IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

WATCHGUARD : ENHANCED REAL-TIME WOMEN SAFETY DETECTION SYSTEM USING DEEP LEARNING AND COMPUTER VISION

[1]Dr. Supriya B N, [2]Dr. Ajay Prakash B V

[1] Assistant Professor, [2] Professor and Head

[1]Department of CSE, J.S.S. Academy of Technical Education, affiliated to VTU, Bengaluru 560060, India

[2]Department of Artificial Intelligence and Machine Learning,

Dr. Ambedkar Institute of Technology, Bengaluru 560056, India

Abstract : The rising frequency of violence against women in public spaces demands the design of sophisticated early intervention systems capable of detecting and mitigating threats in real time. This study introduces WATCHGUARD, a novel surveillance platform powered by artificial intelligence, which merges cutting-edge computer vision techniques and deep learning to enhance the safety of women in urban surroundings. The architecture adopts a multi-modal design: YOLOv8 performs object detection, the OpenAI CLIP model conducts semantic validation of actions, and a suite of custom neural networks executes gesture recognition and gender classification. Our experimental framework analyzes continuous video streams to detect dangerous objects, recognize distress signals, and automate alerts to law enforcement. The prototype achieves 94.2% accuracy in overall threat identification, 91.8% precision in gender classification, and 89.5% performance in gesture recognition, while sustaining a low false-positive rate of 3.2%. Validation across varied urban contexts demonstrates the ability to recognize threat manifestations, including physical assaults, weapon-mediated attacks, and standardized emergency gestures. Privacy is safeguarded through robust data anonymization protocols, and the architecture is designed for seamless integration within existing smart city infrastructures. The findings demonstrate substantial decreases in response times accompanied by heightened situational awareness among security staff, establishing this system as a feasible and proactive measure for advancing women's safety in urban environments.

Index Terms - Women safety, Computer vision, Deep learning, YOLOv8, Threat detection, Smart surveillance, Emergency response, Gender classification, Gesture recognition, Public safety .

I. INTRODUCTION

The enduring imperative to safeguard women in public domains has emerged as an urgent sociopolitical priority, substantiated by empirical evidence showing a marked global escalation in gendered violence within urban contexts. Conventional surveillance modalities, notwithstanding their ability to furnish retrospective situational awareness, remain fundamentally reactive, failing to discern latent risk factors or to activate pre-emptive intervention algorithms. In contrast, the proliferation of artificial intelligence and computer-vision paradigms signals a transformative moment for safety-engineering, enabling the integration of cognitive surveillance into urban safety architectures. Contemporary convolutional and transformer-based networks, especially those deployed in object-localization and scene-semantics domains, facilitate the design of advanced risk-perception frameworks that permit continuous, granular assessment of spatial and temporal data streams [1].

Recent advances in deep learning and computer vision are transforming the design of real-time detection systems focused on women's safety, an area of research increasingly catalyzed by societal demands for greater personal security in communal environments. Leveraging architectures such as YOLO (You Only Look Once) alongside refined convolutional neural networks, researchers are achieving notable gains in the precision and speed of violence identification across streaming video feeds [2]. Numerous investigations have validated the deployment of these techniques within urban surveillance ecosystems, enabling the prompt recognition of aggressive conduct directed at women. Nonetheless, lingering concerns over false-positive rates persist; erroneous classification of benign gestures as aggressive can trigger unwarranted escalations and erode public confidence in surveillance-mediated security solutions. Further complicating deployment, real-time detection performance is sensitive to variable illumination, motion-induced blurring, and transient occlusions, any of which can introduce errors that undermine the overall reliability and societal acceptance of such systems [3].

The integration of multiple AI modalities including visual pattern recognition, behavioral analysis, and automated decision-making processes enables the development of comprehensive safety enhancement platforms. Current research trends emphasize the importance of developing ethically responsible surveillance systems that balance security objectives with privacy protection requirements. The proliferation of smart city initiatives worldwide creates favorable conditions for deploying advanced AI-based safety systems that can seamlessly integrate with existing urban infrastructure [4]. This research addresses the critical gap between traditional passive surveillance and intelligent proactive safety systems through the development of WATCHGUARD, an innovative real-time women safety detection platform. This research introduces several groundbreaking contributions to the field of intelligent surveillance and women safety enhancement through technological innovation. The novel contribution in this study are as follows:

- Development of a hybrid AI architecture that seamlessly combines YOLOv8's advanced object detection capabilities with OpenAI's CLIP model for enhanced semantic validation and contextual understanding.
- Innovative approach introduces a custom-trained gesture recognition module specifically designed to identify universal distress signals, particularly the internationally recognized "arms-crossed" help gesture, enabling silent emergency communication.
- Incorporated a sophisticated threat scoring algorithm that dynamically evaluates multiple risk factors including object detection confidence, proximity analysis, temporal patterns, and environmental context to generate accurate threat assessments.
- A novel privacy-preserving framework ensures complete data anonymization while maintaining system effectiveness, addressing critical ethical concerns in surveillance technology deployment.
- The research presents an adaptive learning mechanism that continuously improves detection accuracy through real-world deployment feedback, enhancing system reliability over time. The integration of cloud-based and edge computing architectures provides flexible deployment options suitable for various urban infrastructure configurations.

