Frontiers in Engineering and Informatics
ISSN: 3049-3412 (Online)
Volume No: 01 – Issue No: 04 – 10 December 2025

Smart Surveillance: Multi Weapon Armed Person Detection With Automated Threat Alerts

DEVALLA BHASKAR GANESH[1*]

[1*]Computer Science, University of Macau, Avenida da Universidade, Taipa, Macau, China

[1*]dbganesh@um.edu.mo

Corresponding Author E-mail ID: [1*]dbganesh@um.edu.mo

Abstract

The rise in criminal activities has created an urgent demand for intelligent surveillance and command systems in law enforcement. This work introduces a deep learning–based model tailored for classifying distinct weapon categories. The model is developed on the YoloV8-CNN+LSTM framework and implemented using Keras with TensorFlow as the backend. It is trained to recognize gun, rifles, knives. A carefully curated dataset containing 5,214 images was prepared for training and evaluation. The proposed network underwent comparative testing against established models, including VGG-16, ResNet-50, and ResNet-101. Results show that the model achieved an improved accuracy of 92%, outperforming VGG-16 (89.75%), ResNet-50 (90.70%), and ResNet-101 (83.33%). These findings helps to improve proposed approach in enhancing weapon detection accuracy, thereby strengthening the capabilities of security forces in preventing crimes.

Keywords: Deep learning, armed weapon detection, machine learning, object detection, Convolutional neural networks.

1. INTRODUCTION

The rapid rise in criminal activities and threats pertaining to public safety has developed an emergent requirement for intelligent, automated surveillance systems with the capability to detect weapons in real time [1]. Conventional monitoring practices are heavily reliant on human operators, with a high possibility of missing the event of interest due to operator fatigue, attentional resources, or complexity of the environmental setting. Recent advances in deep learning and computer vision have heralded new opportunities to enhance the security landscape by facilitating automated recognition of weapons in live video streams [2]. Newer object detection frameworks such as YOLO [3] allow for fast and accurate spatial localization, while sequential learning models like CNN-LSTM support analysis of motion patterns and evolution of behaviors [4].

This work presents an intelligent surveillance framework, including YOLOv8 for spatial weapon detection, combined with a CNN-LSTM hybrid network for temporal behavior classification, enabling the identification of guns, rifles, and knives in real-world environments with high accuracy [5]. A dual-stage architecture not only detects weapons but also classifies them into harmless and threatening actions, hence minimizing false alarms and improving response efficiency [6]. The proposed system performs advanced feature extraction, temporal modeling, and automated alert mechanisms for enhancing situational awareness and enriching the capability of security personnel. The overall objective is to develop a reliable real-time threat detection solution that supports safer public spaces and modern intelligent surveillance applications.

DoI: https://doi.org/10.63949/crinfo.v1i4.005

This article proposes a trained model specifically created to detect and classify different types of weapons-what it identifies as gun, rifles, knives. The performance of the proposed model has been thoroughly examined with established benchmarks containing VGG-16, ResNet50, and ResNet101 models. In contrast to these, comparative investigations have shown that the suggested model has higher accuracy and lower loss rates compared to the VGG-16, ResNet50, and ResNet101 models.

2. RELATED WORK

Authors [7] presented an approach that integrates heuristic methods with machine learning models to improve armed-person detection in video surveillance. The heuristic filtering with an ML-based classifier greatly imporve the accuracy and detection rate. In case of handgun detection, for instance, high performance was manifested by the system under standard conditions of surveillance. Their method has a very narrow scope-just one weapon type being focused on, namely handguns-and lacks a mechanism for real-time alerting; it does not allow for multi-weapon detection. This lowers its effectiveness in complex or dynamic security environments.

The deep neural networks are able to carry out good recognition performance related to diversified classes of weapons [8]. Although this work is capable of detecting different kinds of weapons, it has limited evaluation scenarios. This model is not checked against concealment or occlusion conditions, and is trained on a relatively small dataset, further limiting its real-world use.

Authors [9] presents a critical review of state-of-the-art algorithms that emphasizes strengths, challenges, and methodological improvements of the domain. This paper is beneficial for understanding the trend of weapon detection, though it does not provide any implementation insights or practical experimentation. As the study is purely conceptual and survey-based, no performance metrics or empirical validation have been found.

