YOLO in Suspicious Human Activity Recognition for Intelligent Environmental Security Systems: A Review

Yohanes Bowo Widodo [1)*)], Sondang Sibuea [2)], Rano Agustino [3)]

[1)] Department of Informatics, Universitas Nusa Mandiri
[2)] Teknik Informatika, Fakultas Komputer, Universitas Mohammad Husni Thamrin
[3)] Sistem Informasi, Fakultas Komputer, Universitas Mohammad Husni Thamrin
[*)]**Correspondence author**: ybowowidodo@gmail.com , DKI Jakarta, Indonesia
DOI: https://doi.org/10.37012/jtik.v12i1.3243

Abstract

The rapid growth of intelligent environmental security systems has intensified the need for accurate and real-time suspicious human activity recognition. Computer vision techniques, particularly deep learning–based object detection models, have emerged as key enablers in addressing these challenges. Among them, You Only Look Once (YOLO) has gained significant attention due to its high detection speed, end-to-end architecture, and suitability for real-time surveillance applications. This review paper presents a comprehensive analysis of the application of YOLO-based models in suspicious human activity recognition for intelligent environmental security systems. It examines the evolution of YOLO architectures, their adaptations for activity and behavior analysis, and their integration with surveillance frameworks. The review further discusses commonly used datasets, performance evaluation metrics, and comparative results reported in existing studies. In addition, key challenges such as occlusion, varying illumination, complex backgrounds, privacy concerns, and computational constraints are highlighted. Finally, the paper outlines future research directions, including hybrid models, multi-modal data fusion, edge-based deployment, and explainable AI, to enhance the robustness and reliability of YOLO-driven security systems. This review aims to provide researchers and practitioners with a structured understanding of current advancements and open issues in YOLO-based suspicious human activity recognition.

Keywords: *Suspicious Activity Recognition, Intelligent Security Systems, Human Behaviour Analysis, Computer Vision, Deep Learning*

Abstrak

Pertumbuhan pesat sistem keamanan lingkungan cerdas telah meningkatkan kebutuhan akan pengenalan aktivitas manusia yang mencurigakan secara akurat dan real-time. Teknik visi komputer, khususnya model deteksi objek berbasis pembelajaran mendalam, telah muncul sebagai kunci utama dalam mengatasi tantangan ini. Di antara teknik tersebut, You Only Look Once (YOLO) telah mendapatkan perhatian yang signifikan karena kecepatan deteksinya yang tinggi, arsitektur ujung-ke-ujung, dan kesesuaiannya untuk aplikasi pengawasan real-time. Makalah tinjauan ini menyajikan analisis komprehensif tentang penerapan model berbasis YOLO dalam pengenalan aktivitas manusia yang mencurigakan untuk sistem keamanan lingkungan cerdas. Makalah ini mengkaji evolusi arsitektur YOLO, adaptasinya untuk analisis aktivitas dan perilaku, dan integrasinya dengan kerangka kerja pengawasan. Tinjauan ini selanjutnya membahas kumpulan data yang umum digunakan, metrik evaluasi kinerja, dan hasil perbandingan yang dilaporkan dalam studi yang ada. Selain itu, tantangan utama seperti oklusi, variasi pencahayaan, latar belakang yang kompleks, masalah privasi, dan kendala komputasi disoroti. Terakhir, makalah ini menguraikan arah penelitian masa depan, termasuk model hibrida, fusi data multi-modal, penerapan berbasis edge, dan AI yang dapat dijelaskan, untuk meningkatkan kekokohan dan keandalan sistem keamanan berbasis YOLO. Tinjauan ini bertujuan untuk

memberikan pemahaman terstruktur kepada para peneliti dan praktisi mengenai kemajuan terkini dan isu-isu terbuka dalam pengenalan aktivitas manusia yang mencurigakan berbasis YOLO.

Kata Kunci: Pengenalan Aktivitas Mencurigakan, *Intelligent Security Systems*, Analisis Perilaku Manusia, *Computer Vision, Deep Learning*

INTRODUCTION

This research is important because the increasing complexity of modern environments—such as smart cities, public transportation hubs, campuses, and critical infrastructure—demands intelligent security systems capable of detecting suspicious human activities accurately and in real time. Traditional surveillance methods rely heavily on manual monitoring, which is time-consuming, error-prone, and ineffective at scale. By reviewing the role of YOLO-based models in suspicious human activity recognition, this research highlights how advanced deep learning techniques can significantly improve situational awareness, reduce response time to potential threats, and enhance public safety. Moreover, understanding the strengths, limitations, and practical challenges of YOLO in real-world security applications helps guide future system design, promotes efficient deployment on resource-constrained devices, and supports the development of more reliable, scalable, and ethical intelligent environmental security solutions.