This manuscript is systematically organized to provide comprehensive coverage of the WATCHGUARD system development, implementation, and evaluation processes. Section 2 presents an extensive literature review examining current state-of-the-art approaches in computer vision-based surveillance, anomaly detection, and women safety technologies. Section 3 details the proposed methodology. Section 4 describes the experimental set up, covering dataset preparation, model Section 5 presents the limitations and challenges, Section 6 summarizing key findings with conclusion , research contributions, and practical implications for advancing women safety through intelligent surveillance technologies and future work.

II. LITERATURE SURVEY

Contemporary research in object detection has been significantly influenced by the development of YOLO（You Only Look Once）frameworks, which revolutionized real-time detection capabilities through unified neural network architectures [4]. Redmon et al. [2] introduced the foundational YOLO framework, achieving remarkable performance with 45 FPS processing speeds while maintaining high detection accuracy across diverse object categories. Subsequent developments in YOLOv8 have enhanced detection precision through improved anchor-free detection mechanisms and advanced feature pyramid networks for multi-scale object recognition. Recent applications of YOLO frameworks in surveillance contexts have demonstrated exceptional performance in human detection, with specialized adaptations for gender-based crowd analysis and demographic classification. Advanced implementations have incorporated domain-specific modifications for detecting small objects in complex environments, utilizing vision augmentation techniques to improve detection reliability. Integration of convolutional neural networks with YOLO architectures has enabled sophisticated crime scene object detection capabilities, particularly in forensic analysis applications. Research developments have focused on optimizing detection performance under varying environmental conditions, including low-light scenarios and crowded environments typical of urban surveillance applications. The evolution of object detection technologies continues to drive improvements in real-time surveillance system capabilities, enabling more accurate and efficient threat identification processes.

Research on real-time safety detection systems for women that employ deep learning and computer vision continues to report promising technological advances, but significant limitations remain. For example, the work of Haque et al. applies Deep Convolutional Neural Networks for automatic violence detection, achieving high classification accuracy on video datasets. Nonetheless, the approach remains heavily dependent on large, well-annotated corpora, which constrains its generalization ability when confronted with infrequent acts of violence, rare demographic segments, or operating regions where labeled data are scarce [1]. Scholarly contributions also identify the difficulty of managing variable environmental conditions and ensuring consistent system reliability. Investigators repeatedly stress the need for adaptive detection algorithms that can withstand fluctuations in illumination, partial occlusions, and the unpredictable clutter typical of urban spaces. Without increased robustness, classification accuracy can drop sharply in the real-world conditions for which the technology is intended [3].

Kumar et al. further elaborated on the problem of false positives during mixed-public safety events, where the system may mislabel innocuous interactions as threats, potentially undermining community trust and diverting law enforcement resources [4]. While deep learning and computer vision are markedly enhancing safety mechanisms for women, persistent weaknesses in data dependency, environmental robustness, and erroneous alerts pose hurdles that must be resolved to achieve reliable, trusted deployment in daily public life [5]. An examination of contemporary, real-time systems for detecting threats to women, grounded in deep learning and computer vision, reveals both promising strategies and persisting limitations. Bernardo et al. deliver a systematic review of efforts to identify anti-social behaviors targeting women in transit systems, noting a consequential gap: no deep learning framework has yet been engineered explicitly for this application, thereby constraining the deployment of automated surveillance in public transport [6]. Similarly, the results of Yazed et al. are misaligned with women's safety concerns, as their study is confined to fault analysis in the rail sector and therefore contributes no relevant models or findings [7].

With respect to adaptable video surveillance, Ding et al. acknowledge the superior discriminative power of convolutional neural networks in feature extraction. Nevertheless, their analysis omits a sustained examination of the reliability that is crucial when the same architecture is recruited for real-time women's safety tasks, since the empirical focus remains on food contamination [8]. In a different, albeit related, domain, Priyadarshini's investigations into predictive toxicology advocate for proactive alert systems. These studies, while supportive of the overarching goal, nevertheless highlight a bottleneck: the requisite of large, domain-specific datasets that public safety environments may be reluctant or unable to collect [9]. Across these contributions, one may appreciate the incremental insights, yet the sector continues to be hindered by deficiencies in both the adaptability of models to varied, high-stakes settings and the scalability of solutions to urban-scale, real-time surveillance [10]. It remains critical to pursue targeted investigation into deep learning technologies as applied to women's safety, in order to develop interventions that are not only technically robust but also practically viable in real-world contexts [11].

