# AI Enhanced Surveillance for Identifying and Recognizing Crowd Behavior A Survey

B.S. Prakash[1], B. Tamilsudar[2], B. Ananthi[3]

*[1,3]Assistant Professor, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamilnadu, India*

*[2]Student, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamilnadu, India*

***Abstract*—Deep learning techniques in AI-enhanced surveillance systems for identifying and recognizing crowd behavior. The paper reviews recent advancements in computer vision and neural network architectures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers, which are pivotal for tasks like crowd density estimation, activity recognition, and anomaly detection. Key topics include supervised and unsupervised learning approaches, transfer learning for domain adaptation, and multimodal data fusion for improving accuracy and robustness. The survey also highlights real-time processing capabilities enabled by advancements in hardware, such as GPUs and edge computing devices. Challenges such as data scarcity, model interpretability, and addressing biases in crowd behavior datasets are discussed. This survey aims to provide a detailed understanding of the field, supporting researchers and practitioners in the development of more effective and ethical crowd monitoring systems.**

***Index Terms*—Deep Learning, AI-enhanced Surveillance, Crowd Behavior Analysis, Anomaly Detection, Real-time Processing.**

## I. INTRODUCTION

Deep learning is part of machine learning, which itself is a branch of artificial intelligence (AI). It focuses on using multilayered neural networks (hence the name “deep”) to model and understand complex patterns and relationships in data. It draws inspiration from the human brain, where each layer of the network acts as a layer of information. In recent years, crowd management and behavior analysis have become critical components of public safety and security, especially in high-traffic areas such as stadiums, airports, shopping malls, and public events. As mass gatherings increase in frequency and scale, the potential risks associated with crowd-related incidents, such as stampedes, riots, and other emergencies, have escalated. These scenarios underscore the urgent need for robust and efficient crowd monitoring systems (CMS) that can ensure public safety while mitigating risks. Traditional surveillance techniques, such as closed-circuit television (CCTV) cameras, often fall short in terms of coverage, real-time processing, and the ability to detect and predict abnormal crowd behaviors. These limitations arise from the manual nature of monitoring, scalability issues, and the challenges of processing vast amounts of video data in real-time.

Artificial intelligence (AI), particularly through deep learning and computer vision technologies, has emerged as a transformative solution in crowd surveillance. AI-enabled systems can autonomously detect, track, and analyze crowds in real-time, offering unprecedented insights into crowd behavior, density, movement patterns, and potential risks. By leveraging advanced algorithms such as deep learning, convolutional neural networks (CNNs), and recurrent neural networks (RNNs), these systems can perform tasks that were previously impossible or inefficient with traditional methods. For instance, AI systems can not only count the number of individuals in a crowd with remarkable accuracy but also analyze their behavior to identify specific patterns indicative of potential issues. These include panic, aggression, unusual clustering, or erratic movement, which might signal emergencies or disturbances. Moreover, deep learning-based systems are capable of predictive analytics, allowing them to foresee potential risks before they escalate. For example, by analyzing historical data combined with real-time inputs, these systems can predict overcrowding, identify bottlenecks, and alert authorities to take preventive measures. Advanced techniques such as optical flow analysis and skeleton tracking enhance the system's ability to interpret complex crowd dynamics and individual behaviors within a collective setting. These insights are particularly valuable for large-scale event organizers and public safety officials who need to make informed, timely decisions to maintain order and prevent incidents. The integration of AI-powered surveillance systems into crowd management not only improves safety but also optimizes resource allocation, reduces human error, and enhances overall operational efficiency. By offering automated, scalable, and intelligent monitoring solutions, deep learning continues to shape the future of public safety, transforming the way we manage and understand crowd behavior in high-stakes environments.

## II.LITERATURE SURVEY

[1] Yangkai Wu the Abnormality Converging Scene Analysis Method (ACSAM) is a system designed to detect unusual group behavior in crowded environments using video or CCTV footage. By leveraging a Convolutional Neural Network (CNN) with specialized training layers, ACSAM identifies and classifies abnormal activities with high accuracy, even in dense crowds. The method extracts video frames, validates abnormal behavior by comparing current patterns to previous data, and improves accuracy through iterative training. Tested on 26 videos and trained on 34 samples, ACSAM outperformed other systems like Deep ROD, MSI-CNN, and PT-2DCNN, achieving 12.55% higher accuracy, 12.97% better recall, and a 10.23% faster convergence rate. This makes ACSAM a robust, real-time solution for detecting anomalous behavior in complex crowd scenarios.

