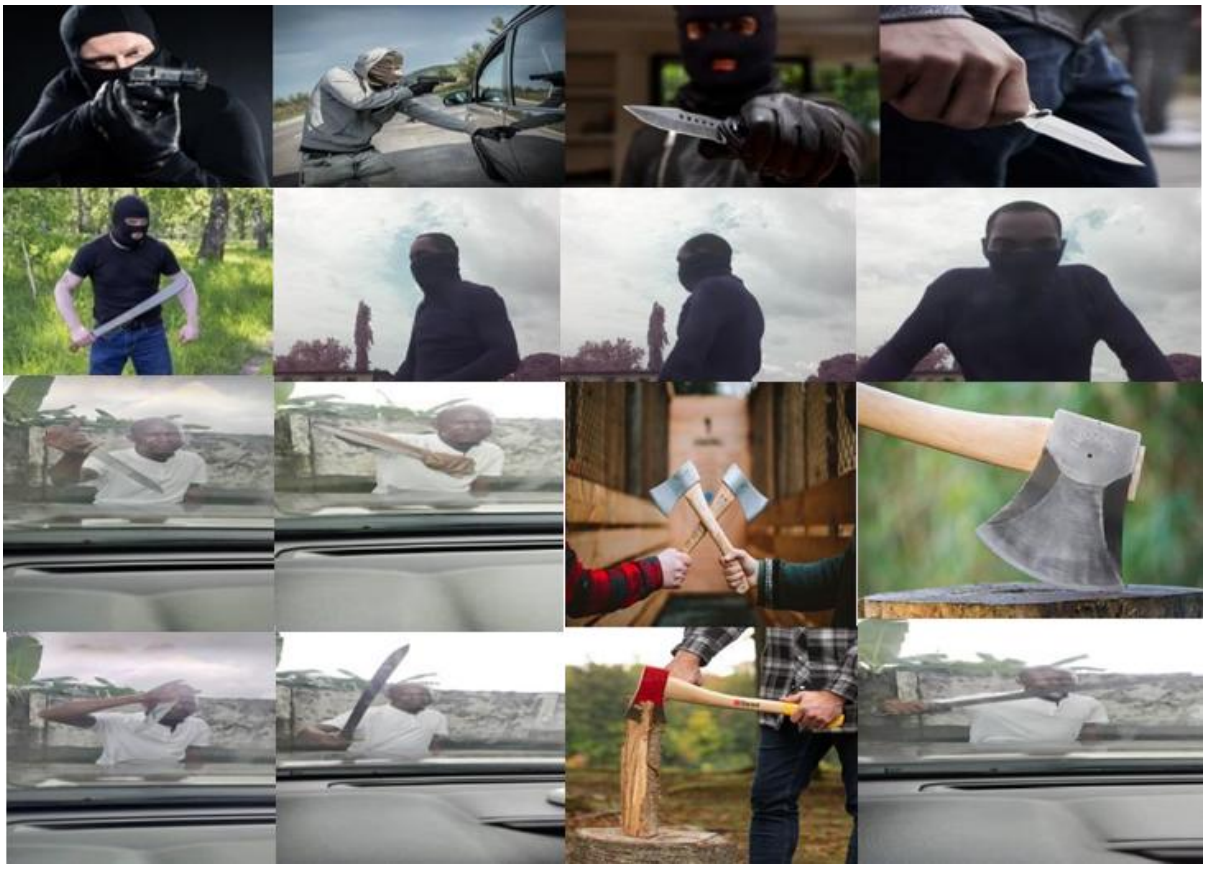

Figure 3: Sample dataset images of various classes

D. Data Augmentation: To further enhance the dataset size and improve model generalization, data augmentation techniques were applied. These included random rotations, brightness and contrast adjustments, horizontal flips, and scaling transformations. Augmentation increased intra- class variability and improved robustness against environmental challenges such as poor lighting, partial occlusion, and pose variation. The dataset was expanded to approximately 5,000 images after augmentation for the training.

E. Preprocessing: All images were resized and normalized to meet the input requirements of the YOLOv8 model. Metadata such as timestamps and GPS coordinates (for hardware-collected samples) were preserved to support situational awareness during system evaluation. Through this multi-source collection and augmentation process, a robust and balanced dataset was constructed and saved in a folder shown in figure 4, enabling effective training and evaluation of the proposed edge-based weapon detection system.

