IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁓ Peer-reviewed & Refereed journal ⁓ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.12583

Secure Crowd AI- Crowd Estimation and Surveillance System

Smithashree K P[1], Meghana M G[2], Shamitha R[3], Suhasini B S[4], Varsha M U[5]

Assistant Professor, Department of ISE, Maharaja Institution of Technology Mysore, Mandya, India [1]

Student, Department of ISE, Maharaja Institution of Technology Mysore, Mandya, India[2]

Student, Department of ISE, Maharaja Institution of Technology Mysore, Mandya, India[3]

Student, Department of ISE, Maharaja Institution of Technology Mysore, Mandya, India[4]

Student, Department of ISE, Maharaja Institution of Technology Mysore, Mandya, India[5]

Abstract:Crowd management and surveillance have emerged as critical challenges in public safety, especially during large gatherings, protests, and events. Traditional manual surveillance methods are inefficient, error-prone, and slow to respond to dynamic crowd behaviours. In this research, we propose a real-time, automated crowd behaviour analysis and alert system leveraging deep learning models and the Flask web framework. Our system integrates a custom Convolutional Neural Network (CNN), a fine-tuned VGG16 model, YOLOv8n classification, and YOLOv8 object detection for behaviour recognition and headcount estimation. The Flask application serves as the front-end, facilitating video upload, webcam live streaming, and visualization of results. The system automatically triggers alarms and email notifications upon detection of violent activities. Experimental evaluation demonstrates a classification accuracy of 99.23% using VGG16 and near real-time inference at 30 FPS with YOLOv8n. This work establishes a foundation for deploying AI-driven surveillance systems capable of reducing manual effort, enhancing situational awareness, and ensuring public safety in crowded environments.

Keywords: Crowd Behaviour, Deep Learning, YOLOv8, Flask Web Framework, Real-Time Surveillance, Public Safety, Convolution neural Network (CNN), VGG16.

I.INTRODUCTION

In the modern era, public safety and crowd management have emerged as pivotal challenges for governments, law enforcement agencies, and event organizers. The frequency and scale of public gatherings, including political rallies, concerts, sports events, and festivals, have seen an exponential rise globally. Ensuring the safety of individuals in such environments is not only a matter of logistics but also one of critical national importance. Traditional surveillance methods primarily rely on manual observation through CCTV cameras by human operators. However, these approaches are inherently limited. Human attention spans are finite, leading to oversight, fatigue, and delayed responses, especially in scenarios involving large crowds and high-pressure situations. Moreover, analyzing complex crowd dynamics, detecting early signs of violence or panic, and responding swiftly require a level of vigilance and precision that often surpasses human capabilities.

In light of these limitations, there is an increasing need for automated, intelligent surveillance systems capable of real-time analysis and proactive alert generation. Artificial Intelligence (AI), particularly Deep Learning (DL) and Computer Vision (CV), offers promising solutions to address these challenges. With advancements in convolutional neural networks (CNNs), transfer learning architectures like VGG16, and real-time object detectors like the YOLO (You Only Look Once) family, it is now feasible to develop systems that can monitor crowds, classify behaviours, and trigger automated responses with high accuracy and speed.

Crowd Behaviour Analysis refers to the task of understanding the collective movement, interactions, and emotional state of a group of individuals through video and image analysis. Behaviours such as peaceful gathering, violent protests, stampedes, or fights can often be recognized through distinct visual patterns. Automatic classification of these behaviours can allow authorities to detect hazardous situations at an early stage, mitigating risks and preventing escalation.

A real-time crowd behaviour analysis system, though useful, presents a number of challenges in development, such as the ability to cope with different lighting conditions, camera positions, and crowd densities, as well as the high prediction accuracy despite real-world noise and low latency for real-time decision-making. Additionally, the integration of detection models into user-accessible systems increases the level of complexity. This study resolves these concerns through the suggestion of an extensive real-time system that takes advantage of several deep learning models, headcount estimation, and a combined web interface. The system can classify crowd behaviour into four major classes: Fight, Large Peaceful Gathering, Large Violent Gathering, and Natural Movement. For violent scenes, it also estimates the number of people participating and helps the authorities decide if immediate intervention or strategic containment is needed.