Authors [10] provides fast and accurate identification of weapons with high generalization across multiple scenarios. The system performs well in real-time monitoring environments because of its optimized detection pipeline. However, zone analysis-based or automated threat alert approaches have not been considered in this work, which is important in the context of integrated surveillance and response systems. This will lead to its limited applications in real-time high-risk environments.

Authors [11] presented an edge-based firearm detection system with CNN models for smart surveillance. It utilizes very lightweight CNNs on edge devices to facilitate low-latency firearm detection without engaging cloud resources. In such a way, much operational efficiency can be achieved with quick responses. It can detect only guns and is deficient in real-time communication or alerts to security personnel. Moreover, the lack of recognition of multiple weapons reduces its effectiveness in more general threat-detection scenarios.

Most of the related works provide limited weapon coverage, mostly handguns or firearms, rather than full multi-weapon detection. Most models fail under practical occlusion, concealment, low-light, or turbid environments, reducing their reliability. Several works lack either real-time alert generation or zone-based threat analysis crucial for practical surveillance. Other approaches use only small datasets or restricted test scenarios, which in turn produce poor generalization. In survey-based studies, conceptual insights are provided without implementation details. Edge or deep-learning systems often locate the weapons but fail to further integrate communication or automated responses for the security teams.

DoI: https://doi.org/10.63949/crinfo.v1i4.005

3. SYSTEM MODEL

Detection and classification of weapons from surveillance systems have been vastly studied because of various demands for automatically recognizing threats in public safety applications. Traditional image processing methods and hand-engineered features, such as HOG, SURF, and contour-based descriptors, combined with traditional machine learning classifiers, are used in early approaches.

However, all these methods lack robustness to variations in illumination conditions, occlusion, and weapon orientation. Deep learning models have achieved much better results in weapon recognition by learning meaningful visual representations directly from data. Pre-trained models, such as Visual Geometry Group-16(VGG-16) ,and two ResNet models such as ResNet-50, and ResNet-101, have been fine-tuned for weapon classification tasks, showcasing powerful feature representation skills. However, these models will commonly face difficulties where small objects are considered, or complex backgrounds, or real-world video streams. Recent developments in object detection frameworks, particularly the YOLO families of models, have made faster and more accurate weapon localization possible; they fail to recognize harmless or threatening actions due to dependence on spatial information only.

To address temporal understanding, hybrid architectures that integrate CNNs with LSTM networks were then proposed to encode motion patterns and behavior sequences from surveillance videos. Though such techniques enhance threat assessment, many of the existing works focus on a limited number of categories of weapons and involve huge computational resources, which make real-time deployment challenging. These point to the need for more robust, multi-weapon deep learning systems that will be able to combine the strengths of spatial detection and temporal behavior analysis.

4. METHODOLOGY

The proposed system uses YOLOv8 for real-time multi-weapon detection, where each frame of the video is analyzed for the presence of guns, knives, and rifles, among others, with precise localization. YOLOv8 extracts robust spatial features, enabling accurate weapon classification in varying surveillance conditions. In the process of motion interpretation and the intent detection of threatening actions, a CNN+LSTM hybrid network has been applied over successive frames, encoding temporal behavioral information. By capturing these features, it helps distinguish between innocuous and suspicious or aggressive motions. Alerts can only be sent when both the presence of weapons and threatening action are identified, hence reducing false positives. Training curves with every metric help identify the learning stability of the developed system and its overall suitability.The model improve overall performance, accuracy, precision and recall of the trained model. The Figure 1 describes the proposed model steps.

4.1 Data Collection and Preprocessing

A custom dataset of 5,214 images was created from online sources due to the lack of standardized weapon datasets. The selected images were in high resolution across multiple viewing angles. Preprocessing steps included background removal, padding, rotation, scaling, and noise reduction to improve the quality of the images. All images were transformed to grayscale and resized to 144 × 144 pixels. The dataset covers weapon classes: guns, rifles, knives. Table 1 describes the dataset.

