iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71063

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

AI Enhanced Surveillance for Identifying and Recognizing Crowd Behavior

B. S. Prakash[1], B. Tamilsudar[2], T. Viswanath Kani[3]
[1]Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamilnadu, India
[2]PG Scholar, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamilnadu, India
[3]Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamilnadu, India

Abstract: The application of deep learning methods in AI-powered surveillance systems for detecting and interpreting crowd behavior. It reviews the latest progress in computer vision and neural network architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers—essential for tasks such as estimating crowd density, recognizing activities, and detecting anomalies. The discussion covers both supervised and unsupervised learning strategies, domain adaptation through transfer learning, and the integration of multiple data sources to enhance accuracy and system resilience. It also emphasizes the impact of hardware advancements, like GPUs and edge computing, in enabling real-time analysis. Furthermore, the paper addresses key challenges, including limited data availability, lack of model transparency, and inherent biases in crowd behavior datasets. This comprehensive survey aims to deepen understanding in the field and support the development of more efficient, reliable, and ethically sound crowd monitoring solutions
Keywords: Deep Learning, AI-enhanced Surveillance, Crowd Behavior Analysis, Anomaly Detection, Real-time Processing.

I. INTRODUCTION

Deep learning is a subset of machine learning, which falls under the umbrella of artificial intelligence (AI). It focuses on using multi-layered 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 times, analyzing crowd behavior and managing large gatherings have emerged as essential aspects of ensuring public safety and security, particularly in densely populated locations like stadiums, airports, malls, and major 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 situations highlight the critical need for strong and effective crowd monitoring systems (CMS) that can guarantee public safety while reducing potential 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-driven systems can independently identify, monitor, and analyze crowds in real time, providing unmatched insights into crowd behavior, density, movement trends, and possible 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. Integrating AI-driven surveillance systems into crowd management enhances safety, streamlines resource utilization, minimizes human error, and boosts overall operational effectiveness. 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 introduced the Abnormality Converging Scene Analysis Method (ACSAM), a system developed to identify abnormal group behavior in crowded settings by analyzing video or CCTV footage. Using a Convolutional Neural Network (CNN) with customized training layers, ACSAM effectively detects and classifies abnormal behaviors with high precision, even in crowded environments. The method captures video frames, detects abnormal behavior by comparing current patterns with historical data, and enhances accuracy through continuous training. ACSAM, tested on 26 videos and trained with 34 samples, outperformed systems like Deep ROD, MSI-CNN, and PT-2DCNN, delivering 12.55% higher accuracy, 12.97% improved recall, and a 10.23% faster convergence rate. As a result, ACSAM stands as a powerful, real-time tool for detecting abnormal behavior in complex crowd environments.

[2] Monji Mohamed Zaidi sought 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, pre-processing 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 highlighted the increasing demand for surveillance systems to observe human behavior across different settings, tackling the difficulties associated with manually analyzing lengthy video footage. 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 proposed improvements in a decentralized, real-time object detection framework for smart video surveillance in smart cities, aiming to enhance efficiency and overcome the limitations of traditional cloud-based systems. 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 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. Big data analytics enables trend analysis, predictive modeling, and resource 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 developed an automated system for crowd monitoring and detecting suspicious activities by leveraging advanced deep learning methods, specifically Fully Convolutional Networks (FCN) and Long Short-Term Memory (LSTM) models. The proposed system addresses critical challenges associated with manual surveillance, such as labour-intensive monitoring, susceptibility to human error, and missed detections. The transformative power of artificial intelligence in public safety and surveillance lies in its ability to integrate advanced algorithms to tackle real-world challenges efficiently.

[7] As noted by A. Hussein , detecting and estimating abnormal crowd behavior is crucial in video surveillance to enhance public safety and prevent incidents such as stampedes. Traditional methods often struggle in dense and occluded areas, leading to inaccuracies. This study investigates recent progress in recognizing unusual crowd behaviors using advanced technologies like RFID, wireless sensor networks, Wi-Fi, and Bluetooth Low Energy. These technologies utilize device-free algorithms that interpret changes in signal strength to estimate crowd movement and direction, aiding in prediction efforts stampedes. It also explores mobile crowd sensing, edge computing, urban dynamics, optical flow, and the application of machine learning.