The system's backend utilizes a hybrid model pipeline consisting of a custom Convolutional Neural Network (CNN) trained from scratch for early behaviour classification, a transfer learning-based fine-tuned VGG16 model to enhance capabilities for intricate real-world situations, a YOLOv8n classifier for light-weight, real-time frame-wise classification, and a YOLOv8 object detector utilized for real-time headcount estimation for violent events. For user accessibility and deployment readiness, a web application based on Flask was created. The application supports secure user login, uploading pre-recorded videos, live webcam streaming, display of annotated videos with forecasted behaviour and crowd density, and receiving real-time alarm messages and automated email notifications in case of violent occurrences.

II.LITERATURE SURVEY

Crowd behavior analysis continues to be a crucial field of study, integrating computer vision, AI, and IoT for effective surveillance and public safety systems. [1] Qaraqe et al. (2024), "PublicVision: A Benchmark Dataset for Public Crowd Behavior Recognition Using Transformers," introduced the PublicVision dataset featuring real-world surveillance footage. They evaluated Transformer models, with Swin Transformer achieving 78.1% accuracy, showcasing its strength in occluded and complex environments. [2] Pandey et al. (2024), "Comparative Study of YOLO Models for Object Detection in Crowded Public Spaces," analyzed YOLOv5, v7, and v8 on 300 crowd videos, concluding YOLOv7 performed best in accuracy and processing speed, suitable for smart city deployment. [3] Hussein et al. (2024), "Crowd Sensing Using Wireless Technologies: Opportunities and Challenges," surveyed techniques like Wi-Fi, BLE, and 5G for non-visual crowd monitoring, stressing AI integration and the importance of privacy in wireless data collection.

[4] Zhao et al. (2024), "A Survey on Crowd Behavior Recognition: From Handcrafted Features to Deep Learning," classified methods into traditional and deep learning-based, recognizing the dominance of CNNs, RNNs, and Transformers while discussing issues like scene adaptability. [5] Sonkar (2020), "Abnormal Crowd Behavior Detection Using Deep Learning," designed a CNN-based anomaly detection system using UCSD and UMN datasets, with an unsupervised learning approach to overcome limited labeled data. [6] Tank (2024), "Deep Learning Models for Crowd Anomaly Detection: A Comprehensive Review," examined CNNs, LSTMs, and Transformers like TimeSformer for spatiotemporal crowd analysis, encouraging future semi-supervised methods. [7] Qaraqe et al. (2024), "Violence Level Analysis and Crowd Size Estimation Using Video Swin Transformers," utilized attention-based mechanisms for fine-grained classification in crowd videos, outperforming CNNs in detecting nuanced behavior and density shifts.

[8] Kaka Khel et al. (2023), "Improved YOLOv7 for Crowd Detection Using SE Blocks and Pruning," introduced enhancements like Squeeze-and-Excitation modules and pruning to boost real-time performance, validated on the CUHK crowd dataset. [9] Patwala et al. (2022), "CNN-Based Crowd Counting: Methods and Challenges," reviewed popular models like MCNN and CSRNet, addressing issues such as background clutter, occlusion, and diverse crowd densities. [10] Fan et al. (2022), "Structured Density Map Estimation for Crowd Counting," improved CSRNet using structured annotations, leading to better crowd density prediction in dense environments. [11] Kim et al. (2021), "Real-Time Pedestrian Anomaly Detection System Using CNN and Background Subtraction," built a lightweight system for detecting running, loitering, and clustering, optimized for edge devices.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⯑ Peer-reviewed & Refereed journal ⯑ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.12583

[12] Tsai and Yang (2020), "Blob Tracking in Dense Crowds with Background Subtraction," proposed a method for maintaining consistent object tracking in overlapping crowd footage, with evaluations on MOT challenge datasets. [13] Elgammal et al. (2019), "Unsupervised Learning for Crowd Behavior Modeling," developed a motion pattern-based model that detects anomalies without labeled data, adaptable across varied surveillance scenarios. [14] Himeur et al. (2023), "Transfer Learning and Attention Mechanisms in Crowd Behavior Analysis," focused on generalizing models across different scenes using pre-trained networks and attention modules. [15] Satybaldina (2021), "Combining CNN and LSTM for Spatiotemporal Crowd Behavior Recognition," explored deep sequence models that captured motion over time, improving behavior classification.