Current methods for suspicious human activity recognition primarily rely on deep learning–based computer vision approaches, combining object detection, human pose estimation, and action recognition techniques. Convolutional Neural Networks (CNNs) and two-stage detectors such as Faster R-CNN have been widely used for accurate human and object detection, while single-stage detectors like YOLO and SSD are preferred for real-time surveillance due to their high speed. In addition, temporal modeling techniques—including Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and 3D CNNs—are commonly employed to capture motion patterns across video frames. Recent studies also integrate pose-based methods using frameworks like OpenPose and graph convolutional networks to recognize complex behaviors. Furthermore, hybrid systems combining YOLO for object detection with tracking algorithms (e.g., SORT, Deep SORT) and anomaly detection models are increasingly adopted to improve robustness in dynamic

and cluttered environments. These methods collectively form the foundation of current intelligent surveillance systems for suspicious activity recognition.

YOLO (You Only Look Once) plays a central role as an efficient and real-time object detection framework for suspicious human activity recognition within intelligent environmental security systems. YOLO treats detection as a single, end-to-end regression problem, enabling it to simultaneously localize and classify multiple objects—such as humans, weapons, or suspicious items—within a single forward pass of the network. Its high processing speed makes it particularly suitable for continuous video surveillance in dynamic environments where timely threat detection is critical. Recent versions of YOLO further enhance detection accuracy through improved feature extraction, multi-scale prediction, and attention mechanisms, allowing better performance under challenges such as occlusion, low lighting, and crowded scenes. By serving as a foundational detection module, YOLO is often integrated with tracking, activity recognition, and temporal analysis models, making it a key enabler for scalable, responsive, and intelligent security monitoring systems.

This review provides a comprehensive analysis of YOLO-based approaches for suspicious human activity recognition by systematically reviewing existing architectures, application scenarios, datasets, and evaluation metrics used in intelligent environmental security systems. It synthesizes comparative findings across different YOLO versions and hybrid frameworks, highlighting performance trade-offs between accuracy, speed, and computational cost. However, several limitations are also identified. Many studies focus primarily on object detection rather than full activity understanding, limiting the ability to recognize complex or subtle behaviors. Performance often degrades in real-world conditions involving severe occlusion, dense crowds, low-resolution footage, or varying illumination. Additionally, the reliance on labeled surveillance datasets raises concerns related to data availability, bias, and generalization across environments. Computational constraints on edge devices and unresolved privacy and ethical issues further restrict large-scale deployment. These limitations underscore the need for more robust, context-aware, and ethically responsible YOLO-based security solutions.

This review contributes to the field by presenting a structured and up-to-date review of YOLO-based methods for suspicious human activity recognition in intelligent

environmental security systems. It consolidates scattered research findings by analyzing the evolution of YOLO architectures and their adaptations for surveillance and activity recognition tasks. The study offers a comparative perspective on existing approaches, datasets, and performance metrics, enabling clearer understanding of current capabilities and limitations. Additionally, it identifies key research gaps and technical challenges, such as real-time deployment constraints, activity-level understanding, and privacy concerns. By outlining future research directions, this review serves as a valuable reference for researchers and practitioners seeking to design more effective, scalable, and reliable YOLO-driven security systems.

RESEARCH METHOD

The method of this research is based on a systematic and structured review of existing literature related to YOLO-based suspicious human activity recognition in intelligent environmental security systems. Relevant research articles were collected from well-established scientific databases, focusing on studies that apply YOLO or its variants to surveillance, activity recognition, and security monitoring. The selected works were analyzed according to key criteria, including YOLO architecture versions, application domains, datasets used, evaluation metrics, and reported performance outcomes. This approach enables a clear understanding of how YOLO has evolved and how it is currently utilized within intelligent security frameworks.

In the second stage, the reviewed studies were comparatively analyzed to identify trends, strengths, limitations, and research gaps. YOLO-based detection methods were examined both as standalone solutions and as components integrated with tracking, temporal modeling, and activity recognition techniques. Challenges such as real-time constraints, environmental complexity, data imbalance, and ethical considerations were also systematically assessed. Based on this comprehensive analysis, future research directions and methodological improvements were formulated to guide the development of more robust, efficient, and scalable suspicious human activity recognition systems.