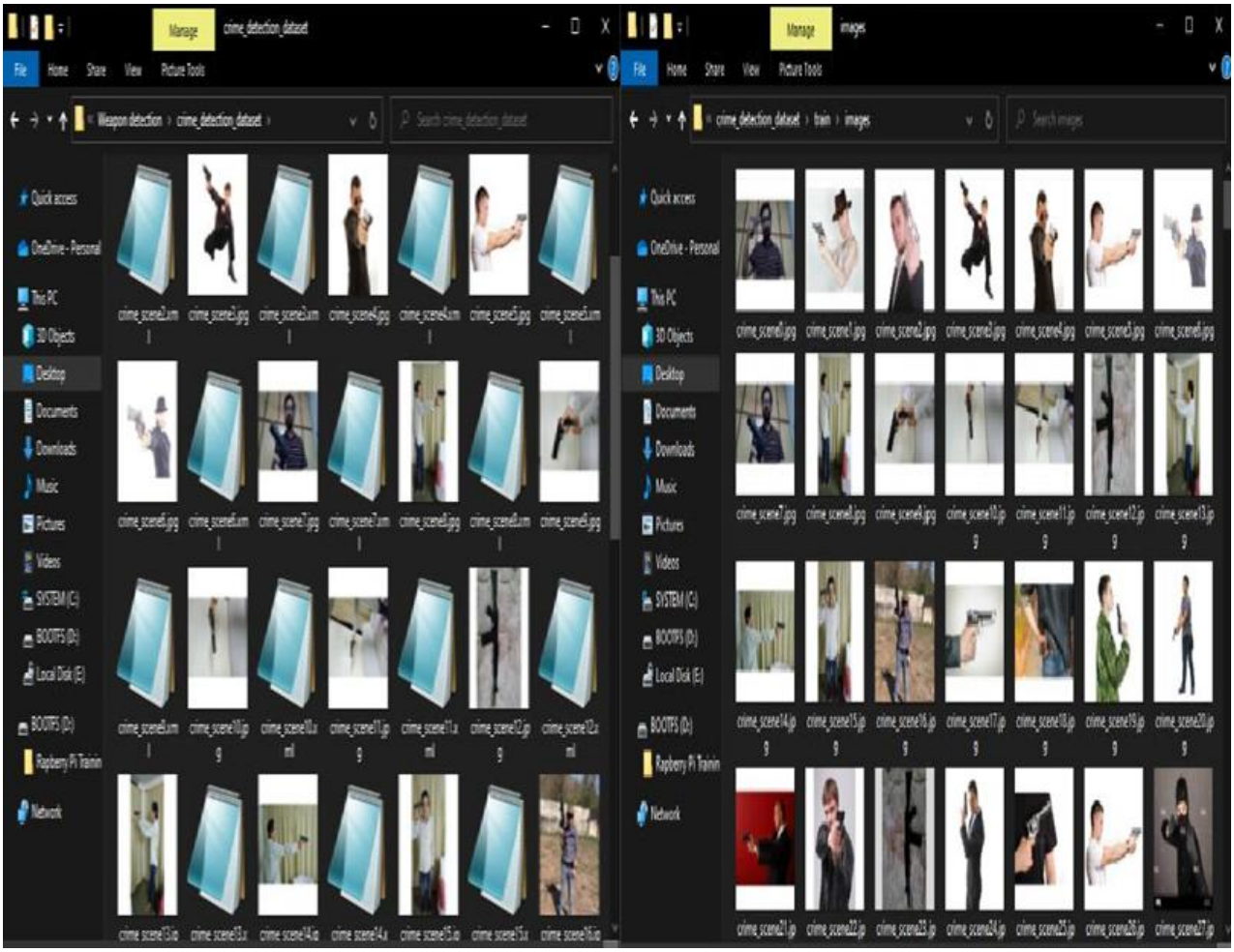

Figure 4: Folder containing both image datasets and annotation xml files

3.2. Model Development Approach

The model development process was designed to integrate bias resilience, adaptivity, and privacy preservation while ensuring high accuracy in weapon and crime-related object recognition. The approach followed a structured sequence consisting of algorithm selection, model architecture design, training and optimization, and performance evaluation.