[2] Monji Mohamed Zaidi to enhance surveillance and security systems. By using advanced deep learning methods, the research overcomes limitations of existing approaches, which often struggle with accuracy and efficiency. The method involves collecting and refining data, preprocessing images, and training models to detect suspicious behaviors. Convolutional Neural Networks (CNNs) and architectures like the time-distributed CNN and Conv3D models were used, achieving high accuracy rates of 90.14% and 88.23%, respectively, surpassing previous methods. The trained models were tested on unseen data and real-world scenarios, such as analyzing YouTube videos, demonstrating their ability to predict and recognize suspicious activities effectively. This approach enhances public safety by enabling more precise and efficient surveillance systems, reducing potential risks in various settings.

[3] P. Kuppusamy, the growing need for surveillance systems to monitor human behavior in various environments, addressing the challenges of manually reviewing long videos. Automated systems powered by Convolutional Neural Networks (CNNs), particularly 3D CNNs, have proven more effective than traditional methods for detecting abnormal behaviors. With the rise of real-time data from surveillance networks, advancements in deep learning and computing resources have improved the accuracy and efficiency of behavior recognition. The research compares different CNN models, exploring feature extraction, dataset variations, and technique limitations while emphasizing their potential to enhance surveillance and security systems.

[4] Yung-yao Chen, vices, to enhance efficiency in decentralized survey distributed real-time object detection framework for smart video surveillance in smart cities, addressing the limitations of traditional cloud-based systems that rely on constant internet connectivity. Using edge computing, data is processed locally on edge devices, enabling faster responses for latency-sensitive tasks while reducing reliance on the cloud. The cloud consolidates data from edge devices, derives global insights using AI, and shares this knowledge with the edge for real-time surveillance. The framework improves responsiveness, minimizes data transmission needs, and balances workloads, with experimental validation showing its effectiveness. Future work includes peer-to-peer workload sharing among edge deillance systems.

[5] Prof. M.S. Khan, a real-time AI-powered crowd surveillance system integrated with big data analytics to enhance urban security. By using advanced AI for live video analysis and edge computing for immediate processing, the system detects abnormal crowd behavior and provides real-time alerts for quick responses. Big data analytics enables trend analysis, predictive modeling, and resource

optimization using historical data. While offering improved threat detection and situational awareness, the system faces challenges like privacy concerns, legal compliance, and infrastructure requirements. Continuous refinement of AI models and balancing security with privacy are key to its success in creating safer urban spaces.

[6] Chaya Jadhav, An automated system for crowd monitoring and suspicious activity detection using advanced deep learning techniques, particularly Fully Convolutional Networks (FCN) and Long Short-Term Memory (LSTM) models. The proposed system addresses critical challenges associated with manual surveillance, such as labor-intensive monitoring, susceptibility to human error, and missed detections. The transformative potential of artificial intelligence in public safety and surveillance, demonstrating how the integration of cutting-edge algorithms can address real-world challenges effectively.

[7] A. Hussein Abnormal crowd detection and estimation are vital for public safety in video surveillance, especially to prevent stampedes. Traditional methods often struggle in dense and occluded areas, leading to inaccuracies. This study reviews recent advancements in identifying unusual crowd behaviors using innovative technologies like RFID, wireless sensor networks, Wi-Fi, and Bluetooth Low Energy. These rely on device-free algorithms that analyze signal strength variations to estimate crowd speed and direction, helping predict stampedes. It also examines mobile crowd sensing, edge computing, urban dynamics, optical flow, and machine learning.

[8] Ajay Kumar Crowd monitoring and behavior analysis are crucial in computer vision research due to their importance in safety and security. Over the past decade, many methods have been developed to estimate crowd size, predict future behaviors, and ensure better management. Despite these advances, there is still a need for real-time analysis, especially for unorganized crowds. existing methods for monitoring and analyzing both organized and unorganized crowds, covering traditional techniques and modern deep learning approaches. It also provides details on datasets used in these studies, highlighting their strengths and limitations. The goal is to help researchers understand the current state of the art and improve crowd analysis methods, particularly for challenging scenarios involving unorganized crowds.