RESULTS AND DISCUSSION

Table 1. Summary of Relevant Survey

Reference	Year	Primary Focus	Main Contributions
[1]	2025	Weapon Detection System (WDS) utilizing advanced deep learning and computer vision techniques, specifically YOLO and CNNs), to enable the real-time recognition and classification of firearms and other hazardous weapons for enhanced public safety and threat detection.	An enhanced YOLO-based Weapon Detection System (WDS) that achieves improvements in real-time threat identification, detection accuracy, and computational efficiency, supported by training on a diverse dataset and exploring the integration of edge AI processing to function efficiently on resource-constrained devices for practical surveillance applications.
[2]	2025	Examine and integrate the use of deep learning technologies, specifically CNNs and YOLO, for real-time facial recognition and weapon detection to Enable proactive crime prevention and improve public safety.	Providing novel knowledge on how to combine YOLO, Faster R-CNN, and CNN-based models to perform real-time facial recognition and weapon detection as incidents occur, thereby closing the gap between detection and action and offering a functional advancement toward improving public safety infrastructure and proactive crime prevention.
[3]	2025	To enhance the automation and effectiveness of real-time threat detection in video surveillance by developing a Mono-Scale CNN-LSTM Fusion Network that integrates ORB feature extraction to accurately identify and classify suspicious	The development of a Mono-Scale CNN-LSTM Fusion Network that integrates ORB (Oriented FAST and Rotated BRIEF) feature extraction to optimize spatial and temporal analysis, achieving an

Reference	Year	Primary Focus	Main Contributions
		criminal activities such as robbery, shoplifting, and fighting.	impressive 99% accuracy rate on the UCF crime dataset while significantly reducing the need for continuous human monitoring through automated real-time threat detection and alarm generation.
[4]	2024	To provide a comprehensive review and comparative analysis of various supervised and unsupervised machine learning techniques, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Hidden Markov Models (HMMs), for video-based Human Activity Recognition (HAR) to enable intelligent surveillance systems to detect and differentiate between normal and suspicious behaviors in real-time	A comprehensive review and comparative analysis of various supervised and unsupervised machine learning techniques for Human Activity Recognition (HAR), providing a detailed investigation of their feature extraction methods, performance parameters, and limitations while identifying future research trends and challenges for practical implementation in real-world surveillance systems
[5]	2022	To develop and evaluate an automated surveillance system that utilizes Human Activity Recognition (HAR) and deep learning algorithms to detect and classify suspicious behaviors in real-time such as loitering, intrusion, and aggression to facilitate proactive crime prevention and minimize the need for manual monitoring.	The development of a new system architecture for real-time suspicious behavior detection that utilizes deep learning and Human Activity Recognition (HAR) to achieve nearly 100% accuracy with low computational complexity, effectively automating the monitoring of multiple camera feeds to provide immediate alerts for activities such as theft, loitering, and violent crimes.
[6]	2022	To provide a comprehensive review and taxonomy of Human Activity	A comprehensive review of ninety-five articles published

Reference	Year	Primary Focus	Main Contributions
		Recognition (HAR) literature published since 2018, systematically evaluating ninety-five articles to categorize their application areas, data sources, and techniques while identifying open research challenges for future study.	since 2018 to generate a detailed taxonomy of Human Activity Recognition (HAR) that categorizes current application areas, data sources, and techniques while highlighting open research challenges such as data collection and hardware limitations to provide a roadmap for future studies.
[7]	2022	To enhance smart city security by developing an Internet-of-Things (IoT) based framework that utilizes a multimodal deep learning approach combining an improved YOLO-v4 for detection and a 3D-CNN for recognition to automatically identify and provide real-time alerts for suspicious human activities such as gun pointing, fighting, and vandalism.	The development of a novel multimodal framework for real-time suspicious activity recognition that utilizes a fine-tuned YOLOv4 for detecting regions of interest and a 3D-CNN for temporal classification, while integrating an Internet-of-Things (IoT) based architecture to manage automated decision-making and alerts in smart city surveillance environments.
[8]	2023	To propose and review an intelligent video surveillance framework that utilizes the YOLOv3 algorithm to monitor multiple zones in real-time, aiming to identify and alert on suspicious human and vehicle activities such as theft, violence, and illegal parking to prevent crimes before they occur.	An intelligent video surveillance framework that utilizes the YOLOv3 algorithm to monitor multiple designated zones in real-time, enabling the automatic detection and classification of suspicious human and vehicle activities to facilitate proactive crime prevention through immediate alerts to authorities.
[9]	2020	To develop an Automated Threat Recognition System that utilizes a hybrid deep learning architecture—	The development of an Automated Threat Recognition System that utilizes a hybrid