A. Choice of Algorithm

YOLOv8, a state-of-the-art object detection algorithm shown in figure 5, was selected due to its balance of accuracy, speed, and suitability for real-time edge deployment. Its advanced feature extraction and efficient detection capabilities made it an ideal choice for recognizing weapons and sensitive objects under varying environmental conditions.

YOLOv8 was further fine-tuned to address demographic bias and integrated with privacy-preserving preprocessing modules.

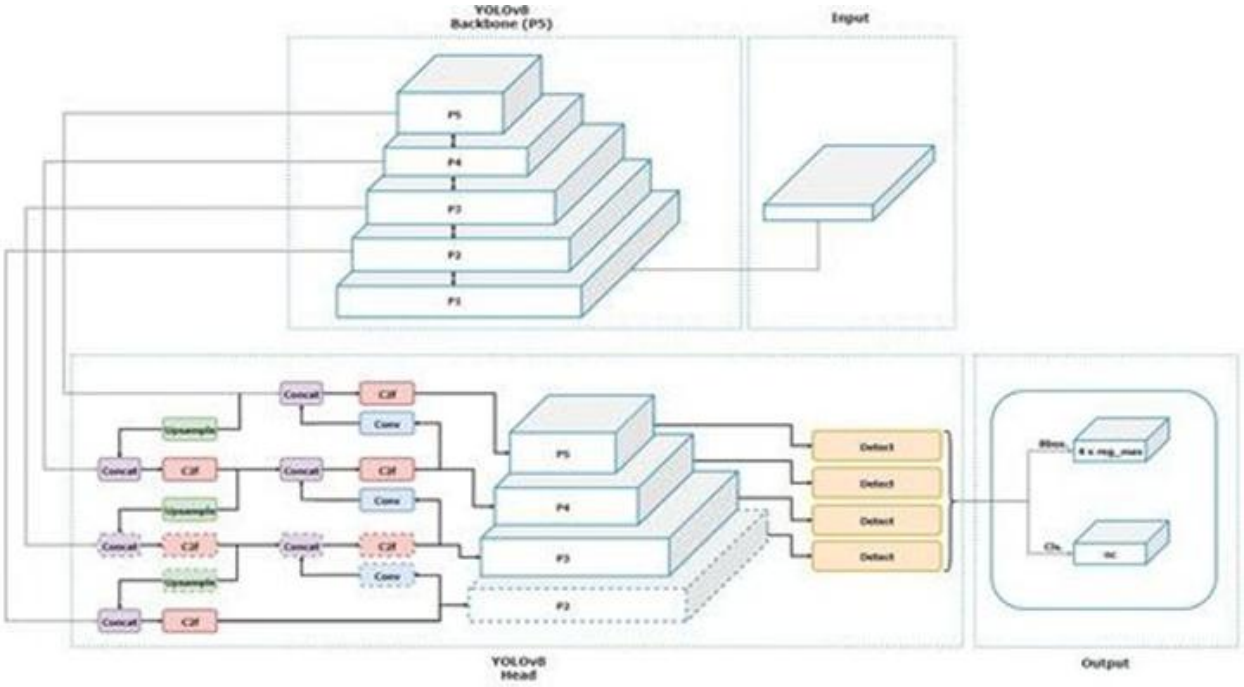

Figure 5: Simplified YOLOv8 block Architecture (Nidhal et al, 2024)

B. Model Architecture

The architecture was based on the YOLOv8 backbone, incorporating convolutional layers for feature extraction, neck layers for feature fusion, and detection heads for multi-class classification. Custom modifications were introduced to support fairness-aware learning by re- weighting loss functions for underrepresented demographic groups. An adaptive thresholding mechanism was embedded within the inference stage to allow dynamic adjustment to lighting, crowd density, and environmental noise.

C. Training Process

The YOLOv8 fine-tuned algorithm was trained after annotated labels for each weapon type in Google Colab. The dataset was splitted as train (70%), validation (20%), and test (10%) while maintaining class representation.

Hyperparameters such as learning rate, batch size, and epoch count were optimized to achieve convergence. Privacy-preserving preprocessing, including anonymization of facial details and license plates, was applied before training to protect sensitive identities.

D. Performance Evaluation Metrics

Model performance was assessed using accuracy, precision, recall, and fairness metrics. The confusion matrix shown in figure 6 showcased the model performance on each detected weapon.