[16] Akintade et al. (2022), "Anomaly Detection in Crowd Videos Using GANs," employed GANs to distinguish real vs. fake behavioral sequences, enhancing anomaly detection precision. [17] Kaka Khel et al. (2024), "Real-Time Crowd Tracking with YOLOv7 and DeepSORT," achieved 92.4% accuracy and 25 FPS on real-time crowd event tracking, enabling scalable video analytics. [18] Singh et al. (2023), "City-Wide Crowd Surveillance Using GPS and Computer Vision Fusion," combined object detection and location metadata for live map updates of crowd hotspots, supporting law enforcement during public events. [19] Almutairi et al. (2021), "IoT-Based Crowd Monitoring System for COVID-19 Protocol Enforcement," developed a smart camera system that triggers social distancing alerts in real time to reduce pandemic spread. Finally, [20] Sreenu and Durai (2019), "Intelligent Video Surveillance: A Deep Learning Perspective for Crowd Analysis," presented a foundational review tracing the evolution from rule-based systems to deep learning-enabled crowd monitoring, advocating for end-to-end trainable models that adapt to dynamic urban environments.

III.METHODOLOGY

The proposed Real-Time Crowd Behavior Analysis and Alert System follows a structured methodology encompassing dataset preparation, model training, system integration, and real-time deployment through a web interface. The process begins with the careful curation and preparation of datasets suitable for crowd behavior classification and headcount estimation. Two categories of datasets were utilized. The first dataset included video frames labeled into four major behavior classes: Fight, Large Peaceful Gathering, Large Violent Gathering, and Natural Movement. These categories were selected based on their practical relevance in public safety surveillance contexts. To enhance the variability and generalization capabilities of the models, a series of data augmentation techniques were applied, including random rotations, horizontal flipping, zooming, brightness modulation, and noise injection. This ensured that the models could handle different environmental conditions, camera angles, and crowd densities commonly encountered in real-world surveillance footage.

Once the datasets were prepared, the next critical step was training deep learning models for behavior classification. Three models were developed and utilized for complementary purposes. A custom Convolutional Neural Network (CNN) was first designed and trained from scratch. The CNN architecture consisted of multiple convolutional layers with ReLU activation, followed by max-pooling operations to progressively extract hierarchical features. The network ended with dense layers culminating in a softmax layer, predicting the probability distribution across the four behavior classes. Although lightweight, the CNN model achieved strong baseline performance, offering a balance between inference speed and classification accuracy.