[9] Dushyant Kumar Singh, The evolving role of automated video surveillance systems for public and private safety. Traditional methods, reliant on manual monitoring, are limited by human attention span, emphasizing the need for autonomous solutions. Modern systems integrate computer vision and machine learning techniques, such as background subtraction, Histogram of Oriented Gradients (HOG), and Support Vector Machines (SVM), to detect anomalies like prohibited area entry and crowd violence in real-time. Techniques like optical flow and Violent Flows (ViF) descriptors enable efficient motion analysis in dynamic settings. Communication modules facilitate swift alerts to police using real-time data and GUI integrations like Google Maps. Despite advancements, challenges remain in handling varied environmental conditions, bandwidth demands, and privacy concerns. Nonetheless, the integration of these technologies promises enhanced security and smarter policing for urban surveillance.

[10] Ali m. Al-shaery AI technology has rapidly advanced and is now widely applied in commercial areas. Intelligent surveillance video analysis is used to predict customer preferences, optimize product placement, and manage advertisements by gathering customer data. This approach uses deep learning techniques to address tasks like object detection, tracking, and human identification. To handle challenges like occlusion, skeleton recognition algorithms are used instead of traditional object detection methods. Additionally, human re-identification (ReID) and multi-human tracking algorithms enable accurate tracking and counting. The results, including density maps and statistical data, help businesses evaluate customer behavior and adjust strategies for better outcomes.

[11] Vishakha L. Bansod, Crowd analysis theory revolves around understanding, monitoring, and predicting the behavior of large groups of people to ensure safety and mitigate risks in public spaces. The core concepts in this field typically include: Crowd Dynamics: This involves the study of how people move and interact in crowds. Understanding crowd dynamics helps predict potential risks such as stampedes, congestion, or violence. Models focus on how individual actions and collective behavior influence the overall crowd movement. Crowd

Attributes: Crowd Counting: Estimating the number of people in a given area to assess the risk of overcrowding. Density Estimation: Analyzing how tightly packed the crowd is, which directly correlates with safety risks (higher density often leads to greater danger).Motion Detection: Identifying unusual or abnormal movement patterns, which can indicate panic or dangerous behavior. Behavior Analysis: Understanding patterns in crowd behavior, such as group formations, movements, and reactions to stimuli, to predict and manage potential hazards. Abnormal behaviors can signal the need for intervention to prevent incidents. Crowd Tracking: This involves following individuals or groups in a crowd using surveillance systems. Effective tracking allows for early detection of issues and helps authorities respond quickly. Deep Learning Models: Advanced techniques such as Convolutional Neural Networks (CNNs) for feature extraction and Recurrent Neural Networks (RNNs) for modeling temporal behavior have enhanced the accuracy of crowd analysis. CNNs are used to identify spatial patterns (e.g., crowd density and movement), while RNNs help predict future behaviors by analyzing time-based sequences. The combination of these theories, supported by video surveillance, aims to improve crowd safety by enabling the detection and understanding of abnormal crowd behavior in real-time. This enables timely interventions to prevent accidents or crises.

[12]Dohun Kim, Heegwang Kim, a real-time surveillance system for abnormal behavior detection in CCTV environments, addressing the computational limitations of deep learning methods. [1] The system combines pedestrian detection and tracking to extract real-time pedestrian information. It identifies abnormal behaviors like intrusion, loitering, falls, and violence through a two-step process: location-based analysis for intrusion/loitering and behavior pattern analysis for falls/violence. The system uses periodic detection coupled with tracking to improve efficiency. Cropped pedestrian images are analyzed by a dedicated module for fall and violence detection. The system incorporates functions for information management and inter-module communication. [2] Performance was evaluated using a KISA dataset and VLC streaming. Results show high accuracy for intrusion, loitering, and violence detection. Fall detection performance was slightly lower but is expected to improve with enhanced algorithms and datasets. The proposed method offers a practical solution for real-time abnormal behavior detection in CCTV systems.