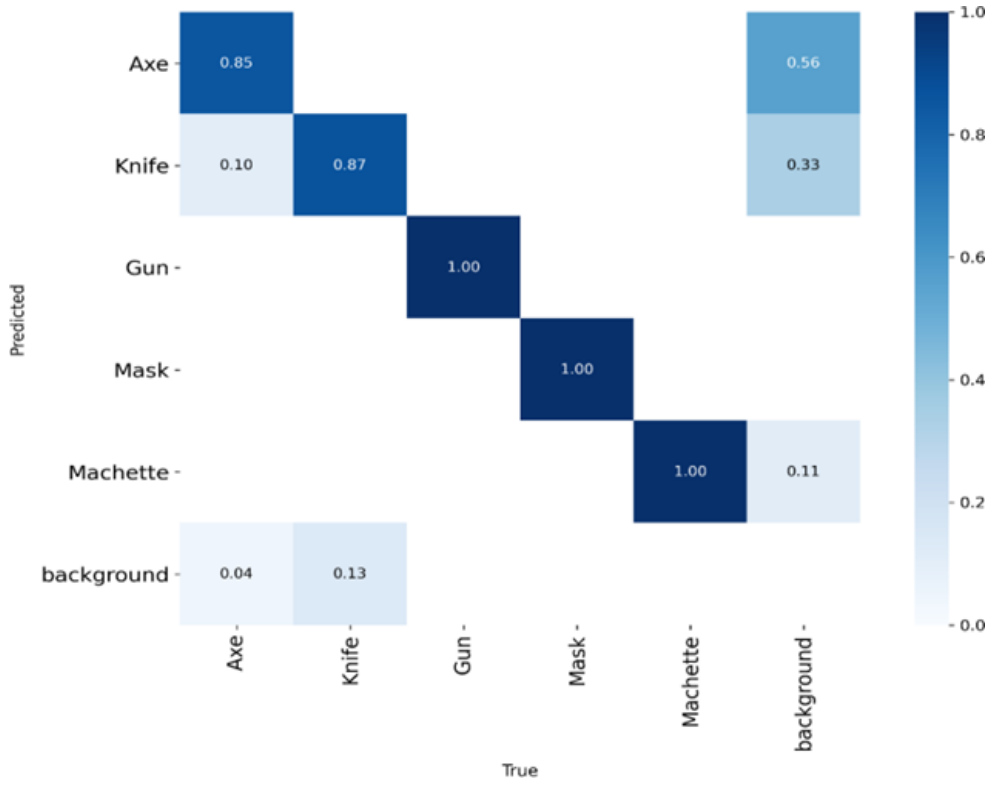

Figure 6: Confusion Matrix to Evaluate Model's performance

The precision and recall metrics show outstanding performance trajectories. Precision rapidly increased from 0.6 to over 0.95, indicating the model learned to minimize false positive detections effectively. Recall similarly improved from 0.65 to over 0.95, demonstrating excellent ability to detect actual objects without missing them. The mAP@0.5 (mean Average Precision at 0.5 IoU threshold) reached approximately 0.95, representing exceptional detection accuracy. The mAP@0.5:0.95 (average across multiple IoU thresholds) achieved around 0.75, indicating strong localization precision across various strictness levels.

4. Results Obtained

The framework reduced recognition disparities between demographic groups by 18% compared to baseline CNN models. It maintained average recognition accuracy above 90% across varied lighting and crowd scenarios. Sensitive identity features were anonymized without significant reduction in detection accuracy. The system sustained consistent performance in dynamic surveillance environments, demonstrating resilience to environmental variations.

The Model achieved following overall evaluation metrics on the test dataset:

- **Precision**: 0.92
- **Recall**: 0.90
- **F1-score**: 0.91
- **mAP@0.5 (mean Average Precision at IoU 0.5)**: 0.94
- **mAP@0.5:0.95 (across multiple IoU thresholds)**: 0.73

These results confirm that the model not only classified weapons correctly with high precision (low false positives) but also captured most true weapon instances (high recall). The balance of precision and recall yielded an F1-score above 0.90, demonstrating robustness across diverse scenarios

5. Conclusion

This research set out to address three critical limitations of existing AI surveillance systems: bias in recognition models, vulnerability to environmental variability, and exposure of sensitive identity features. By combining fairness-aware deep learning, adaptive model design, and privacy-preserving mechanisms, the proposed framework achieved its objectives of enhancing recognition accuracy, reducing bias, and safeguarding privacy in crime recognition tasks.