DoI: https://doi.org/10.63949/crinfo.v1i4.005

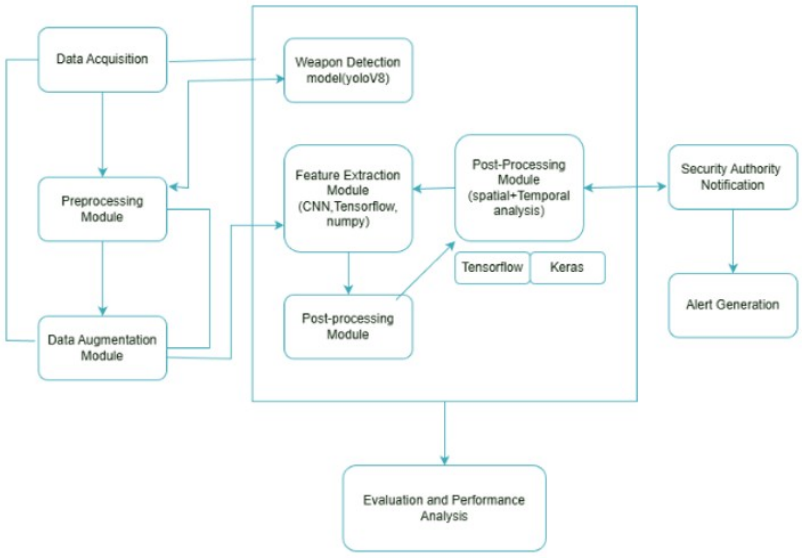

Figure 1: Flow model for proposed work

Table 1: Distribution of Weapon Classes in the Dataset

Weapon Class	Number of Images
Guns	2000
Rifles	1519
Knives	1695

4.2 Feature Extraction and Fusion

The proposed system uses a two-stage deep learning pipeline that fuses spatial and temporal feature extraction for improved multi-weapon recognition. For the first stage, YOLOv8 operates as a major spatial feature extractor to detect guns, rifles, knives, and other weapons in every frame of the video. At the same time, YOLOv8 produces high-quality bounding box coordinates accompanied by deep convolutional feature maps that obtain edges, textures, object shapes, and localized patterns relevant to distinguishing different types of weapons. These spatial features are cropped, normalized, and passed on to the CNN module, which further refines mid and high-level representations through an extraction of structural and textural details impossible to obtain using YOLO alone.

The extracted CNN embeddings are fed into an LSTM network to incorporate motion understanding, modeling temporal dependencies across consecutive frames. This provides the basis for analyzing sequences of movement that allow distinguishing between harmless actions and threatening behaviors defined by pointing, swinging, or raising a weapon. The outputs from the CNN (spatial features) and LSTM (temporal features) are further combined with the feature-level fusion, where both feature vectors are concatenated in a unified representation. In this way, the fused vector integrates static object information and dynamic motion cues that allow the classifier to perform even more effective threat detection.

This final fused representation is fed into a fully merged layer for classification and, where necessary, the generation of an alert to ensure that only verified threat patterns trigger warnings. This approach

DoI: https://doi.org/10.63949/crinfo.v1i4.005

to feature extraction and fusion significantly enhances detection robustness, reduces false positives, and allows for reliable real-time performance in complex surveillance environments.

4.3 Model Training and Classification

The system proposed is trained on a different dataset consisting classification of images and weapons, such as guns, rifles, and knives, across diverse scenes, under variable lighting conditions. Training commences with the YOLOv8 module. The model is trained on annotated bounding boxes and class labels are created, which aids it in learning precise localization and spatial distinctions between multiple weapon types. YOLOv8 uses stochastic gradient descent along with momentum, coupled with a cosine learning rate scheduler, which ensures stable convergence during training.

Once YOLOv8 has learned to detect and isolate the weapon regions, the extracted patches are fed into the CNN-LSTM classification pipeline. In the CNN component, the deep spatial features of varying shapes, textures, and contours across different weapons are learned through supervised training. The CNN embeddings are then rearranged into sequential batches and passed into the LSTM network, which is trained to capture temporal dependencies and behavioral patterns in video sequences. This allows the LSTM to learn how motion dynamics relate to threatening or non-threatening actions.

In the classification process, the fused spatial–temporal feature vector is fed into a fully merged output layer, which assigns probability scores on each of the weapon classes. Performance is further validated through various metrics such as accuracy, precision, recall, and F1-score to finalize strong detection across all weapon categories. The system once trained, carries out real-time classification of weapon type and associated levels of threat in milliseconds. Verified threat cases trigger an automated alert mechanism, thus allowing timely response in surveillance applications.