[13] Nidhi Shetty, a crowd management system using computer vision techniques on a Raspberry Pi 3. The system aims to address overcrowding issues in public gatherings by detecting and counting people in a scene. It employs OpenCV-Python and a Haar cascade classifier trained for human head detection. Human tracking is implemented by analyzing the direction of movement. The system utilizes a Raspberry Pi 3 with an ARMv8 CPU for processing. Head detection is performed using Haar features and the Adaboost algorithm. Optical flow is used for tracking individuals within the video feed. Testing was conducted using video footage from the authors' institution. Increased training data improved the efficiency of the system. The proposed method has potential applications in surveillance and crowd control scenarios

[14] B.Ganga, deep learning's impact on object detection (OD), highlighting its superiority over traditional methods relying on handcrafted features. Deep learning models directly learn features from data, effectively addressing spatiotemporal challenges for enhanced object recognition in images and videos. The survey categorizes deep learning-based OD algorithms into two-stage (e.g., Faster R-CNN) and one-stage (e.g., YOLO, SSD) methods, all built upon Convolutional Neural Networks (CNNs). It explores diverse OD applications, with a strong emphasis on crowd analysis, a field benefiting significantly from these advancements. A review of existing literature reveals that CNNs are the dominant architecture in OD research (28% of papers). Within crowd analysis research, efforts are distributed across counting (24%), categorization (25%), individual behavior analysis (25%), and other related areas (27%). The survey concludes that deep learning has significantly advanced OD capabilities, providing effective solutions for various applications, particularly in complex crowd analysis scenarios. It provides a valuable overview of related review papers and specific deep learning-based OD algorithms and their optimization strategies.

[15] N Fadzil1, crowd monitoring systems and technologies, focusing on their application in preventing the spread of COVID-19, particularly in

airport terminals. It addresses the importance of public social distancing and the need for effective crowd management in confined spaces. The report reviews commercially available and developing crowd monitoring products from various countries, including Malaysia, China, Korea, Japan, and Europe. It categorizes crowd monitoring methods into counting, localization, and behavior analysis. The study emphasizes non-contact sensor technologies, including thermal imaging for temperature measurement and video camera imaging (using a Microsoft Kinect) for crowd counting. The Kuala Lumpur International Airport (KLIA) is presented as a potential case study location. The report highlights the need for efficient crowd monitoring solutions capable of handling large data volumes and providing real-time analysis. The reviewed technologies aim to provide online monitoring of crowds based on non-contact sensors for movement counting to mitigate COVID-19 outbreaks. It presents a comparison between different product and technology approaches. The study emphasizes the growing importance of crowd monitoring and the need for further development to meet current public health needs. The use of thermal imaging and video analysis as non-contact methods for public health monitoring is a key focus. The paper contributes a valuable survey of crowd monitoring products and technologies in the context of pandemic control.

## III PERFORMANCE METRICS

| Model/Method | Accuracy | Dataset Used | Applications |
|---|---|---|---|
| Convolutional Neural Network (CNN) | Real-time object detection 12.55% higher accuracy, 12.97% | Custom Dataset (26 videos, 34 samples) | Abnormal crowd behavior detection |
| Time-distributed CNN, Conv3D | Accuracy: 90.14% (Time-dist.), 88.23% (Conv3D) | YouTube videos, custom suspicious activity dataset | Suspicious human activity recognition |
| 3D CNNs | Achieved 12.55% higher accuracy, Deep ROD had an accuracy of 85%. | Surveillance network data | Real-time abnormal behavior detection |
| Distributed real-time object detection | Accuracy of 85%, Decentralized System Accuracy 93.5%. | Edge computing simulation | Smart city surveillance |
| Big data analytics, AI-based models | Accuracy of 95%, New System Accuracy=85%. | Historical and live surveillance data | Urban security and crowd behavior |
| FCN, LSTM | Accuracy: 92.3%<br>Precision: 90.7%<br>Recall: 89.5%<br>F1-Score: 90.1% | Public crowd datasets | Suspicious activity detection |
| RFID, WSN, Wi-Fi, BLE | Accuracy: 85%–92%<br>Detection: 88%–94% | Signal strength datasets | Stampede prevention |
| Optical flow, ML methods | Abnormal Behavior Detection: 75%–90% accuracy | Various crowd datasets | Organized and unorganized crowd analysis |
| ML/DL Histogram of Oriented Gradients (HOG), SVM. | Accuracy: 75%–90%<br>Controlled Environments: Traditional Methods: 75%–90%<br>Modern Techniques: 90%–97% | UMN Dataset, Violence Flow Dataset (ViF) | Anomaly Detection in Crowds, Crowd Flow and Motion Analysis. |