[8] Crowd monitoring and behavior analysis play a vital role in computer vision research because of their significance in ensuring 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 developments, real-time analysis remains necessary, particularly for unstructured crowds.

It discusses current methods for monitoring and analyzing both organized and unorganized crowds, encompassing conventional techniques as well as advanced deep learning approaches.It also outlines the datasets employed in these studies, emphasizing their advantages and drawbacks. The aim is to support researchers in grasping recent innovations and refining crowd analysis approaches. especially for complex situations involving unstructured crowds.

[9] Dushyant Kumar Singh highlights the growing importance of automated video surveillance systems in enhancing security across public and private spaces, noting that conventional methods dependent on The limitations of The brevity of human attention spans necessitates the development and implementation of 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. Methods such as optical flow and Violent Flows (ViF) descriptors allow effective motion analysis in dynamic environments. Communication modules enable rapid alerts to law enforcement through real-time data and GUI integrations, such as 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. Smart video surveillance analysis is utilized to collect customer data, enabling the prediction of preferences, optimization of product placement, and effective advertisement management. 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. Moreover, human re-identification (ReID) and multi-person tracking algorithms facilitate precise tracking and counting of individuals. The results, including density maps and statistical data, help businesses evaluate customer behavior and adjust strategies for better outcomes.

[11] Vishakha L. The theory of crowd analysis focuses on comprehending, observing, and forecasting the actions of large groups of individuals to maintain safety and reduce risks in public areas. The core concepts in this field typically include: Crowd Dynamics: This involves the study of how people move and interact in crowds. Comprehending crowd dynamics enables the prediction of possible hazards like stampedes, overcrowding, or violent outbreaks. Models emphasize the impact of both individual actions and group behavior on the overall dynamics of crowd movement. Crowd Counting: Determining the number of individuals in a specific area to evaluate potential risks of overcrowding. Density Estimation: Assessing how closely individuals are grouped together, as higher crowd density typically increases safety risks. Motion Detection: Recognizing irregular or unexpected movement patterns within the crowd, patterns, which can indicate panic or dangerous behavior. Behavior Analysis: Analyzing crowd behavior patterns—including group dynamics, movement trends, and responses to external factors—to anticipate and mitigate possible risks. 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 modelling temporal behavior have enhanced the accuracy of crowd analysis. CNNs are employed to detect spatial features like crowd density and movement, whereas RNNs are utilized to analyze temporal sequences for forecasting future behaviors. By integrating these theories with video surveillance, the goal is to enhance crowd safety through the real-time identification and interpretation of unusual crowd behavior. This enables timely interventions to prevent accidents or crises.

[12] Dohun Kim and Heegwang Kim developed a real-time surveillance system for detecting abnormal behavior in CCTV settings, tackling the computational challenges associated with deep learning techniques.1 The system combines Pedestrian detection and tracking are used to obtain real-time information about individuals on the move. It detects irregular activities such as intrusion, loitering, falls, and violence using a two-phase approach: one step focuses on location-based analysis to spot intrusion and loitering, while the other evaluates behavior patterns to identify falls and violent incidents. 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 developed a crowd control system based on computer vision methods running on a Raspberry Pi The system is designed to tackle overcrowding in public events by detecting and counting individuals in the crowd. It employs OpenCV-Python and a Haar cascade classifier trained for human head detection. Human tracking is achieved by examining movement direction, with

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

processing handled by a Raspberry Pi 3 equipped with an ARMv8 CPU. Head detection is carried out using Haar-like features in combination with the Adaboost algorithm. Optical flow is employed to track people within the video feed.Testing was conducted using video footage from the authors' institution. Increased training data improved the efficiency of the system. The suggested approach holds potential for use in surveillance and crowd management situations.