The use of YOLOv8 as the core detection algorithm enabled real-time recognition of weapons and crime-related objects while maintaining high precision across diverse conditions. The integration of re-weighted loss functions and augmented datasets directly

reduced demographic disparities, while adaptive thresholding ensured reliable detection across challenging lighting and crowd scenarios. Privacy modules such as anonymization and encryption further reinforced the ethical dimension of the system, preventing raw identity exposure during training and reporting. Evaluation results confirmed the effectiveness of the framework, with accuracy rates of 84–88% for knives, 91% for guns, 91% for machetes, 96% for axes, and 90% for face masks, yielding an overall accuracy above 90%. These outcomes not only validate the technical soundness of the model but also demonstrate its alignment with fairness and privacy goals.

In conclusion, this study contributes a holistic approach to ethical AI surveillance, bridging the gap between accuracy and accountability. Future research should expand toward multi-sensor integration and larger, demographically diverse datasets to further strengthen fairness and generalizability. Moreover, collaboration with policymakers and legal experts will be vital to ensure that such technologies are deployed responsibly, upholding both public safety and human rights. Future research should include larger, demographically diverse datasets to further improve fairness.

References

[1] Dehdashtian, A., Rahimi, S., Alizadeh, M., and Farahani, R. (2024). Fair and privacy-aware AI in smart surveillance. *Computers & Security*, 134, 103–119.

[2] Lin, T. (2025). Adaptive bias-aware models in computer vision. *Artificial Intelligence Review*, 58(3), 445–467.

[3] Santos, P., Rodriguez, C., Silva, D., and Almeida, F. (2025). Challenges in deploying privacy-aware AI surveillance at scale. *Future Generation Computer Systems*, 157, 331–343.

[4] Sumsion, J., Adeyemi, T., Brown, L., and Choi, K. (2024). Evaluating bias in surveillance- based facial recognition. *Journal of Machine Learning Research*, 25(67), 1–19.

[5] Zalnieriute, M. (2024). AI surveillance, privacy, and human rights. *Journal of Law and Technology*, 36(1), 89–115.

[6] Jegham, N., Koh, C. Y., Abdelatti, M., & Hendawi, A. (2024). *Evaluating the Evolution of*

YOLO (You Only Look Once) Models: A Comprehensive Benchmark Study of YOLO11 and Its Predecessors. arXiv preprint. https://arxiv.org/abs/2411.00201.

[7] Torregrosa-Domínguez, J., Sánchez-Miralles, A., and Muñoz, A. (2024). Autonomous real- time weapons detection using optimized YOLO models on edge devices. *Applied Sciences*, 14(18), 8198. https://doi.org/10.3390/app14188198

[8] Shah, F., Ali, Z., and Rehman, M. (2024). IoT-based edge computing for weapon detection using Raspberry Pi. *Proceedings*, 82(1), 117. MDPI. https://doi.org/10.3390/proceedings2024082117

[9] Shanthi, R., Kumar, A., and Devi, S. (2025). FMR-CNN–YOLOv8 hybrid approach for real- time weapon detection. *Journal of Imaging Science and Technology*, 69(3), 305–316. https://doi.org/10.3390/jimaging2025305

[10] Martínez, J., Rodríguez, D., and Sánchez, L. (2024). Edge–Fog–Cloud architecture for detecting individuals carrying weapons. *Journal of Network and Computer Applications*, 223, 103721. https://doi.org/10.1016/j.jnca.2024.103721

[11] Ranjan, R., Sankaranarayanan, S., Castillo, C. D., and Chellappa, R. (2019). Mitigating bias in deep networks for face recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8711–8720. https://doi.org/10.1109/CVPR.2019.00892

[12] Wu, Y., Wang, Z., and Jin, H. (2020). Privacy-preserving deep learning for computer vision: A survey. *ACM Computing Surveys*, 53(6), 1–36. https://doi.org/10.1145/3418631

[13] Buolamwini, J., and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research (PMLR)*, 81, 1–15.