| | | | |
|---|---|---|---|
| Skeleton recognition, ReID | Accuracy: 85%–95% Human Re-identification (ReID) Accuracy: 85%–95% | Real store surveillance videos | Customer behavior and business analytics |
| Crowd analysis theory (CNNs, RNNs) | Accuracy through deep learning 90%–97% | Video surveillance | Real-time detection of abnormal crowd behavior |
| Pedestrian detection & tracking, location & behavior pattern analysis | High accuracy for intrusion, loitering, and violence; lower for fall detection accuracy of 85%. | KISA dataset, VLC streaming | Real-time abnormal behavior detection in CCTV environments (intrusion, loitering, falls, violence) |
| Haar cascade classifier (head detection), optical flow (tracking) | Improved efficiency with increased training data | Video footage from author's institution | Crowd management (people counting and tracking) on Raspberry Pi 3 |
| Deep learning-based object detection (Faster R-CNN, YOLO, SSD) | Superior to traditional methods, not specifically quantified | Not specified, review of existing literature | Various object detection applications, with emphasis on crowd analysis (counting, categorization, behavior analysis) |
| Counting, localization, and behavior analysis methods using non-contact sensors (thermal imaging, video camera imaging) | real-time analysis and handling large data volumes Accuracy: 85%–95% | Various commercially available and developing products, KLIA as potential case study | Crowd monitoring for COVID-19 prevention in airport terminals |

## IV. ANALYSIS

Various models used for surveillance and crowd analysis. CNNs and 3D CNNs showed up to 12.55% higher accuracy for abnormal behavior detection. Time-distributed CNNs and Conv3D achieved ~90% accuracy for suspicious activity recognition, while FCN and LSTM reached 92.3% accuracy in public crowd datasets. Distributed systems improved smart city surveillance with 93.5% accuracy, and technologies like RFID and Wi-Fi achieved 85%–94% accuracy for stampede prevention. Optical Flow and HOG-SVM analyzed crowd motion with 75%–97% accuracy. Big data and AI models reached 95% accuracy for urban security, while skeleton recognition and ReID achieved 85%–95% for customer behavior analysis.

## V. CONCLUSION

The transformative impact of deep learning and AI-driven surveillance systems on public safety, urban security, and commercial applications. By leveraging advanced architectures like Convolutional Neural Networks (cnns), Fully Convolutional Networks (fcns), and Long Short-Term Memory (LSTM) models, these systems effectively detect abnormal behaviors and suspicious activities in real time. This research highlights the superiority of AI methods over traditional techniques, particularly in handling dense crowds, occlusions, and unorganized crowd scenarios. Moreover, edge computing and big data analytics enhance system responsiveness and provide predictive insights while addressing latency and data transmission challenges. However, the studies also underscore the importance of balancing security with privacy and legal compliance. As AI models and computational resources continue to evolve, these technologies will become increasingly vital for efficient surveillance, crowd management, and customer behavior analysis, setting a benchmark for future advancements in intelligent monitoring systems.

## VI. REFERENCES


[1] Yangkai Wu, Luhua Qiu, Jinming Wang, Shujuan Feng, The use of convolutional neural networks for abnormal behavior recognition in crowd scenes, Information Processing and Management 62 103880, 2024, https://doi.org/10.1016/j.ipm.2024.103880.

[2] Monji mohamed Zaidi, Gabriel Avellino sampedro, ahmad almadhor shtwai alsubai, abdullah al hejaili, michal gregus, Suspicious Human Activity Recognition from Surveillance Videos Using Deep Learning, in Proc. IEEE Int., VOLUME 12, 2024, DigitalObject.Identifier,10.1109/ACCESS.2024. 3436653.https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10620187.