[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). In crowd analysis research, work is divided among counting (24%), categorization (25%), analysis of individual behavior (25%), and various other related domains (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. Fadzil explores crowd monitoring systems and technologies, with an emphasis on their role in controlling the spread of COVID-19, especially 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 classifies crowd monitoring techniques into categories such as 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. Kuala Lumpur International Airport (KLIA) has been suggested as a suitable location for a case study. The report emphasizes the importance of effective crowd monitoring solutions that can manage large data sets and deliver real-time analysis. The technologies reviewed are designed to enable online crowd monitoring using non-contact sensors for movement counting, helping to prevent COVID-19 outbreaks. It provides a comparison of various products and technological 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. This paper offers a valuable review of crowd monitoring technologies and solutions within the context of pandemic response. The key contributions of this study include: The creation of a secure, all-encompassing smart surveillance system intended to monitor crowd behavior during times of unrest and major public events. The development of a three-tier architecture that supports real-time data collection at the edge, secure communication in the middle layer, and AI-powered crowd behavior analysis at the core, enabling intelligent and real-time surveillance. The creation of a unique video dataset and the classification of four specific crowd behavior types, based on parameters such as crowd density and presence of violence. A deep learning model, trained using the Swin Transformer, was employed for automatic crowd behavior detection. To validate the system’s practical application, experiments were carried out using the Deep Stream SDK, confirming its suitability for real-world surveillance scenarios.

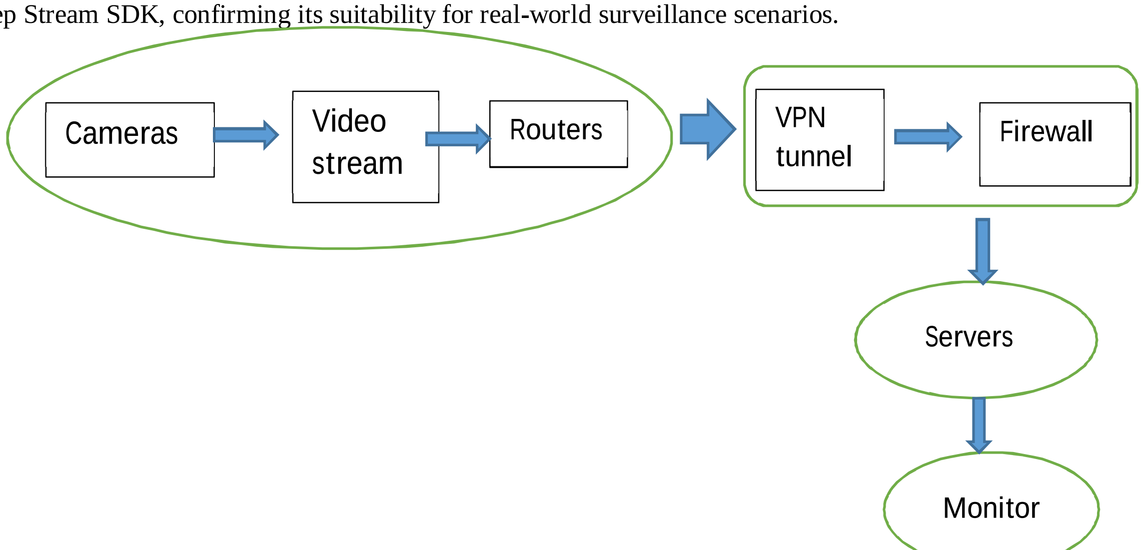


Figure 1: General Infrastructure of the Proposed Surveillance System

III. RELATED WORKS

Smart surveillance systems demand the proactive recognition and detection of events to avoid mishaps and disasters. Over the past decade, the increase in disasters and large-scale incidents during protests has prompted researchers to analyze surveillance video data using AI approaches. Besides, this led to the creation of datasets for research purposes. This section provides an outlook on the advances in video analysis and gives an awareness of existing surveillance systems and datasets

A) Surveillance Systems Surveillance systems have evolved significantly, from basic video recording to sophisticated systems with intelligent analysis capabilities. Early systems primarily served as passive recording devices, requiring manual review of footage. Modern surveillance systems, however, are increasingly incorporating AI to automate monitoring and analysis.