Reference	Year	Primary Focus	Main Contributions
		combining InceptionV3 (CNN) for high-level feature extraction and RNN with LSTM cells for temporal sequence analysis—to automatically identify and categorize 12 types of real-world anomalies (such as violence, theft, and arson) in real-time CCTV surveillance recordings.	deep learning architecture—integrating InceptionV3 for high-level feature extraction and a Recurrent Neural Network (RNN) with LSTM cells for temporal analysis—trained on a unique 128-hour large-scale dataset to accurately classify twelve types of real-world anomalies with an overall performance of 97.23% accuracy.
[10]	2023	To develop and evaluate a real-time human activity recognition system that classifies behaviors into criminal, suspicious, and normal categories by comparing the performance of a novel 2D-CNN architecture against pre-trained VGG16 and ResNet50 models using both standard training and transfer learning techniques.	The development of a real-time human activity recognition system that classifies behaviors into criminal, suspicious, and normal categories by systematically comparing the performance of a novel 2D-CNN architecture against pre-trained VGG16 and ResNet50 models utilizing both with and without transfer learning.
[11]	2023	To provide a comprehensive review and literature evaluation of state-of-the-art suspicious human activity recognition systems, detailing their general structures, common solution approaches such as feature extraction and object classification, and the inherent technical challenges like illumination changes and occlusions.	A comprehensive literature review and evaluation of state-of-the-art suspicious human activity recognition systems, providing a detailed overview of their general structures, common solution approaches such as object tracking and feature extraction, and the technical challenges including illumination changes and occlusions inherent in the field.

Golande et. al. (Golande et al., 2025) introduced the necessity for intelligent surveillance systems to counter the increasing threat to public safety, drove the development of the Weapon Detection System (WDS), designed for real-time identification of firearms and other hazardous weapons in public areas. The system's methodology leverages advanced deep learning techniques, primarily employing the YOLO (You Only Look Once) object detection model to swiftly detect and localize potential threats in live video streams or images. Once detected, a Convolutional Neural Network (CNN) is used to classify the objects into specific categories, such as handguns, rifles, or knives. Experimental evaluations demonstrate that the proposed model achieves a high detection rate with minimal false alarms, offering a robust solution that is validated using metrics like precision and recall. However, the system faces several limitations, including the significant challenge of detecting concealed weapons hidden under clothing due to occlusion, the impact of environmental variability (such as varying lighting and crowded surroundings) on accuracy, and constraints related to the scalability of real-time processing when applied to large-scale surveillance networks.

Santhi et. al. (Shanthi & Manjula, 2025) driven by the urgent need for proactive crime prevention and intelligent surveillance, focuses on developing and evaluating deep learning models capable of performing real-time facial recognition and weapon detection in public settings, moving beyond the limitations of traditional post-event analysis. The methodology primarily utilizes an integrated framework evaluating Convolutional Neural Networks (CNNs) for feature extraction and classification, in conjunction with the YOLO (You Only Look Once) and Faster R-CNN object detection models to achieve quick and accurate localization of weapons and faces in video feeds. Results confirm the high performance of these deep learning architectures, with CNNs achieving up to 98% accuracy and the YOLO model reaching 96.53% accuracy in various detection tasks. Specifically regarding real-time application, YOLO models (such as YOLOv5 and YOLOv7) demonstrated superior inference rates of approximately 10 to 16 FPS, making them highly applicable for live CCTV streams. However, the system faces several significant limitations, including the fundamental challenge of accurately detecting weapons hidden under clothing or nonmetallic objects, poor performance under varying lighting conditions, occlusions, or in

complex crowd density scenarios typical of real-world CCTV environments, and the overall difficulty in minimizing false alarms while addressing the substantial computational demands of real-time processing.

Aas et. al. (Aas et al., 2025) Conducted study on the escalating threat of organized crime and the shortcomings of traditional, human-intensive video surveillance systems necessitate the Mono-Scale CNN-LSTM Fusion Network (Introduction), a novel approach aimed at enhancing the automation, sustainability, and high-accuracy of real-time threat detection in CCTV environments. The methodology involves integrating Convolutional Neural Networks (CNN) to extract spatial features and Long Short-Term Memory (LSTM) networks for temporal analysis, optimizing the model by employing the Oriented FAST and Rotated BRIEF (ORB) feature descriptor during pre-processing for improved image processing speed. Experimental results using the UCF crime image dataset demonstrated the model's exceptional capability, achieving an accuracy rate of approximately 99% which significantly surpasses traditional models like CNN, VGG-16, and ResNet-50. However, the paper identifies several limitations inherent in current crime detection technologies, including the persistent challenge of achieving high accuracy for real-time processing due to latency concerns, the difficulty in handling varying camera angles and low-resolution footage, and issues related to scalability when deploying the model in large-scale systems with diverse video sources.