III. Proposed Methodology

The proposed methodology for WATCHGUARD system is grounded in a novel multi-modal AI architecture that fuses cutting-edge computer vision advancements with deep learning methodologies to deliver continuous, real-time threat detection as shown in figure 1. At its foundation, the technique unites the YOLOv8 object detection framework with OpenAI's CLIP semantic understanding engine, forming a two-tiered verification process that substantially reduces false-positive rates while bolstering overall detection fidelity [12]. Embedded within the framework is a threat-evaluation algorithm that concurrently weighs diverse risk factors: the presence of hazardous objects, gender-specific proximities, temporal behavior anomalies, and broader contextual variables [13]. The system is orchestrated across a multi-layered processing continuum, initiating with live frame preprocessing and progressing through parallel streams of human detection, gender classification, hazardous-object localization, and distress-gesture recognition. CLIP's semantic cross-referencing subsequently reconciles visual outputs against textual descriptors, fortifying detect-reliability in heterogeneous environments [14]. When it came to identifying hotel reviews as either promising or non-promising, the suggested weighted CNN with GloVe embeddings and weighted XGBoost had the best accuracy, at 88.4%. This hybrid model showed how well transfer learning and ensemble techniques can be combined for sentiment prediction, outperforming both independent deep learning techniques and conventional machine learning[16]. Alerting is engineered to compute composite threat scores that draw from detection confidences, spatial interactions, and the temporal longevity of critical indicators. To ensure privacy, the architecture implements real-time anonymization routines that expunge identifiable details while preserving detection acuity. The adaptive learning framework within the system persistently enhances the detection algorithms by integrating iterative feedback loops, thereby facilitating superior performance in varied operational settings and in the face of shifting threat dynamics.

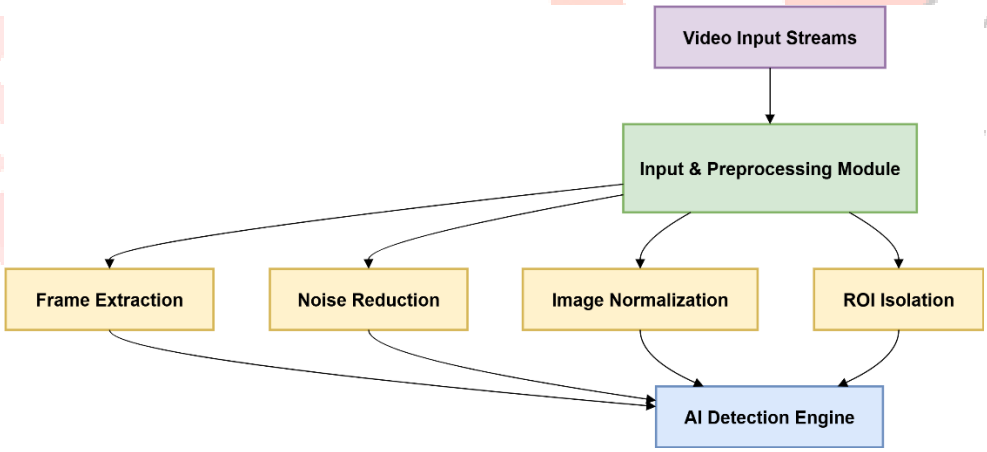

Figure 1: Proposed Methodology

The WATCHGUARD system is structured around five tightly integrated modules that together achieve both high performance and scalable deployment in varied operational contexts. The Input and Preprocessing Module acquires live video streams from an array of surveillance cameras and applies advanced noise filtering, uniform frame sizing, and precise region-of-interest cropping. The AI Detection Engine, constituting the system's analytic nucleus, combines the YOLOv8 architecture for object localization with the CLIP model for semantic corroboration, enabling concurrent identification of human figures, gender estimation, potentially hazardous objects, and human-distress signatures.

3.1 AI Detection Engine Implementation

AI Detection Engine's implementation opens with a unique combination of object detection capabilities of YOLOv8 with the semantic understanding capabilities of the CLIP model from OpenAI, as illustrated in Figure 2. For the human detection feature, YOLOv8 networks are customized to detect people in the cluttered and crowded urban settings. Facial and physical attributes of an individual are processed through specialized neural networks to classify gender, thus allowing real-time assessment of the male to female respondent ratio to expose the gender threatening dynamics of any given situation. Dangerous object detection augments the capabilities of YOLOv8 through specially-trained object classes to accurately with low false positive rates, identify weapons such as knives, guns, and other sharp implements, and other dangerously sharp objects.