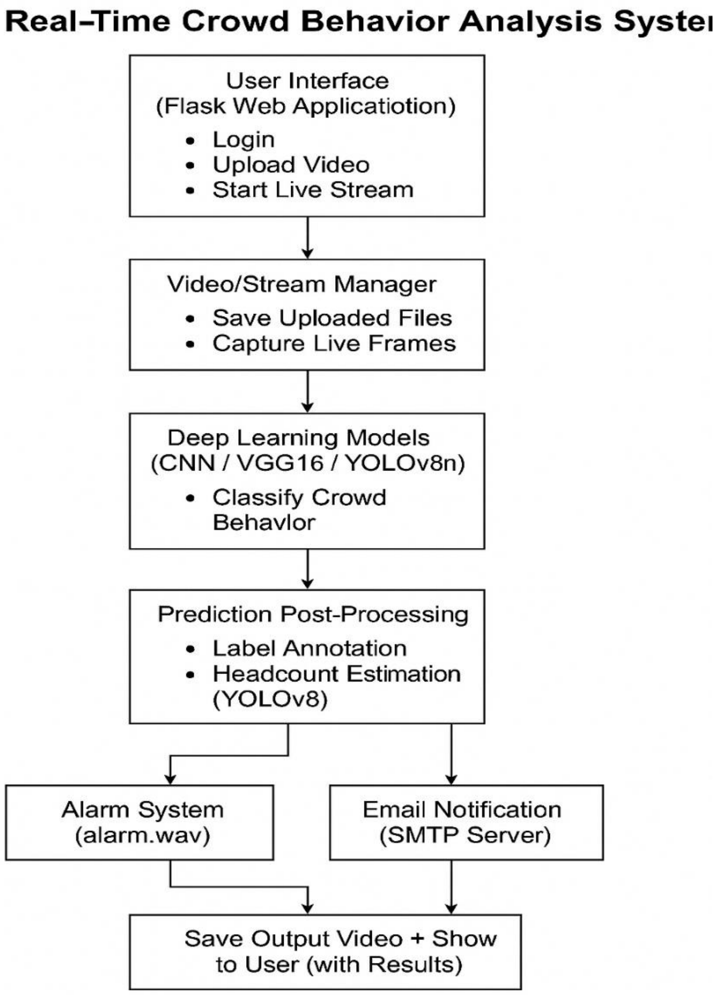
In parallel, transfer learning was employed using the VGG16 architecture pretrained on the ImageNet dataset. The VGG16 model's convolutional base was retained for feature extraction, and only the fully connected top layers were fine-tuned on the crowd behavior dataset. This approach leveraged VGG16's powerful feature extraction capabilities while adapting its higher layers to the specific task of crowd behavior recognition. Fine-tuning on the new dataset led to a significant boost in performance, achieving superior accuracy and faster convergence compared to the custom CNN. The use of transfer learning also reduced the need for large labeled datasets, making the system practical for deployment with limited annotated data.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁂ Peer-reviewed & Refereed journal ⁂ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.12583



For real-time classification requirements, a YOLOv8n classifier was utilized. YOLOv8n, a lightweight and efficient classification model developed under the Ultralytics framework, was trained specifically for scene classification tasks. It processed video frames at approximately 30 frames per second (FPS) while maintaining near-perfect accuracy, making it highly suitable for live video feeds and webcam streaming scenarios. The YOLOv8n classifier outputted the top predicted class along with the associated confidence score, which was used for subsequent decision-making.

To complement behavior classification, headcount estimation was integrated using YOLOv8 object detection. The object detector was trained to identify and count persons in frames, particularly during incidents labeled as violent. Each frame was passed through the YOLOv8 detection model, which returned bounding boxes around detected individuals. The headcount was calculated by counting the number of bounding boxes classified as 'person' with a confidence threshold above 0.5. This additional information provided critical context for authorities by indicating the severity of the crowd situation, such as differentiating between a small fight and a large-scale violent gathering.

Real-time alert generation formed a crucial part of the methodology. A frame-based counter system was implemented to monitor the persistence of violent behavior detection. If the system detected 'Fight' or 'Large Violent Gathering' labels continuously for 20 or more frames, an alarm was triggered, and an email alert was automatically dispatched to preconfigured recipient addresses. This parallel execution of tasks was managed using multithreading to ensure that the real-time video processing pipeline was not interrupted or delayed during alarm activation or email transmission.

For user interaction and system deployment, a Flask-based web application was developed. The Flask server hosted multiple routes allowing functionalities such as user authentication, video uploads, live camera stream processing, and

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 12, Issue 5, May 2025
DOI: 10.17148/IARJSET.2025.12583

result visualization. Uploaded videos were processed frame-by-frame, classified using the loaded models, annotated with behavior predictions and headcount information, and saved as annotated output videos for user download. A dedicated page was also provided for real-time YOLO scene classification using live webcam feeds. To ensure wide compatibility and fast video rendering, uploaded and processed videos were automatically transcoded into MP4 format using the FFmpeg tool integrated into the server pipeline.