Jahan et. al. (Jahan et al., 2024) presented the increasing prevalence of suspicious activities in crowded public areas like airports and banks has created an urgent need for intelligent surveillance systems capable of real-time video analysis to replace manual, human-intensive monitoring. This paper presents a comprehensive review of video-based Human Activity Recognition (HAR), systematically evaluating the progression from basic preprocessing and motion detection to advanced supervised and unsupervised learning models such as CNNs, RNNs, SVMs, and K-means clustering. Comparative analysis within the sources highlights that deep learning architectures have proven groundbreaking, with CNN-based models achieving up to 99% accuracy and RNN-LSTM networks reaching 98% recognition rates on various activity datasets. Despite these successes, the paper identifies critical limitations, including the lack of specialized labeled datasets for diverse surveillance

anomalies, high computational demands that restrict real-time processing on low-end edge devices, and performance degradation caused by real-world environmental factors like occlusions, camera motion, and variable lighting.

Dekkati at. al. (Dekkati et al., 2022) driven by the alarming increase in global crime rates and the resulting demand for enhanced video surveillance, developed automated surveillance systems capable of Human Activity Recognition (HAR) to identify suspects in real-time. The researchers proposed a new system architecture that leverages deep learning and machine learning algorithms to label video footage based on human behavior, utilizing a semantic approach and background subtraction to extract and analyze foreground blobs. Experimental results conducted on public datasets demonstrate that this approach achieves an accuracy of approximately 100% with low computational complexity, successfully detecting suspicious activities such as loitering, intrusion, and violent crimes. However, the paper identifies critical limitations, including the scarcity of high-quality, professional datasets for unique behaviors, the lack of standardized performance evaluation criteria, and the current unreliability of the system when monitoring behavior over extended durations.

Arshad et. al. (Arshad et al., 2022) presents a comprehensive review and updated taxonomy of Human Activity Recognition (HAR) research published since 2018. The researchers utilized a systematic methodology to analyze ninety-five articles selected from major scientific databases, categorizing them by application areas, data sources, and algorithmic techniques. The results demonstrate that while research into daily living activities (42%) is well-established, there is a notable lack of literature regarding real-time activities such as suspicious behavior detection and healthcare monitoring (24%). The review further identifies Convolutional Neural Networks (CNN) as the most prominent technique used in 25% of studies, with CCTV cameras (52%) and mobile phone sensors (26%) serving as the primary data sources. Finally, the paper highlights significant open challenges and limitations, including the scarcity of large-scale labeled datasets, high computational costs for hardware, and the misalignment of activities during data annotation.

Rehman et. al. (Rehman et al., 2022) addresses the challenges of manual monitoring and the need for enhanced security in smart cities, this research proposes an automated system for human activity recognition (HAR) to identify suspicious behaviors in real-time.

The researchers developed a multimodal framework utilizing an improved YOLO-v4 detector to identify specific regions of interest and a 3D-CNN to classify activities by analyzing both spatial and temporal features. This system is integrated within an Internet-of-Things (IoT) architecture using Ethernet communication and centralized GPU servers to manage automated decision-making and provide immediate alerts. Experimental evaluations conducted on benchmark datasets like UCF-Crime demonstrated that the proposed architecture achieves an overall accuracy of 94.21%, performing exceptionally well in detecting activities such as gun pointing and smoking. However, the paper notes that the deep learning models exhibited overfitting during training, and the system faced challenges with activity confusion, resulting in slightly lower performance when distinguishing between complex actions like fighting and vandalism.

Yadav et. al. (Yadav et al., 2023) recognized that traditional CCTV is often used only for post-crime investigations, this research introduces an intelligent video surveillance framework designed to proactively identify and alert authorities to suspicious behavior in real-time. The methodology utilizes the YOLOv3 algorithm to monitor multiple designated zones—specifically separating human and vehicle activity—by processing video frames through a deep convolutional neural network and pre-trained weights to detect specific objects and behaviors. The results indicate that this system can effectively classify activities as routine or odd, facilitating immediate action against threats such as theft, vandalism, and illegal parking. Despite these benefits, a primary limitation of the system is its absolute dependency on camera hardware, meaning the entire security framework becomes non-functional if a camera is damaged or suffers a technical crash.

Singh et. al. (Singh et al., 2020) addresses the rising frequency of disruptive activities and the impossibility of manual real-time CCTV monitoring, this research proposes an Automated Threat Recognition System designed to filter irregularities from normal behavior. The methodology employs a hybrid deep learning architecture, utilizing InceptionV3 (CNN) for high-level feature extraction through transfer learning and a Recurrent Neural Network (RNN) with LSTM cells to analyze the temporal sequence of movements. This system was trained on a significant, rarely explored large-scale dataset comprising 128 hours of surveillance recordings categorized into 13 groups, including 12

specific anomalies such as fighting, robbery, and arson. Experimental results demonstrate that the optimized model achieves an overall accuracy of 97.23% with reduced overfitting, effectively identifying realistic threats and significantly decreasing the time complexity of manual searches. However, the paper identifies critical limitations, specifically the challenge of real-time execution and the dependency on high processing power, noting that hardware constraints and transmission latency can potentially hinder performance unless optimized through expensive onboard GPUs or specialized software.