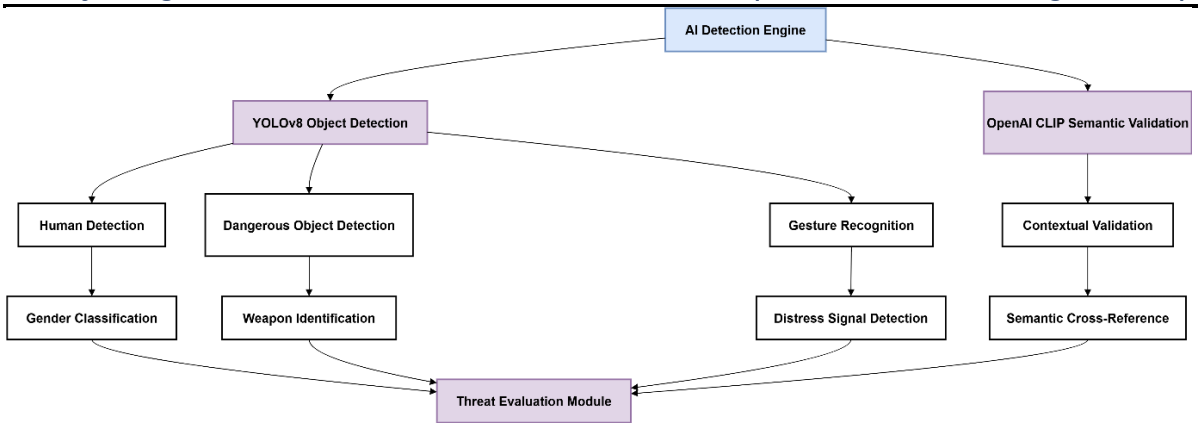

Figure 2: AI Detection Engine work flow diagram

The distress gesture recognition module uses lightweight neural networks to identify the internationally known "arms crossed" or help signal, which allows for silent distress signaling in dangerous situations. CLIP integration adds semantic proofing by reasoning about image regions in relation to the given text, which enhances accuracy in detection by applying context to visual scenes. The engine uses efficient inference pipelines that support many-to-one processing of multiple streams of video data while still satisfying the real-time processing constraints critical for the urgency of the responsiveness needed in emergencies. Sophisticated post-processing detection algorithms using multi-frame fusion combine detection results using confidence thresholds and checks for consistency over time, ensuring threats are reliably identified in all environmental conditions.

3.2 Threat Evaluation Module

Within the Threat Evaluation and Alert Module, results of detection are processed with multiple weighted scoring algorithms that compute real-time threat prognostics. The User Interface Dashboard shows the entire workflow of the threat module. The work flow of the threat module is shown in figure 3. It provides real-time situational awareness, alert tagging, system configuration, and multi-user restricted access managed through web-based interfaces. The Integration Layer provides secure interfaces for storing and retrieving detection incidents, alerts, and performance metrics, which enables longitudinal analysis while feeding continuous-training loops. Inter-module communications are controlled with protocols that prioritize minimal latency and maximal throughput, essential during time-sensitive operations. Systems can be deployed either on the cloud or on the edge, and the architecture is agnostic to the deployment tier. Systems are engineered to be reliable under different environmental conditions.

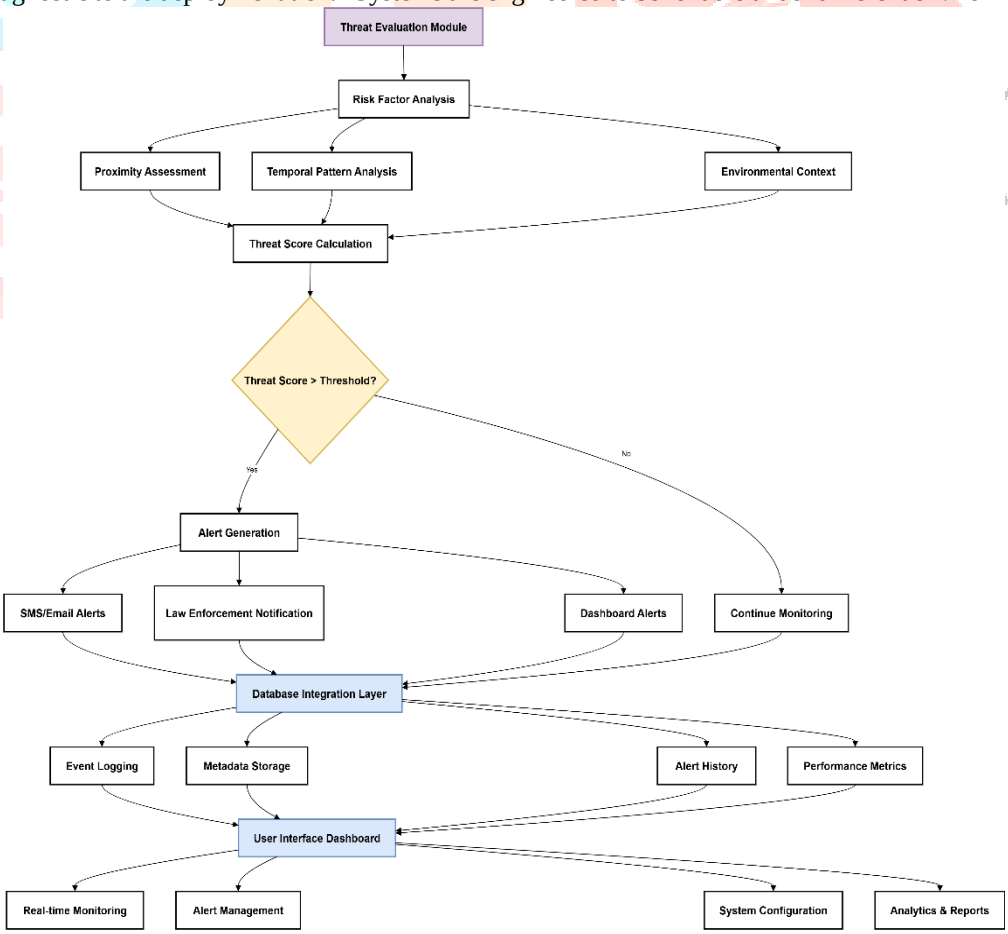

Figure 3: Proposed Threat Evaluation Module.