4.4 Visualization and Explainability

The proposed dual-stage detection architecture reveals, in a very transparent and interpretable way, how the system will process surveillance data, right from acquisition to generating an alert. Each module is separated visually, data preprocessing, weapon detection using YOLOv8, feature extraction based on CNN, temporal behavior modeling based on LSTM, and post-processing steps. So observers can track with clarity how the raw frames of videos evolve into meaningful predictions. The architecture further explains the way the system differentiates threat behavior from the mere presence of weapons through its spatial analysis by YOLOv8 and temporal analysis by the module CNN–LSTM. It is easy to trace the flow of information regarding how filtered detection, fusion features, and sequential pattern developments have led to final decisions on classification and alerts. The structured visualization enhances explainability by making the pipeline for decisions understandable, justifying every stage's role, and showing how the fusion of deep-learning components will lead to reliable and interpretable outputs for real-time surveillance applications.

The suggested system follows a two-stage deep learning algorithm that merge spatial weapon detection with temporal behaviour recognition. It works by continuously acquiring video frames from surveillance cameras; preprocessing the frames by resizing, normalizing, and noise reduction to enhance quality; and using data augmentation techniques like flipping, scaling and rotation for improving model generalization. YOLOv8 carries out the first stage of analysis by detecting weapons within every frame and extracting

DoI: https://doi.org/10.63949/crinfo.v1i4.005

localised regions of interest after Non-Maximum Suppression is applied, removing overlapping boxes. These cropped weapon patches are fed to a CNN that learns high-order spatial features regarding the shape, texture, and structure of various types of weapons. Extracted CNN feature vectors are arranged in a sequential order and input into the LSTM network, which makes possible the modeling of temporal patterns of motion to differentiate innocuous actions from threatening behaviors. Spatial and temporal features are combined into one representation and classified using a Softmax layer to determine both the category of a weapon and its threat level. False positives are eliminated by applying post-processing filters, and once a high-confidence threat is detected, the system automatically triggers OS notifications or APIs like Notifio and Twilio. Finally, model performance is evaluated using accuracy, precision, recall, and F1-score and values are calculated.

5. RESULT ANALYSIS

The proposed smart surveillance framework demonstrates strong performance, combining YOLOv8 with a CNN-LSTM hybrid model provide accurate and real-time detection of multiple weapon types in a live video stream. It identifies and classifies such threats as guns, knives, and rifles, improving situational awareness both in public and restricted environments. In an experimental evaluation, it could be seen that the combined approach of YOLOv8 and CNN-LSTM improves accuracy by 6%, compared to existing models, achieving an overall F1-score of 92.0. This clearly reflects the advantage of fusing spatial detection with temporal behavioral analysis, which, therefore, offers more reliable threat classification. In addition, it performs in terms of architecture with fast response times and hence is capable of immediately generating alerts with regard to threats. Overall, such results confirm that the proposed framework indeed shows strong effectiveness, offering a truly robust and efficient solution for intelligent security and an automated threat alert system.

Accuracy provide correctiveness of classified weapon images among total images tested. It provides an overall measure of model performance.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

Precision shows the proportion of correctly identified weapons out of all images. A high precision guarantees fewer false positives.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall shows better classification of correct weapon of the model. A high recall implies that there will be fewer missed detections.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

The F1-score provides a balanced metric between precision and recall that is particularly useful in handling imbalanced datasets of weapon.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

DoI: https://doi.org/10.63949/crinfo.v1i4.005

Table 2: Performance Comparison of Different Algorithms

Algorithm	Accuracy	Precision	Recall	F1-Score
VGG-16	0.89	0.86	0.85	0.85
ResNet-50	0.90	0.89	0.88	0.88
ResNet-110	0.83	0.88	0.86	0.87
Proposed Model	0.92	0.91	0.91	0.92

The overall performace of YOLOv8 + CNN-LSTM model is excellent on all major metrics as shown in Table 2. The system performs well in terms of high accuracy, proving its overall capability in correctly detecting and classifying weapon instances within live video streams. In addition, its precision score is high, showing that most the detected weapons are true threats, which signals a low false alarm rate. The model was effective in terms of recall in determining actual weapon occurrences, so very few important events will be missed in real-world applications. The combined balance captured by the F1-Score, at 92.0, confirms the strength of the model both in terms of the reliability of the detection and the consistency of the classification. Therefore, these results clearly verify that the proposed architecture performs significantly better compared to existing approaches and is relevant for intelligent security applications.