[3] P. Kuppusamy, V.C. Bharat, Human abnormal behavior detection using cnns in crowded Uncrowded surveillance, Measurement: Sensors, Measurement:24(2022100510.2024https://doi.org/10.1016/j.measen.2022.100510.

[4] Yung-yao chen, yu-hsiu lien Yu-chen hu chih-hsien, Distributed Real-Time Object Detection Based on Edge-Cloud Collaboration for Smart Video Surveillance Applications, in Proc. IEEE Int., VOLUME 10, 2022, Digital Object Identifier10.1109/ACCESS.2022.3203053,https://ieeexplore.ieee.org/document/980797.

[5] Chaya Jadhav, Rashmi Ramteke, Rachna K. Somkunwar, Smart Crowd Monitoring and Suspicious Behavior Detection Using Deep Learning, International information engineering association Vol. 37, No. 4, August, 2023, https://iieta.org/journals/ria/paper/10.18280/ria.370416

[6] Prof. M.S. Khan, Bhalerao Chaitrali, Bangar Gaurav, Kharat Sakshi, Pingale Yukta, Real Time AI Enhanced Crowd Surveillance International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences, Volume 11, Issue 6, Nove-Dec 2023, https://www.ijirmps.org/researchpaper.php?id=230623.

[7] A. HHussein, M.W. Raed , A. Al-Shaikhi , M. Mohandas , B. Liu, Crowd anomaly estimation and detection, Franklin Open, Volume 8, September 2024, 100169, https://doi.org/10.1016/j.fraope.2024.100169.

[8] Ajay Kumar Arunnehru J, Crowd Behavior Monitoring and Analysis in Surveillance, Turkish Journal of Computer and Mathematics Education Vol.12 No. 7 (2021), 2322-2336, https://turcomat.org/index.php/turkbilmat/article/view/3441.

[9] Dushyant Kumar Singh, Sumit Paroothi, Mayank Kumar Rusia, Mohd. Aquib Ansari,

[10] Human Crowd Detection for City Wide Surveillance, Science Direct, Procedia Computer Science 171 (2020) 350–359, https://www.sciencedirect.com/science/article/pii/S1877050920310012 .

[11] Ali m. Al-shaery, shroug s. Alshehri, norah s. Farooqi, and mohamed o. Khozium, monitor and manage crowd, in Proc. IEEE, 2, 2020,_**DOI:** 10.1109/ACCESS.2020.3038334,https://ieeexplore.ieee.org/document/9260138.

[12] Vishakha L. Bansod, Asha Ambhaikar , Crowd Analysis System for Images of CCTV Camera, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-5, January 2020. https://www.ijrte.org/wpcontent/uploads/papers/v8i5/E6283018520.

[13] Dohun KKim,Heegwang Kim, Yeongheon Mok, Joonki Paik, Real-Time Surveillance System for Analyzing Abnormal Behavior of Pedestrians,. 2021, *11*(13),6153; https://doi.org/10.3390/app11136153

[14] Nidhi Shetty,2Pooja Sankpal,3Priyanka Shripad,4Sandip Zade., CROWD DETECTION AND MANAGEMENT IN SURVEILLANCE SYSTEM, International Journal of Creative Research Thoughts (IJCRT), © 2020 IJCRT |

Volume 8, Issue 3 March 2020 | ISSN: 2320-2882. https://ijcrt.org/papers/IJCRT2003013.

[15] B.GANGA ,LATA B.T, VENUGOPAL K.R., Object Detection And Crowd Analysis Using Deep Learning Techniques, Neurocomputing Volume 597, 7 September 2024, 12793, Https://Www.Sciencedirect.Com/Science/Article/Abs/Pii/S0925231224007033.

[16] N Fadzil1, N H Abu Bakar1, S M Idrus1*, A I Azmi1, S H Mohammad1 and N. Ali2, Recent Development of Crowd Monitoring Technology Solution for Covid-19 Prevention at Airport Terminal, International Journal of Nanoelectronics and Materials Volume 14 (Special Issue) December 2021[227236],https://ijneam.unimap.edu.my/images/PDF/InCAPE2021/Vol_14_SI_Dec_2021_227-236.