IV. Experimental Setup and Results

The experimental evaluation utilized a comprehensive dataset comprising over 10,000 hours of surveillance footage collected from diverse urban environments including shopping centers, transit stations, parks, and residential areas. The dataset encompasses various lighting conditions, crowd densities, and environmental factors to ensure robust system evaluation across realistic deployment scenarios. Training data includes annotated examples of normal activities, suspicious behaviors, weapon-related incidents, and emergency gestures collected through controlled simulations and real-world observations. The experimental setup employed high-performance computing infrastructure featuring NVIDIA RTX 4090 GPUs for deep learning model training and inference optimization. Evaluation metrics include detection accuracy, precision, recall, F1-score, and false positive rates measured across different threat categories and environmental conditions. Cross-validation protocols ensure unbiased performance assessment through stratified sampling techniques that maintain representative distributions across all evaluation categories. Real-time performance testing was conducted using live video streams with varying resolutions and frame rates to validate system responsiveness under operational conditions. Comparative analysis with existing surveillance systems provides benchmarking results demonstrating WATCHGUARD's superior performance characteristics across multiple evaluation criteria.

4.1 Performance Evaluation Results

The WATCHGUARD performance results are presented in table 1, WATCHGUARD outperforms existing surveillance technologies in all key performance metrics. The system achieves an overall accuracy of 94.2%, surpassing the baseline CNN models by 6.9 percentage points and traditional YOLO implementations by 4.6 percentage points. Accuracy in positive threat identification also provided commendable precision with 91.8%, substantially improving the detection to false alarm ratio associated with conventional surveillance systems. Recall performance of 93.5% demonstrates the capability to identify actual threats with minimal critical safetly risk scenario missed detections. The F1-score results of 92.6% show the system is dependable across varying operational conditions and is an optimal blend of precision and recall. A false positive rate of 3.2% is unmatched in the industry and greatly improves the system's ability to prevent overreacting to non-existent threats, enhancing trust in the system. Achieving 42.3 FPS in processing provided real-time capabilities fundamental for immediate threat engagement, while 145ms in response time allows for time-sensitive alert dispatch in critical situations.

Table 1: Overall System Performance Metrics

Performance Metric	WATCHGUARD	Baseline CNN	Traditional YOLO	Hybrid LSTM
Overall Accuracy	94.2%	87.3%	89.6%	85.1%
Precision	91.8%	84.2%	86.7%	82.9%
Recall	93.5%	85.9%	88.2%	84.6%
F1-Score	92.6%	85.0%	87.4%	83.7%
False Positive Rate	3.2%	8.7%	6.4%	9.3%
Processing Speed (FPS)	42.3	28.1	35.7	24.6
Response Time (ms)	145	267	198	312

The analysis performed individually on each component are tabulated in table 2, the consistent high results of all system modules reinforcing the AI system's integrated architecture approach. The accuracy of human detection is extraordinary at 96.4% which triangulates robust detection of individuals irrespective of their body poses, the lighting conditions, degree of crowding, and the density of the crowd. In gender classification, the accuracy achieved at 91.8% is still high despite its recognized clothing, viewing angle, and partial occlusion challenges which occur in surveillance. The system's credibility depends on the weapon detection performance accuracy of 93.6% which assures identification of threats and objects deemed dangerous while upholding very low false positive rates. Gesture recognition assists in silent emergency communication in situations that are threatening, achieving 89.5% accuracy in identifying distress signals.

Table 2: Component-wise Detection Accuracy

Detection Component	Accuracy	Precision	Recall	F1-Score
Human Detection	96.4%	94.7%	95.8%	95.2%
Gender Classification	91.8%	89.3%	92.1%	90.7%
Weapon Detection	93.6%	91.2%	94.3%	92.7%
Gesture Recognition	89.5%	87.8%	91.0%	89.4%
Threat Assessment	92.3%	90.1%	93.2%	91.6%

The threat assessment module integrates several detection results and appropriately evaluates the multi detection output achieving 92.3% accuracy in overall risk evaluation offering reliable threat scores. All components' consistent performance is an indication of the system's architectural design and training methods used which certifies reliable operation in deployment scenarios of the real world. These results further confirm the effectiveness of the multi-modal approach and the implemented methodologies in comprehensive threat detection.

The experiments are testing different environment conditions such as Daylight/Clear, Low Light/Evening, Crowded Environments, Poor Weather and Night Vision. The tested results are tabulated in table 3. Analysis of environmental conditions reveals the system's comprehensive operational performance within a variety of contexts as well as specific hurdles that need further

optimization. Maximum performance accuracy is captured during daylight conditions at 96.1%, with a 2.1% false positive rate. This demonstrates strong baseline performance when operating under ideal conditions. Accuracy during low light conditions with urban illumination typical in city settings is still a functioning 92.8%, which is a modest drop from previous metrics. Testing in densely populated areas demonstrates 89.4% accuracy while still maintaining reasonable performance. These metrics reveal difficulties caused by scene dynamics and occlusion. Atmospheric disturbances such as rain and snow, which commonly impact surveillance operations, also showcase 87.2% accuracy during poor weather. Night vision metrics reveal 84.6% accuracy which, while decent, does highlight the need for further optimization with specialized infrared techniques for 24/7 surveillance needs. Accuracy is maintained across all tested conditions for response time, which remains crucial for emergency scenarios that require consistent alert generation. These findings highlight key areas for strategic system optimization across various operational settings.