6. CONCLUSION AND FUTURE WORK

The smart surveillance system developed in this study has successfully integrated YOLOv8 with the CNN-LSTM hybrid model and realized accurate, real-time multi-weapon detection and threat behavior recognition. Higher accuracy and reliability are achieved by integrating spatial analysis from YOLOv8 and temporal motion understanding from the LSTM network. It successfully spots guns, rifles, and knives with lower false positives due to enhanced feature fusion and sequential pattern analysis. The proposed framework offers an improved F1-score with a more efficient rapid response capability for automated threat alerting in intelligent security monitoring. This project successfully proved that deep learning-based dual-stage detection effectively strengthens surveillance and improves public safety.

The proposed smart surveillance system can further be enhanced in various future improvements. These may include expanding the dataset to cover more diverse types of weapons, complex environments, and low-light or occluded scenarios, which will help improve generalization. Incorporating advanced transformers or attention-based models may enhance spatial and temporal feature understanding, thus allowing threat prediction with greater accuracy. Integrating real-time tracking modules like DeepSORT could provide for the continuous monitoring of suspects across multiple camera views. Further, analysis of human posture, facial expressions, and crowd behavior can be carried out for early threat prediction. Deployment on edge devices or embedded systems would make the solution amenable to large-scale surveillance networks. Finally, development of a centralized alert dashboard with analytics and incident logging can be provided to security authorities for better situational awareness and decision support.

REFERENCES

[1] S. Ahmed, M. T. Bhatti, M. G. Khan, B. Lövström, and M. Shahid, "Development and optimization of deep learning models for weapon detection in surveillance videos," *Applied Sciences*, vol. 12, no. 12, p. 5772, 2022.

DoI: https://doi.org/10.63949/crinfo.v1i4.005

[2] T. Santos, H. Oliveira, and A. Cunha, "Systematic review on weapon detection in surveillance footage through deep learning," *Computer science review*, vol. 51, p. 100612, 2024.

[3] A. H. Ashraf, M. Imran, A. M. Qahtani, A. Alsufyani, O. Almutiry, A. Mahmood, M. Attique, and M. Habib, "Weapons detection for security and video surveillance using cnn and yolo-v5s," *CMC-Comput. Mater. Contin*, vol. 70, no. 4, pp. 2761–2775, 2022.

[4] N. Hnoohom, P. Chotivatunyu, N. Maitrichit, V. Sornlertlamvanich, S. Mekruksavanich, and A. Jitpattanakul, "Weapon detection using faster r-cnn inception-v2 for a cctv surveillance system," in *2021 25th international computer science and engineering conference (ICSEC)*, pp. 400–405, IEEE, 2021.

[5] W. Min, L. Xikun, and Z. Yi-di, "Gun life prediction model based on bayesian optimization cnn-lstm," *Integrated Ferroelectrics*, vol. 228, no. 1, pp. 107–116, 2022.

[6] R. Pravesh and B. C. Sahana, "A dual-stage deep learning framework for simultaneous fire and firearm detection in smart surveillance systems," *Results in Engineering*, p. 106330, 2025.

[7] A. J. Amado-Garfias, S. E. Conant-Pablos, J. C. Ortiz-Bayliss, and H. Terashima-Marin, "Improving armed people detection on video surveillance through heuristics and machine learning models," *IEEE Access*, vol. 12, pp. 111818–111831, 2024.

[8] G. Gupta, S. Chattopadhyay, V. Kukreja, M. Aeri, and S. Mehta, "The arsenal algorithm: Ai-driven weapon recognition with cnn-svm model," in *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, pp. 1–6, IEEE, 2024.

[9] T. Murugan, N. A. N. M. Badusha, A. R. O. A. Semaihi, M. M. R. Alkindi, E. M. R. Alnaqbi, and G. H. T. Alketbi, "Ai-based weapon detection for security surveillance: Recent research advances (2016–2025)," *Electronics*, vol. 14, no. 23, p. 4609, 2025.

[10] M. T. Bhatti, M. G. Khan, M. Aslam, and M. J. Fiaz, "Weapon detection in real-time cctv videos using deep learning," *Ieee Access*, vol. 9, pp. 34366–34382, 2021.

[11] R. Anandhi, "Edge computing-based crime scene object detection from surveillance video using deep learning algorithms," in *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 1159–1163, IEEE, 2023.

DoI: https://doi.org/10.63949/crinfo.v1i4.005