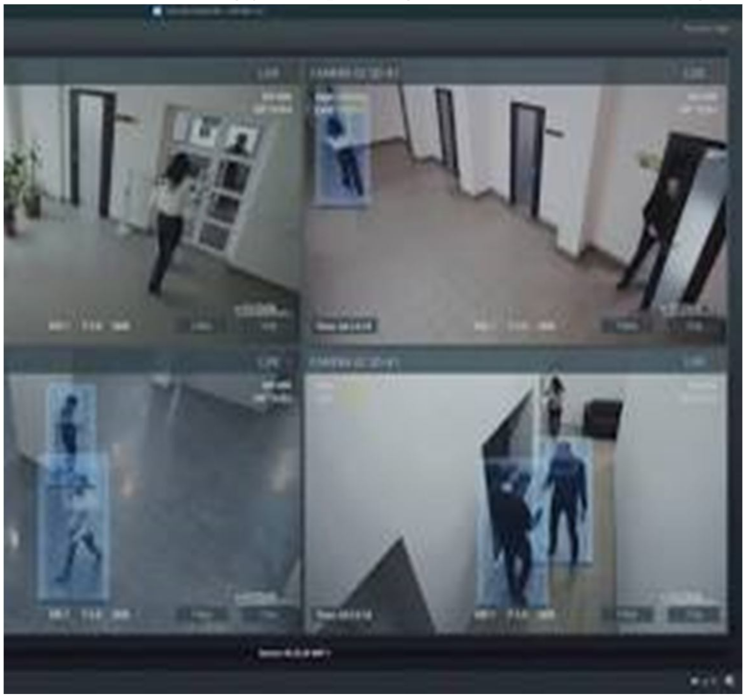


Figure 2: Sample Frames for Surveillance Camera Detection

B) Advances in Video Analysis Deep learning has significantly advanced the field of video analysis, enabling machines to understand video content with unprecedented accuracy. Key advancements include Convolutional Neural Networks (CNNs): CNNs excel at extracting spatial features from video frames, making them ideal for object detection and image classification. Recurrent Neural Networks (RNNs): RNNs, particularly Long Short-Term Memory (LSTM) networks, are designed to process sequential data, such as video frames, and capture temporal dependencies.3D Convolutional Neural Networks (C3D) C3D networks extend 2D convolutions to the temporal dimension, allowing them to learn spatio-temporal features directly from video clips. Object Detection Models: Models like YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), and Faster R-CNN have revolutionized object detection, enabling real-time identification of objects within video frames.

C) Existing Systems Existing crowd behavior analysis systems vary in their capabilities. Some systems rely on traditional computer vision techniques, such as optical flow and background subtraction, which can be limited in complex scenarios. Deep learning-based systems have shown superior performance, but challenges remain in areas such as Real-time processing of high-resolution video, Handling occlusions and varying crowd densities, Accurate detection of subtle behavioral cues, Generalizing to diverse crowd scenarios. Existing crowd behavior analysis systems vary in their capabilities. Some systems rely on traditional computer vision techniques, such as optical flow and background subtraction, which can be limited in complex scenarios. Deep learning-based systems have shown superior performance, but challenges remain in areas such as: Real-time processing of high-resolution video is essential for analyzing dynamic environments, allowing for the capture and interpretation of fast-moving events. It also involves handling occlusions and varying crowd densities, which can pose challenges in accurately identifying individuals or specific behaviors. The ability to detect subtle behavioral cues is crucial for understanding crowd dynamics, as small changes in movement or interaction can convey significant information. Furthermore, the system must be capable of generalizing to diverse crowd scenarios, ensuring its effectiveness across different settings and situations without requiring extensive retraining or adaptation.

Table 1: Existing Action Recognition and Surveillance Crowd Dataset

DATASET NAME	CLASSES	CARDINALITY (NUMBER OF CLASSES) - SIZE/NUMBER OF SAMPLES (AS MENTIONED)
KTH	Walking, Jogging, Running, Boxing, Hand Waving, Hand Clapping	6
Weizmann	Walking, Running, Jumping, Bending, Hand Waving, Jack Jumping, Pacing, Skipping, Side Gallop, Walking on the spot	10
HMDB51	51 action categories	51 - 6,849 samples
UCF101	101 human action categories	101 - 13,320 videos
Kinetics (400/600/700)	Human actions (singular, person-person, person-object)	400+ (Kinetics 400) - 400+ clips per class (Kinetics 400)
JHMDB	21 activities	21 - 923 videos
Hollywood-2	12 action & 10 scene classes	22 - 3,669 video clips
(Unspecified ACSAM Test)	Abnormal vs. Normal Group Behavior	2 - 26 videos
(Unspecified ACSAM Train)	Abnormal vs. Normal Group Behavior	2 - 34 samples
KISA dataset	Intrusion, Loitering, Falls, Violence	4

IV. PROPOSED DESIGN OF SURVEILLANCE SYSTEM

The proposed AI-enhanced surveillance system comprises several key components that work together to acquire, process, analyze, and interpret crowd behavior in video streams. The overall system architecture is designed to be modular, scalable, and adaptable to different environments and requirements.