Indhumathi et. al. (Indhumathi et al., 2023) Recognized that ensuring individual safety is a primary societal concern, this research addresses the difficulty of distinguishing between normal, suspicious, and criminal human behaviors through automated real-time surveillance. The methodology involves comparing a novel 14-layer 2D-CNN architecture against pre-trained VGG16 and ResNet50 models, employing transfer learning through layer freezing and fine-tuning on a combination of Kaggle-sourced images and 9,000 real-time video frames. The results demonstrate that ResNet50 with transfer learning achieved the highest performance with a 99.18% accuracy rate, while the inclusion of batch normalization and dropout layers successfully mitigated model overfitting. However, the paper identifies that identifying human activities under varying lighting conditions and different real-time scenarios remains a complex issue, and future improvements are needed to specifically distinguish suspicious individuals from the general activities detected.

Gupta et. al. (Gupta & Agarwal, 2023) addresses the rising need to remotely monitor public spaces to prevent threats like terrorism, theft, and vandalism, this paper provides a comprehensive literature evaluation of state-of-the-art suspicious human activity (SHA) recognition systems. The methodology outlines a hierarchical framework for these systems, typically starting with splitting video into frames, followed by background subtraction for object detection, feature extraction of motion or shape, and concluding with object classification using algorithms such as SVM, KNN, or CNNs. The review indicates that intelligent systems can effectively categorize behaviors into normal and abnormal classes, with specific reviewed models achieving efficiency ratings as high as 0.99. Despite these advancements, significant limitations remain, including the difficulty of processing

illumination fluctuations, handling object occlusions, and managing the increased processing time required for complex backgrounds in real-time scenario.

In addition to the articles above, there are several articles related to the recognition of suspicious human activities for smart environmental security which can be explained as follows. Golestani et. al. (Golestani & Moghaddam, 2020) introduces an innovative human activity recognition (HAR) system that utilizes magnetic induction (MI) signals and deep recurrent neural networks (LSTM) to accurately identify physical movements while overcoming the power and cost limitations of conventional wearable sensors. Experimental results demonstrate that the system achieves high performance, such as 98.9% accuracy on the MHAD dataset, and provides superior reliability in lossy dielectric media like the human body compared to standard radio-wave technologies.

Mandalapu et. al. (Mandalapu et al., 2023) provides a comprehensive systematic review of over 150 articles to analyze how machine learning and deep learning algorithms are utilized to identify patterns, trends, and hotspots in criminal activity. The study contributes to the field by archiving publicly available crime datasets, identifying that classification tasks are the primary focus of existing research, and highlighting critical future directions such as the need for interpretable models and ethical considerations regarding privacy and bias.

Yuan et. al. (Yuan et al., 2022) describes PRF-PIR, an interpretable, passive, multi-modal sensor fusion system that integrates Software-Defined Radio (SDR) and a novel motion-induced PIR sensor to mitigate the privacy, cost, and interference issues inherent in single-modality monitoring. This framework employs a Recurrent Neural Network (RNN) with LSTM units to achieve high accuracy rates—98.66% for human identification and 96.23% for activity recognition—while utilizing SHAP methodologies to provide transparency and validate the efficacy of the sensor fusion approach.

Vijeikis et. al. (Vijeikis et al., 2022) introduces a lightweight violence detection framework designed for real-time surveillance that combines a U-Net-like spatial feature extractor using a MobileNet V2 encoder with an LSTM network for temporal classification. The proposed model achieves high performance across various benchmarks, notably an 82%

average accuracy on the complex RWF-2000 dataset, while maintaining a small footprint of 4.07 million parameters suitable for low-power edge devices.

Ashraf et. al. (Ashraf et al., 2022) proposes an automated weapon detection system for real-time surveillance that utilizes the YOLO-v5s architecture combined with Gaussian blur preprocessing to minimize false negatives and improve detection efficiency. The results demonstrate that this framework is 19 times faster than existing Faster R-CNN models—achieving a speed of 0.010 seconds per frame—while maintaining high reliability with a recall rate of up to 99% on images and 94% on video.

Burnayev et. al. (Burnayev et al., n.d.) presents an autonomous weapon detection system that utilizes edge computing on a Raspberry Pi and the EfficientDet-Lite0 architecture to identify armed threats locally without requiring an internet connection. By integrating real-time audio notifications and visual overlays for augmented reality glasses, the system achieves an execution time of 1.48 seconds while significantly improving data privacy and reducing network bandwidth compared to cloud-based solutions.