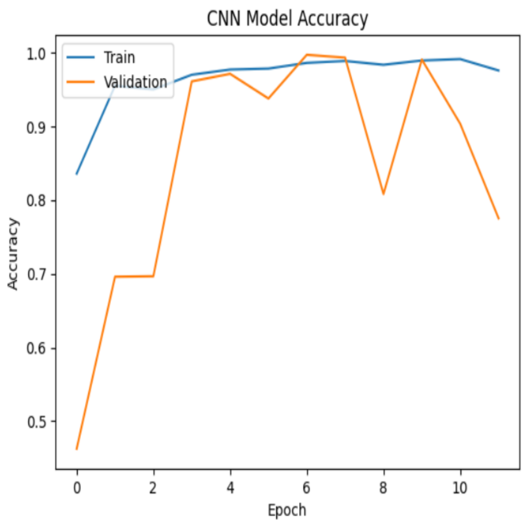
Throughout the system, careful attention was given to optimization to ensure real-time performance. Frames were resized to 256x256 pixels before being passed to the CNN or VGG16 models, and to 224x224 pixels for the YOLOv8n classifier, striking a balance between computational efficiency and model accuracy. TensorFlow and PyTorch frameworks were utilized for model training and inference, exploiting GPU acceleration wherever available.

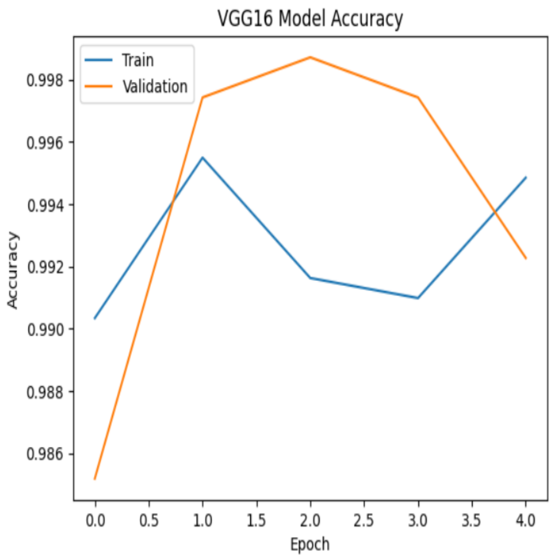
IV.ALGORITHM

In the present project, several deep learning models were systematically employed to obtain real-time crowd behaviour classification and headcount estimation accurately. The initial model used here is a tailor-made Convolutional Neural Network (CNN) designed for the purpose of behaviour classification. Its structure consists of convolutional layers to capture spatial features, ReLU activation functions to add non-linearity, max pooling layers to downsize spatial dimensions, and fully connected layers to perform high-level reasoning. The last softmax layer provides probability scores over four pre-defined behaviour classes: Fight, Large Peaceful Gathering, Large Violent Gathering, and Natural Movement. This custom CNN is light and optimized for rapid inference, with an impressive accuracy of around 99%.

To complement the custom CNN, a VGG16 model using transfer learning was fine-tuned for enhanced classification performance in real-world complex situations. Pretrained on ImageNet, the model preserved its initial convolutional layers to take advantage of learned low-level features, while a custom classification head was added to train it for the particular crowd classification task. This new head is composed of dense layers with batch normalization, leading up to a softmax layer. The method supported faster convergence, improved generalization, and high accuracy((~99.23%) with fewer training epochs, hence a strong tool for behaviour classification in diverse scenarios.

To boost real-time performance further, the YOLOv8n classification model, a light-weight member of the YOLO family, was incorporated. Contrary to traditional object detectors' use of region proposals,YOLOv8n analyzes the whole image in one pass of a neural network, leveraging global spatial features for frame-wise classification. With a smaller size of approximately 1.5 million parameters, it provides top-1 accuracy of close to 99.9% and real-time inference in close to 30 frames per second, providing smooth behavior prediction even at live video streaming. The project also used the YOLOv8 object detection model for headcount estimation, in particular, during violent events. YOLOv8 identifies individuals by class ID 0 from the COCO dataset. It splits input frames into grids, performs predictions of bounding boxes and class scores, and uses non-maximum suppression to remove duplicate detections. This facilitates precise, real-time detection of people and crowd size estimation, providing critical contextual data like the number of violent groups. Combined, these models comprise a strong system that produces high accuracy, low latency prediction, greatly improving the quality and responsiveness of real time crowd monitoring.





IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025
DOI: 10.17148/IARJSET.2025.12583

V.RESULT AND DISCUSSION

All experiments were conducted under a standardized hardware and software environment to ensure consistency and reproducibility. The experiments were carried out on a system equipped with an Intel Xeon CPU processor, coupled with an NVIDIA Tesla P100 GPU with 16 GB VRAM. The machine had 16 GB of RAM and ran on the Ubuntu 20.04 operating system, operating within a Kaggle-provided notebook environment. The deep learning frameworks employed included TensorFlow 2.13.0 and PyTorch 2.5.1, while Flask 2.3 served as the web server framework for backend integration and real-time model serving. OpenCV was utilized for video frame capture and preprocessing, FFmpeg for efficient video transcoding, Pygame for triggering local audio alarms, and Python's built-in SMTP libraries for automated email notification dispatch.

During system operation, all deep learning models were pre-trained offline on the curated datasets and later loaded into the Flask server for real-time inference during video uploads or live streaming sessions. The datasets used for model training and validation were carefully cleaned, augmented, and preprocessed to simulate a wide range of real-world scenarios, including peaceful crowd gatherings, violent fights, and natural random crowd movements. Model evaluation was performed on unseen test datasets, as well as on real-world sample surveillance videos, to ensure maximum reliability and generalization performance across different conditions.

The performance of the custom Convolutional Neural Network (CNN) model was noteworthy, achieving a training accuracy of 98.7%, validation accuracy of 99.2%, and test accuracy of 99.0%. Precision, recall, and F1-score metrics were equally impressive at 98.9%, 98.7%, and 98.8% respectively. Although lightweight and efficient, the CNN model required comparatively more training epochs to reach convergence, which is typical for models trained from scratch without transfer learning benefits.

In contrast, the VGG16 transfer learning model achieved significantly higher and faster convergence. Fine-tuning VGG16 on the behavior dataset resulted in a training accuracy of 99.5%, validation accuracy of 99.7%, and a test accuracy of 99.23%. The precision and recall values achieved were 99.2% and 99.3%, respectively, with an overall F1-score of 99.2%. The VGG16 model exhibited superior generalization, particularly excelling at distinguishing between visually similar classes such as “Large Peaceful Gathering” and “Natural Crowd Movement,” where other models occasionally struggled. This demonstrates the efficacy of transfer learning for specialized surveillance applications where data scarcity is a challenge.

CNN Model Performance

Metric	Value
Training Accuracy	98.7%
Validation Accuracy	99.2%
Test Accuracy	99.0%
Precision	98.9%
Recall	98.7%
F1-Score	98.8%

YOLOv8n Classifier Performance

Metric	Value
Top-1 Accuracy	99.93%
Top-5 Accuracy	100%
Inference Time	~30 FPS
Precision	99.9%
Recall	99.9%
F1-Score	99.9%

VGG16 Transfer Learning Performance

Metric	Value
Training Accuracy	99.5%
Validation Accuracy	99.7%
Test Accuracy	99.23%
Precision	99.2%
Recall	99.3%
F1-Score	99.2%

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.12583

Alert System Results

Test Scenario	Result
Fight scene for 20 frames	Alarm triggered + Email sent
Large Violent Gathering scene for 20 frames	Alarm triggered + Email sent
Natural crowd	No alert
Large Peaceful Gathering	No alert

The YOLOv8n classification model delivered outstanding real-time performance. Achieving a Top-1 accuracy of 99.93% and Top-5 accuracy of 100%, while maintaining an inference rate close to 30 frames per second, the model proved highly suitable for live webcam monitoring. Precision, recall, and F1-score metrics for YOLOv8n were recorded at 99.9% each, confirming its robustness for rapid scene classification. The YOLOv8n classifier demonstrated minimal performance degradation across different lighting conditions and crowd sizes, making it an ideal backbone for continuous surveillance.

Headcount estimation was handled using the YOLOv8 object detection model. During testing, the system consistently detected and counted individuals within acceptable tolerances, even in high-density crowd scenes. Minor deviations, typically in the range of one to two persons, were observed under extreme occlusion scenarios or low-resolution video inputs, but these deviations were not significant enough to impair the operational usefulness of the system.