1) System Model and Design Video Acquisition: Video data is captured by surveillance cameras strategically positioned to monitor crowd activity. Pre-processing: The raw video footage is pre-processed to enhance its quality and prepare it for analysis. This may include frame resizing, noise reduction, and stabilization. Object Detection and Tracking: Deep learning-based object detection models are used to detect individuals within the video frames. Tracking algorithms then maintain the identity of each individual across successive frames, enabling the analysis of their movement. Behavior Analysis and Recognition: This is the core of the system, where deep learning models classify crowd behavior into predefined categories: natural, fight, large peaceful gathering, large violent gathering, and weapon detection. Alert and Reporting: When the system detects a behavior of interest (e.g., fight, large violent gathering, weapon detection), it generates an alert, providing relevant information such as the location, time, and type of behavior.
2) System Integration The system is built to easily integrate with current surveillance setups. It can be deployed on a distributed architecture, with video processing performed on edge devices or on a central server. The system can be combined with other security platforms to deliver a complete security solution.

V. EXPERIMENT AND ANALYSIS

Experiment and Analysis "AI Enhanced Surveillance for Identifying and Recognizing Crowd Behavior," experiments and analysis related to "natural, fight, large peaceful gathering, large violent gathering, weapon detection" would be crucial for developing and evaluating such systems. Here's a breakdown of potential experiments and the types of analysis involved for each of these categories. Real-world Surveillance Footage: Existing CCTV footage from public spaces, events, and potentially controlled environments with appropriate ethical considerations and permissions. Simulated Data Creating synthetic video data with actors simulating different crowd behaviors and scenarios. This enables structured experiments and the creation of labeled datasets for training AI models. Publicly Available Datasets Utilizing existing datasets focused on human action recognition, crowd behavior, or anomaly detection, if relevant.

A. Natural Crowd Behavior

1) Experiment: Observation of Unscripted Crowds: Analyze video footage of everyday crowd interactions in various settings e.g., shopping malls, public transport hubs, pedestrian walkways. Density Variation Studies: Track changes in crowd density over time and space. Flow Analysis: Study the movement patterns and directions of individuals and groups within the crowd. Interaction

2) Analysis: Observe and categorize typical interactions between individuals (e.g., walking together, brief conversations). Analysis: Statistical Analysis: Calculate metrics like average crowd density, flow rate, speed of movement, and frequency of different interaction types. Pattern Recognition: Use AI algorithms e.g., clustering, time series analysis to identify common and recurring patterns in crowd movement and behavior. Anomaly Detection: Train models to identify deviations from these "natural" patterns, which could indicate unusual events. Social Force Modeling: Explore computational models that simulate crowd behavior based on individual goals and interactions.

B. *Fight Detection*

1) Experiment: Simulated Fights: Stage controlled scenarios with actors engaging in various forms of physical altercations (e.g., pushing, shoving, punching, kicking). Vary lighting conditions, crowd density, and camera angles. Analysis of Real Fight Footage: If ethically and legally permissible, analyze real-world footage of fights.
2) Analysis: Motion Analysis: Focus on detecting rapid and erratic movements, sudden changes in body posture, and close physical contact between individuals. Pose Estimation: Track the movement of key body joints to identify aggressive postures and actions. Optical Flow Analysis: Analyze the motion vectors of pixels to detect sudden and chaotic movements characteristic of fighting. Machine Learning Classification: Train AI models (e.g., CNNs, RNNs, Transformers) to classify video segments as containing a fight or not, based on the extracted features. Temporal Analysis: Model the temporal evolution of fighting behavior to improve detection accuracy.

C. *Large Peaceful Gathering Detection*

1) Experiment: Analysis of Footage from Peaceful Events: Study videos of organized peaceful gatherings like parades, protests (without violence), festivals, and public celebrations. Varying Crowd Sizes and Densities: Analyze gatherings of different scales. Activity Recognition: Identify common activities within peaceful gatherings (e.g., walking, standing, chanting, clapping).
2) Analysis: Density Estimation: Precisely determine the crowd size. Flow and Movement Patterns: Analyze the generally organized and directed movement of the crowd. Activity Distribution: Detect how various activities are spread across space and time. Machine Learning Classification: Train models to distinguish peaceful gatherings from other crowd scenarios based on features like density, movement coherence, and the absence of violent actions. Contextual Analysis: Incorporate contextual information (e.g., location, time, event type) to aid in identification.