Ahmad et. al. (Ahmad et al., 2021) presents an automatic weapon detection system for smart CCTV surveillance that utilizes the Scaled YOLOv4 algorithm and Convolutional Neural Networks to identify threats like knives and firearms in real-time. The framework is integrated with an interactive dashboard built on Node.js and MongoDB, achieving a high accuracy of 0.89 and a processing speed of 35 FPS to provide immediate security notifications.

Ahmed et. al. (Ahmed et al., 2022) introduces an enhanced real-time weapon detection system utilizing the Scaled-YOLOv4 algorithm, which achieves a high 92.1 mAP score and 85.7 FPS on high-performance GPUs. To facilitate practical deployment, the authors optimized the model using TensorRT for resource-constrained edge-computing devices like the Jetson Nano, providing a comparative analysis of performance across various hardware configurations.

Hnoohom et. al. (Hnoohom et al., 2022) presents an automatic weapon detection system for smart CCTV surveillance that utilizes the Scaled YOLOv4 algorithm and Convolutional Neural Networks to identify threats like knives and firearms in real-time. The framework is integrated with an interactive dashboard built on Node.js and MongoDB,

achieving a high accuracy of 0.89 and a processing speed of 35 FPS to provide immediate security notifications.

Research findings indicate that AI-Driven Decision Support Systems (AI-DSS) have a positive and significant impact on organizational performance. Statistical analysis indicates that implementing AI-DSS significantly improves organizational performance, as measured by three key indicators: operational efficiency, decision-making quality, and strategic performance achievement. Organizations utilizing AI-DSS are able to integrate and analyze large amounts of data in real time, resulting in faster, more accurate, and evidence-based decision-making processes. This strengthens the organization's ability to respond to changes in the dynamic business environment in the digital economy.

Furthermore, these findings confirm that AI-DSS functions not only as a technical tool but also as a strategic instrument for improving organizational performance. AI-based decision support systems assist management in identifying patterns, predicting risks, and objectively evaluating various decision alternatives. The impact is seen in the increased effectiveness of organizational strategies, optimized resource utilization, and the organization's ability to achieve short-term and long-term targets. Thus, the adoption of AI-DSS has proven to be a key factor in strengthening the competitiveness and sustainability of organizational performance in the digital economy. The results of these findings can be seen in Table 1.

CONCLUSIONS AND RECOMMENDATIONS

In conclusion, this research demonstrates that YOLO-based models play a vital role in advancing suspicious human activity recognition for intelligent environmental security systems due to their real-time performance and detection efficiency. The review highlights that continuous improvements in YOLO architectures have significantly enhanced accuracy and robustness, making them suitable for dynamic and complex surveillance environments. However, the analysis also reveals that YOLO alone is insufficient for comprehensive activity understanding and must be complemented with temporal modeling, tracking, and contextual analysis techniques. Addressing challenges related to data diversity,

environmental variability, computational limitations, and ethical concerns remains essential for practical deployment. Overall, this research concludes that YOLO-based frameworks, when integrated with complementary models and responsibly deployed, offer a promising foundation for next-generation intelligent security systems.

Future research in YOLO-based suspicious human activity recognition should focus on developing more holistic and context-aware frameworks that go beyond object detection to achieve deeper activity understanding. Integrating YOLO with advanced temporal models, such as transformers and graph-based networks, can improve the recognition of complex and long-duration human behaviors. Multi-modal data fusion—combining video with audio, thermal, depth, or sensor data—offers promising opportunities to enhance robustness under challenging conditions like low lighting or occlusion. Further research is also needed on lightweight and energy-efficient YOLO variants for edge and embedded deployment in large-scale surveillance systems. In addition, improving generalization through self-supervised learning, domain adaptation, and the use of synthetic data can reduce dependency on labeled datasets. Finally, future studies should address privacy preservation, fairness, and explainable AI to ensure that intelligent environmental security systems are trustworthy, transparent, and ethically deployed.

REFERENCES

Aas, A., Naveed, H., Asghar, J., Khaleel, S., Khanum, Z., & Noureen, T. (2025). Efficient and Sustainable Video Surveillance Using CNN-LSTM Model for Suspicious Activity Detection. *VFAST Transactions on Software Engineering*, *13*(1), 60–71. https://doi.org/10.21015/vtse.v13i1.2023

Ahmad, I., Dananjaya, D., Muhammad, A., Arghanie, A., Versantariqh, M. A., David, M., & Fatmawati, U. D. (2021). Sistem Deteksi Senjata Otomatis Menggunakan Deep Learning Berbasis CCTV Cerdas. In *Jurnal Sistem Cerdas*.