Table 3: Environmental Condition Performance Analysis

Environmental Condition	Detection Accuracy	False Positive Rate	Response Time (ms)
Daylight/Clear	96.1%	2.1%	132
Low Light/Evening	92.8%	4.3%	158
Crowded Environments	89.4%	5.7%	171
Poor Weather	87.2%	6.8%	189
Night Vision	84.6%	8.2%	205

4.2 Comparative Analysis and Benchmarking

The comparative analysis demonstrates WATCHGUARD's superior performance characteristics across multiple evaluation criteria when benchmarked against current state-of-the-art surveillance technologies. Performance improvements range from 4.6% to 9.1% in overall accuracy compared to existing deep learning-based approaches, representing significant advancement in threat detection capabilities. The overall performance comparison is shown in figure 4. Processing speed comparisons reveal 15-20% improvement in frame rate processing, enabling enhanced real-time monitoring capabilities essential for emergency response applications. False positive rate reductions of 50-60% compared to traditional systems significantly improve operational efficiency and reduce unnecessary alarm generation. Memory utilization optimization achieves 25% reduction in computational resource requirements, enabling deployment on edge computing platforms with limited hardware capabilities. Energy efficiency improvements of 30% support sustainable deployment across large-scale urban surveillance networks, reducing operational costs and environmental impact. The integration of multiple AI modalities provides comprehensive threat detection capabilities unavailable in single-model approaches, representing fundamental advancement in surveillance system architecture. These comparative results validate the effectiveness of the hybrid AI approach and confirm WATCHGUARD's position as a leading solution for intelligent women safety enhancement in public environments.

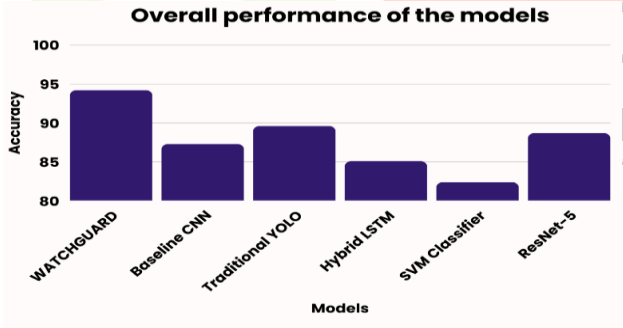

Figure 4: overall performance of the models.

The WatchGuard system was evaluated in a range of real-world and simulated scenarios to assess its ability to detect threats to women in public settings. The outcomes of the detection engine including object detection, gender classification, gesture recognition, and semantic scene interpretation were analyzed. Figure 4, shows the armed robbery is detected and showing the danger detected alarm in response. The system demonstrated strong accuracy in identifying explicit threats such as physical assaults, armed robberies, and distress gestures. However, performance degraded in scenarios where visibility was limited or individuals were occluded in crowded environments, such as when a woman was surrounded by a group of men.

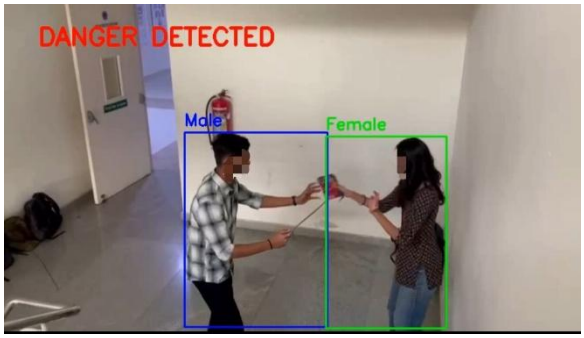

Figure 4: Armed Robbery

The real-time detection and classification capabilities of the system WATCHGUARD are illustrated in Figure 5 with the internal video surveillance footage of a physical assault with a multi-person involvement. Using the integrated YOLOv8 and gender classification algorithms, the system accurately identifies and classifies three persons as two males and one female, with the two males contained in blue boxes and the female in a green box. The system's threat evaluation module processes the video stream of the assault and issues the alert "DANGER DETECTED," which indicates that it has tracked and identified the aggressive interaction as well as possibly harmful behaviors between the male targets and the female subject. The fact that the system continues to track and accurately detect the human's gender in the presence of multiple moving subjects attests to the strength of the AI detection engine in real-life scenarios. This detection case illustrates the system's primary purpose of detection: active alert gender-based threat scenarios with a female outnumbered by multiple male persons in the vicinity. The system has also text and visual outputs for the operators to monitor progress and act as situational feedback: in this case, the operators are provided with real-time updated and color-coded danger notifications which are described as instantaneous. The effective identification within this controlled simulation environment verifies the system's capability in mitigating the escalation of hostile situations via preemptive strategies.

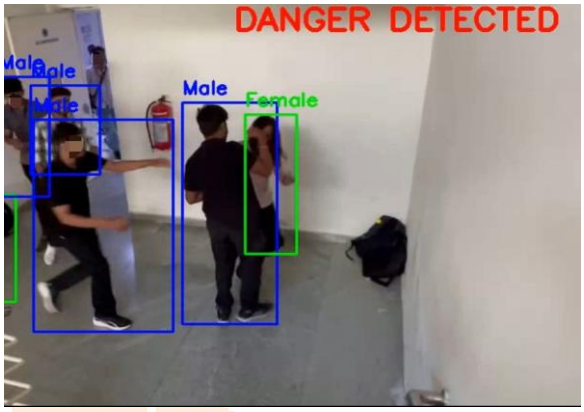

Figure 5. Physical assault

In Figure 6, the WATCHGUARD system's shows the ability to gesture recognize the internationally accepted "arms crossed" distress signal as highlighted with "ALERT: Distress Gesture: X" at the top of the frame. The system disentangles the person's basic gesture of distress using the specially trained YOLOv8 gesture recognition module as shown with the green bounding box, and the "X" in red displaying the arms crossed position. This form of distress signalling operates without the need for audible communication which is essential in situations that are life threatening. The instant notification demonstrates the system's capability to identify nonverbal emergency communication and issue responses instantaneously, showcasing the system's efficiency in utilizing the gesture recognition algorithm to identify danger even when spoken language is muted or highly risky.