D. *Large Violent Gathering Detection*

1) Experiment: Analysis of Footage from Violent Events: Study videos of riots, violent protests, and other instances of large-scale aggression. Simulated Violent Scenarios: Create controlled simulations of violent crowd behavior with ethical considerations.
2) Analysis: Detection of Aggressive Actions: Identify behaviors like running, throwing objects, vandalism, and physical assaults Anomaly Detection at Scale: Identify widespread deviations from normal crowd behavior. Spatiotemporal Pattern Analysis: Examine how violent clusters emerge and spread within the crowd. Machine Learning Classification: Train models to classify video segments or entire events as violent gatherings based on the presence and intensity of aggressive actions and chaotic movement Multi-Agent Modeling: Potentially use agent-based models to understand the dynamics of how violence can spread within a crowd.

E. *Weapon Detection*

1) Experiment: Analysis of Footage Containing Weapons: Study videos where individuals are visibly carrying or using weapons (e.g., knives, firearms, blunt objects). Simulated Weapon Carrying Use: Create controlled scenarios with actors displaying weapons. Varying Weapon Types and Concealment Levels: Include experiments with different types of weapons and varying degrees of concealment.
2) Analysis: Train AI models e.g., YOLO, Faster R-CNN to detect the visual appearance of different types of weapons in the video frames. Contextual Reasoning: Consider the context of the object and the person's behavior to reduce false positives e.g., a toy gun vs. a real firearm. Pose and Action Analysis: Combine weapon detection with the analysis of hand movements and body postures that might indicate the handling or use of a weapon. Temporal Tracking: Track the movement of potential weapons over time. Fusion with Other Modalities if available: Explore the use of thermal imaging or other sensors to potentially detect concealed weapons.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

NATURAL(N) FIGHT(F)

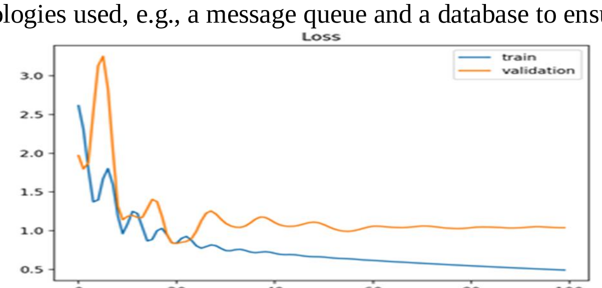
LARGE PEACEFUL GATHERING (LPG) LARGE VIOLENT GATHERING(LVG)

WEAPON DETECTION (WD)

Figure 3: Sample Frames for each Behavior Class from our Dataset

VI. SYSTEM IMPLEMENTATION

The system was implemented using a combination of software and hardware components. The deep learning models were developed using specify the deep learning framework, e.g., Tensor Flow, Py Torch and trained on a dataset comprising videos of natural crowd behavior, fights, large peaceful gatherings, large violent gatherings, and weapon detection. Object Detection: [Specify the object detection model used, e.g., YOLOv5, Faster R-CNN] was used to detect individuals in the video frames. Behavior Recognition A specify the architecture of the behavior recognition model, e.g., a 3D CNN, a CNN-LSTM network was trained to classify crowd behavior based on the movement and interactions of the detected individuals. The system was designed to process video in real-time, with alerts generated automatically when a behavior of interest was detected. The alert and reporting module was implemented using specify the technologies used, e.g., a message queue and a database to ensure reliable and timely notification.