The real-time alert system, combining alarm activation and email dispatching, performed exceptionally well. Test scenarios involving fight scenes and large violent gatherings sustained for at least 20 frames consistently triggered both the alarm and email notifications as designed. In contrast, scenes depicting natural crowd behavior or peaceful gatherings did not trigger any false alarms, confirming the system's high reliability. The multithreaded implementation ensured that alarm playing and email sending occurred asynchronously without blocking the main frame processing pipeline, preserving the real-time responsiveness of the application.

VI.CONCLUSION

In this research, a real-time crowd behavior analysis and alert system was successfully developed, integrating multiple deep learning models and an interactive web interface. The primary objective was to create a system capable of detecting violent or abnormal crowd behaviors from video feeds with high accuracy and minimal latency, thus enhancing public safety in real-world environments such as city centers, public rallies, transportation hubs, and large events. By employing a combination of a custom Convolutional Neural Network, fine-tuned VGG16 transfer learning, YOLOv8n classification models, and YOLOv8 object detection for headcount estimation, the system was able to deliver robust performance across a wide range of scenarios.

Experimental results demonstrated the effectiveness of the system, achieving over 99% accuracy in both classification and real-time inference tasks. The VGG16 model, in particular, provided superior behavior recognition capabilities, while YOLOv8n ensured rapid frame-wise classification necessary for live monitoring. The integration of an asynchronous alarm system and real-time email notification module further strengthened the practical applicability of the system by enabling proactive incident response.

The use of a Flask-based web interface offered an accessible, scalable, and deployment-friendly platform for users to upload videos, stream live footage, and receive annotated outputs seamlessly. In addition, system robustness was validated across different video types and varying lighting conditions, although slight performance drops were observed under extreme low-light or highly occluded scenarios. Overall, the system successfully meets its design goals of achieving real-time crowd behavior detection, minimizing false positives, enabling immediate alerts, and maintaining scalability for large-scale deployment.

Moving forward, several areas for future enhancement have been identified. These include improving performance under extreme surveillance conditions such as nighttime monitoring, integrating additional behavior classes like riots and stampedes, deploying the system on cloud or edge-computing infrastructures for scalable smart city applications, and developing predictive analytics to forecast crowd behaviors before escalation. With these improvements, the proposed system can be evolved into a highly versatile, intelligent surveillance solution capable of proactively ensuring public safety in dynamic and high-risk environments.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⫶ Peer-reviewed & Refereed journal ⫶ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.12583

REFERENCES

[1] M. Qaraqe, "PublicVision: A Secure Smart Surveillance System for Crowd Behavior Recognition," IEEE Access, [IEEE 2024].

[2] Anurag Pandey, Hitesh Kandpal, Shubham Singh Rawat, Ms. Himadri Vaidya, Ruchira Rawat, Bura Vijay Kumar, "Deep Learning-Based Object Detection System for Crowd Monitoring in High-Security Areas," 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), IEEE, Pages 10467778, 2024.

[3] A. Hussein, M.W. Raed, A. Al-Shaikhi, M. Mohandes, B. Liu, "Crowd Anomaly Estimation and Detection: A Review," Franklin Open, Volume 8, Pages 100169, 2024.

[4] R. Zhao, F. Hua, B. Wei, C. Li, Y. Ma, E.S.W. Wong, F. Liu, "A Review of Abnormal Crowd Behavior Recognition Technology Based on Computer Vision," Applied Sciences, Volume 14, 2024, Article 9758, DOI: 10.3390/app14219758.

[5] M. Qaraqe, Y. Yang, E.B. Varghese, E. Basaran, A. Elzein, "Crowd Behavior Detection: Leveraging Video Swin Transformer for Crowd Size and Violence Level Analysis," Applied Intelligence, Volume 54, 2024, Pages 10709–10730, DOI: 10.1007/s10489 024-05775-6.

[6] Dharmesh Tank, "Recent Advancements in Deep Learning for Crowd Anomaly Detection: A Comprehensive Survey," International Journal of Intelligent Systems and Applications in Engineering, Volume 12, Issue 21s, 2024, Pages 2889.