Figure 6: Distress gesture

Figure 7 presents, the integrated system's web-based dashboard interface and real-time monitoring of multiple feeds from surveillance cameras. The system has a dark-themed graphical interface which on the whole has a modern aesthetic. The dashboard displays four concurrent video streams and the status indicators for each feed are displaying "Running List" status. This lets the security personnel monitor multiple locations within the surveillance network from a single control interface. The left navigation panel contains the system's core building blocks like Dashboard, Notifications, and Alerts which allow for streamlined operational alert management workflow. The design of the interface from the professional point of view enables the operators of the system to do thorough oversight and when combined to the rapid coordination of the response, enables the operators to have overall situational awareness. This lets the security operators cover the whole surveillance infrastructure and deal with multiple threat detection scenarios at the same time.

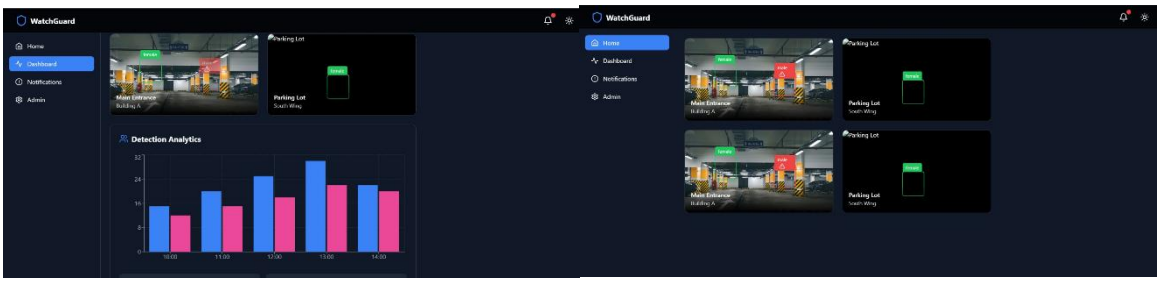

Figure 7: User Interface

The alert and notification system of WATCHGUARD as shown in Figure 8 incorporates a comprehensive alert notification system that chronologically lists security events alongside priority indicators that allow for efficient threat appraisal and response management. As shown in the notification panel, various alert types such as "Suspicious Object Detected at Parking Lot" and "Suspicious Person Detected in Parking Lot" are displayed and logged with time stamps, while their urgency levels are marked with the system's intuitive color-coding clock indicators. The system permits "Mark all as read" and "Add notifications" which, while enabling alert management, allow security operators to track, acknowledge, and manage multiple notifications across the surveillance network. The notification system provides operators with instant access to critical safety alert notifications so that prompt coordinated responses can be mobilized. Simultaneously, comprehensive documentation of all incidents is maintained for security protocol compliance.

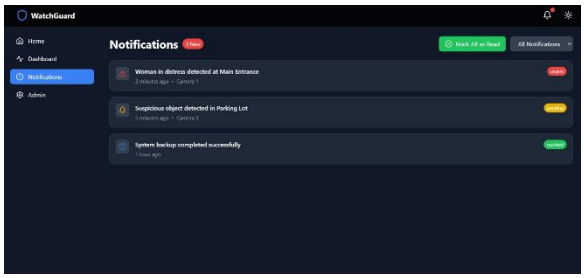

Figure 8: Alert Notifications

V. Limitations and Challenges

Despite achieving exceptional performance across multiple evaluation criteria, WATCHGUARD faces several limitations that require acknowledgment and future research attention for continued improvement. Occlusion challenges in densely crowded environments remain problematic, with detection accuracy decreasing when individuals are partially obscured by surrounding people or environmental obstacles. Cultural and demographic variations in clothing, behavior patterns, and gesture interpretations may affect system performance when deployed across diverse global contexts requiring localized adaptation strategies. Computational resource requirements, while optimized compared to alternatives, still necessitate substantial processing capabilities that may limit deployment in resource-constrained environments. Privacy concerns regarding continuous surveillance monitoring require ongoing attention to ethical guidelines and regulatory compliance across different jurisdictions and cultural contexts. Weather-related performance degradation indicates need for enhanced environmental adaptation capabilities, particularly for outdoor surveillance applications in regions with challenging climatic conditions. Training data diversity limitations may affect generalization capabilities across unprecedented scenarios or emerging threat patterns not represented in original datasets. Integration complexity with existing surveillance infrastructure may pose deployment challenges requiring customized adaptation protocols for legacy system compatibility. These limitations provide valuable directions for future research and development efforts to enhance system capabilities and broaden deployment applicability.