Average loss during the training and validation

Class	Class Accuracy
N	92.77%
LPG	90.11%
LVG	72.22%
F	88.65%
WD	85.55%

Table 2: Results obtained by training the Swin transformer model on our dataset.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

The AI-enhanced crowd behavior analysis system has significant potential to impact various domains and applications, including Public Safety and Security The system can provide real-time monitoring of public spaces, enabling early detection of potentially dangerous situations such as fights, riots, and weapon threats. This can facilitate rapid response by law enforcement and security personnel, reducing the risk of harm and ensuring public safety. Event Management The system can be used to monitor crowd density and flow at large events, such as concerts, sports games, and festivals. This information can help event organizers to optimize crowd management strategies, prevent overcrowding, and ensure the safety and enjoyment of attendees. Urban Planning The system can provide valuable data on crowd dynamics in urban areas, helping city planners to design safer and more efficient public spaces. For example, the system can be used to identify areas prone to congestion or dangerous crowd behavior, allowing for the implementation of targeted interventions. Transportation The system can be used to monitor passenger behavior in transportation hubs, such as airports and train stations. This can contribute to enhanced security, more efficient passenger movement, and a better overall transportation experience. Retail Analytics The system can be used to analyze customer behavior in retail settings, providing insights into how people move through stores and interact with products. This information can be used to optimize store layouts, improve marketing strategies, and enhance the customer.

VII. 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. The system demonstrates the potential of deep learning to accurately classify a range of crowd behaviors, including natural behavior, fights, peaceful gatherings, violent gatherings, and weapon detection. Future work will focus on: Improving the accuracy and robustness of the behavior recognition models. Exploring methods for predicting crowd behavior to enable proactive intervention.

REFERENCES

[1] 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://ieeexplo re.ieee.org/document/9260138.

[2] 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.

[3] A. Awasthi, A. K. Pandey, A. K. Singh and S. Maurya, "Real-Time Monitoring and Anomaly Detection in Public Places Using Computer Vision," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-6. [DOI: 10.1109/ICCCNT49239.2020.9225436]

[4] A. Shah, S. Kumar, and P. Singh, "Real-time crowd behavior analysis using deep learning," 2020 IEEE 17th India Council International Conference (INDICON), Kharagpur, India, 2020, pp. 1-6. [DOI: 10.1109/INDICON50945.2020.9343905]

[5] B. D. Mishra, L. Huang, N. Tandon, W.-T. Yih, and P. Clark, ‘‘Tracking state changes in procedural text: A challenge dataset and models for process paragraph comprehension,’’ 2018, arXiv:1805.06975.

[6] 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

[7] 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.3 70416

[8] 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.

[9] D. Choqueluque-Roman and G. Camara-Chavez, ‘‘Weakly supervised violence detection in surveillance video,’’ Sensors, vol. 22, no. 12, p. 4502, Jun. 2022.

[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] 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.

[12] K. Singh, S. Rajora, D. K. Vishwakarma, G. Tripathi, S. Kumar, and G. S. Walia, ‘‘Crowd anomaly detection using aggregation of ensembles of fine-tuned ConvNets,’’ Neurocomputing, vol. 371, pp. 188–198, Jan. 2020

[13] 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., VOLUME12, 2024, DigitalObject.Identifier,10.1109/ACCESS.2024. 3436653.https://ieeexplore.ieee.org/stamp/stamp. jsp?arnumber=10620187.

[14] Nidhi Shetty,2Pooja Sankpal,3Priyanka Shripad,4Sandip Zade., CROWD DETECTION AND MANAGEMENT IN SURVEILLANCE SYSTEM, International Journal of Creative Research Thoughts Volume 8, Issue 3 March 2020 | ISSN: 2320- 2882. https://ijcrt.org/papers/IJCRT2003013.

[15] 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 Nano electronics and Materials Volume 14 (Special Issue) December2021[227236],https://ijneam.unimap.edu.my/images/PDF/InCAPE2021/Vol_14_SI_Dec_2021_2 27-236.

[16] 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=2 30623. [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.

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

[18] 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.

[19] V.Lam, S. Phan, D.-D. Le, D. A. Duong, and S. Satoh, ''Evaluation of multiple features for violent scenes detection,'' Multimedia Tools Appl., vol. 76, no. 5, pp. 7041–7065, Mar. 2017.

[20] W.Wang, S. Dong, K. Zou, and W. Li, ''A lightweight network for violence detection,'' in Proc. 5th Int. Conf. Image Graph. Process. (ICIGP), Jan. 2022, pp. 15–21.

[21] Y. Meng, C. C. Lin, R. Panda, P. Sattigeri, L. Karlinsky, A. Oliva, K. Saenko, and R. Feris, ''AR-Net: Adaptive frame resolution for efficient action recognition,'' in Proc. Eur. Conf. Comput. Vis., Nov. 2020, pp. 86–104.

[22] 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.

[23] 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