Ahmed, S., Bhatti, M. T., Khan, M. G., Lövström, B., & Shahid, M. (2022). Development and Optimization of Deep Learning Models for Weapon Detection in Surveillance Videos. *Applied Sciences (Switzerland)*, *12*(12). https://doi.org/10.3390/app12125772

Arshad, M. H., Bilal, M., & Gani, A. (2022). Human Activity Recognition: Review, Taxonomy and Open Challenges. In *Sensors* (Vol. 22, Issue 17). MDPI. https://doi.org/10.3390/s22176463

Ashraf, A. H., Imran, M., Qahtani, A. M., Alsufyani, A., Almutiry, O., Mahmood, A., Attique, M., & Habib, M. (2022). Weapons detection for security and video surveillance using CNN and YOLO-V5s. *Computers, Materials and Continua*, *70*(2), 2761–2775. https://doi.org/10.32604/cmc.2022.018785

Burnayev, Z. R., Toibazarov, D. O., Atanov, S. K., Canbolat, H., Seitbattalov, Z. Y., Kassenov, D. D., Kazakhstan -Elbasi, of, & -Elbasi, K. (n.d.). Weapons Detection System Based on Edge Computing and Computer Vision. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 14, Issue 5). www.ijacsa.thesai.org

Dekkati, S., Srujan Gutlapalli, S., Thaduri, U. R., Koteswara, V., & Ballamudi, R. (2022). AI and Machine Learning for Remote Suspicious Action Detection and Recognition. In *ABC Journal of Advanced Research* (Vol. 11, Issue 2).

Golande, R., Bhapkar, R., Nalawade, A., Rashinkar, A., & Mane, S. (2025). Weapon Detection System: Real-Time Object Recognition for Threat Detection. *International Journal for Research in Applied Science and Engineering Technology*, *13*(3), 3514–3521. https://doi.org/10.22214/ijraset.2025.68105

Golestani, N., & Moghaddam, M. (2020). Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-15086-2

Gupta, N., & Agarwal, B. B. (2023). Recognition of Suspicious Human Activity in Video Surveillance: A Review. In *Technology & Applied Science Research* (Vol. 13, Issue 2). www.etasr.com

Hnoohom, N., Chotivatunyu, P., & Jitpattanakul, A. (2022). ACF: An Armed CCTV Footage Dataset for Enhancing Weapon Detection. *Sensors*, *22*(19). https://doi.org/10.3390/s22197158

Indhumathi, J., Balasubramanian, M., & Balasaigayathri, B. (2023). Real-Time Video based Human Suspicious Activity Recognition with Transfer Learning for Deep Learning.

International Journal of Image, Graphics and Signal Processing, *15*(1), 47–62. https://doi.org/10.5815/ijigsp.2023.01.05

Jahan, S., Roknuzzaman, & Islam, M. R. (2024). *A Critical Analysis on Machine Learning Techniques for Video-based Human Activity Recognition of Surveillance Systems: A Review*. http://arxiv.org/abs/2409.00731

Mandalapu, V., Elluri, L., Vyas, P., & Roy, N. (2023). Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions. *IEEE Access*, *11*, 60153–60170. https://doi.org/10.1109/ACCESS.2023.3286344

Rehman, A., Saba, T., Khan, M. Z., Damaševičius, R., & Bahaj, S. A. (2022). Internet-of-Things-Based Suspicious Activity Recognition Using Multimodalities of Computer Vision for Smart City Security. *Security and Communication Networks*, *2022*. https://doi.org/10.1155/2022/8383461

Shanthi, P., & Manjula, V. (2025). A systematic review on CNN-YOLO techniques for face and weapon detection in crime prevention. In *Discover Computing* (Vol. 28, Issue 1). Springer Science and Business Media B.V. https://doi.org/10.1007/s10791-025-09715-x

Singh, V., Singh, S., & Gupta, P. (2020). Real-Time Anomaly Recognition Through CCTV Using Neural Networks. *Procedia Computer Science*, *173*, 254–263. https://doi.org/10.1016/j.procs.2020.06.030

Vijeikis, R., Raudonis, V., & Dervinis, G. (2022). Efficient Violence Detection in Surveillance. *Sensors*, *22*(6). https://doi.org/10.3390/s22062216

Yadav, P., Ghodke, M., Dhokane, V., & Chavan, S. (2023). Predict, Identify and Alert on Suspicious Activity by Multiple Zone. *International Journal for Research in Applied Science and Engineering Technology*, *11*(5), 3454–3457. https://doi.org/10.22214/ijraset.2023.52283

Yuan, L., Andrews, J., Mu, H., Vakil, A., Ewing, R., Blasch, E., & Li, J. (2022). Interpretable Passive Multi-Modal Sensor Fusion for Human Identification and Activity Recognition. *Sensors*, *22*(15). https://doi.org/10.3390/s22155787