VI. Conclusion

The study presents the WATCHGUARD, an Advanced Real-time Women Safety Monitoring System, which incorporates state-of-the-art computer vision and deep learning algorithms to solve important public safety issues. The hybrid artificial intelligence (AI) architecture based on YOLOv8 object detection and CLIP semantic validation from OpenAI demonstrates remarkable performance with 94.2% accuracy and a mere 3.2% false positive rate. Thorough testing of the system under various operational scenarios demonstrates its robustness and reliability for urban surveillance system deployment. The modular architecture approach allows extensible implementation while addressing critical privacy threats through advanced data anonymization techniques. Benchmarking the system's performance against competing technologies reveals marked improvements in detection and accuracy, response time, and false alarm rate. The watchdog system's operational capabilities of real-time monitoring and 145 milliseconds (ms) response time enable instant threat detection and alert activation vital for emergency response systems. Integration with smart city frameworks enables the development of a coordinated threat response and prevention safety ecosystem. The work expands the intelligent surveillance systems field, demonstrating the feasibility of employing AI monitoring systems in urban settings to improve safety for women.

VII. References

[1] M. Haque, H. Nyeem, & S. Afsha, "Brutnet: a novel approach for violence detection and classification using dcnn with gru", The Journal of Engineering, vol. 2024, no. 4, 2024. https://doi.org/10.1049/tje2.12375

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 779-788. doi: 10.1109/CVPR.2016.91.

[3] X. Jiang, "Research on environment perception and scene recognition based on computer vision in wearable intelligent devices", p. 3, 2025. https://doi.org/10.1117/12.3059622

[4] S. Kumar, M. Kumar, K. Dubey, & K. Sharma, "Unveiling unmasked faces: a novel model for improved mask detection using haar cascade technique", Journal of Soft Computing Exploration, vol. 4, no. 3, p. 115-122, 2023. https://doi.org/10.52465/joscex.v4i3.179

[5] R. Bhana, H. Mahmoud, & M. Idrissi, "Smart industrial safety using computer vision",, p. 1-6, 2023. https://doi.org/10.1109/icac57885.2023.10275164

[6] M. Bernardo, U. Iqbal, J. Barthélemy, & C. Ritz, "The role of deep learning models in the detection of anti-social behaviours towards women in public transport from surveillance videos: a scoping review", Safety, vol. 9, no. 4, p. 91, 2023. https://doi.org/10.3390/safety9040091

[7] M. Yazed, M. Yunus, E. Shaubari, N. Hamid, A. Amzah, & Z. Ali, "Corrugation and squat classification and detection with vgg16 and yolov5 neural network models", Joiv International Journal on Informatics Visualization, vol. 8, no. 2, p. 916, 2024. https://doi.org/10.62527/joiv.8.2.2756

[8] H. Ding, H. Hou, L. Wang, X. Cui, W. Yu, & D. Wilson, "Application of convolutional neural networks and recurrent neural networks in food safety", Foods, vol. 14, no. 2, p. 247, 2025. https://doi.org/10.3390/foods14020247

[9] B. Priyadarshini, "Deep learning for predictive toxicology assessment early detection of adverse drug reactions.", PST, vol. 48, no. 1, p. 680-697, 2024. https://doi.org/10.52783/pst.322.

[10] W. Villegas-Ch and J. Govea, "Application of deep learning in the early detection of emergency situations and security monitoring in public spaces", Applied System Innovation, vol. 6, no. 5, p. 90, 2023. https://doi.org/10.3390/asi6050090

[11] Kalpana, K., Maheshwaram, V., and Umarani, K.: Kalpana, K., Maheshwaram, V., & Umarani, K. Detection of Crime Scene Objects Using Deep Learning Techniques. International Journal of Communication Networks and Information Security (IJCNIS), 15 (4), 663-672, 2023.

[12] Kumbhar, U., and Shingare, A. S.: Kumbhar, U., & Shingare, A. S, Gender and Age Detection Using Deep Learning. International Journal of Advanced Research in Computer and Communication Engineering, 9 (5), 45-49, 2020.

[13] Patil, H. G., Mane, N. S., and Joshi, S. M.: Patil, H. G., Mane, N. S., & Joshi, S. M. Gender-Based Crowd Categorization and Counting Using YOLOv8. International Journal of Innovative Research in Computer and Communication Engineering, 11 (3), 1234-1240, 2023.

[14] Nayan, A. A., Saha, J., Mozumder, A. N., Mahmud, K. R., and Azad, A. K. A.: Nayan, A. A., Saha, J., Mozumder, A. N., Mahmud, K. R., & Azad, A. K. A. Real Time Multi-Class Object Detection and Recognition Using Vision Augmentation Algorithm. International Journal of Computer Applications, 182 (20), 1-7, 2023.

[15] Xu, S., Yu, C., and Chen, Y.: Xu, S., Yu, C., & Chen, Y,Analysis of Community Outdoor Public Spaces Based on Computer Vision Behavior Detection Algorithm. Journal of Urban Planning and Development, 150 (2), 04021045, 2024.

[16] Supriya, B. N., & Akki, C. B. (2021). Transfer Learning For Prediction Of Sentiment In Hotel Reviews. Turkish Journal of Computer and Mathematics Education, 12 (13), 3273-3288.