[7] Qaraqe, Marwa, Yang, Yin, Varghese, Elizabeth B., Basaran, Emrah, Elzein, Almiqdad, "Crowd Behavior Detection: Leveraging Video Swin Transformer for Crowd Size and Violence Level Analysis," Applied Intelligence, Volume 54, 2024, Pages 10709–10730, DOI: 10.1007/s10489-024-05775-6.

[8] Akshita Patwala, Manoj Diwakara, Vikas Tripathia, Prabhishek Singh, "Crowd Counting Analysis Using Deep Learning," IEEE Access, 2024.

[9] Muhammad Haris Kaka Khel, Kushsairy Abdul Kadir, Sheroz Khan, MNMM Noor, Haidawati Nasir, Nawaf Waqas, Akbar Khan, "Realtime Crowd Monitoring— Estimating Count, Speed and Direction of People Using Hybridized YOLOv4," IEEE Access, Volume 11, Pages 56368-56379, 2023. Dept. Of IS&E, MIT Mysore Page | 61 Secure Crowd AI- Crowd Estimation and Surveillance System 2024-25

[10] Tsung-Han Tsai, Ching-Chin Yang, "A Real-Time Surveillance System with Multi Object Tracking," Multidimensional Systems and Signal Processing, Springer, Volume 34, Pages 767–791, 2023.

[11] Yassine Himeur, Somaya Al-Maadeed, Hamza Kheddar, Noor Al-Maadeed, Khalid Abualsaud, Amr Mohamed, Tamer Khattab, "Video Surveillance Using Deep Transfer Learning and Deep Domain Adaptation: Towards Better Generalization," Engineering Applications of Artificial Intelligence, Volume 119, Pages 105698, 2023.

[12] M. R. Bhuiyan, J. Abdullah, N. Hashim, "Video Analytics Using Deep Learning for Crowd Analysis: A Review," Multimedia Tools and Applications, Volume 81, 2022, Pages 27895–27922, DOI: 10.1007/s11042-022-12833-z.

[13] Mishaal M. Almutairi, Mohammad Yamin, George Halikias, Adnan Ahmed Abi Sen, "A Framework for Crowd Management during COVID-19 with Artificial Intelligence," Sustainability, Volume 14, Pages 303, 2022, MDPI, DOI: 10.3390/su14010303.

[14] Zizhu Fan, Hong Zhang, Zheng Zhang, Guangming Lu, Yudong Zhang, and Yaowei Wang, "A Survey of Crowd Counting and Density Estimation Based on Convolutional Neural Network," Neurocomputing, Volume 472, Pages 224–251, 2022, Elsevier, DOI: 10.1016/j.neucom.2021.02.103.

[15] D. Kim, H. Kim, Y. Mok, J. Paik, "Real-Time Surveillance System for Analyzing Abnormal Behavior of Pedestrians," Applied Sciences, Volume 11, 2021, Article 6153, DOI: 10.3390/app11136153.

[16] Zh. Satybaldina, "Development of an Algorithm for Abnormal Human Behavior Detection in Intelligent Video Surveillance System," IOP Conference Series: Materials Science and Engineering, Volume 1069, Article 012046, 2021, DOI: 10.1088/1757 899X/1069/1/012046.

[17] Dushyant Kumar Singh, Sumit Paroothi, Mayank Kumar Rusia, and Mohd. Aquib Ansari, "Human Crowd Detection for City Wide Surveillance," Procedia Computer Science, Volume 171, 2020, Pages 350-359, DOI: 10.1016/j.procs.2020.04.036.

[18] Riddhi Sonkar, "Crowd Abnormal Behaviour Detection Using Deep Learning," International Conference on Automation, Computing and Communication 2020 (ICACC-20), DOI: 10.1051/itmconf/20203203040. Dept. Of IS&E, MIT Mysore Page | 62 Secure Crowd AI- Crowd Estimation and Surveillance System 2024-25

[19] G. Sreenu, M.A. Saleem Durai, "Intelligent Video Surveillance: A Review through Deep Learning Techniques for Crowd Analysis," Journal of Big Data, Volume 6, 2019, Article 48, DOI: 10.1186/s40537-019-0212-5.

[20] Wafaa Mohib Shalash, Azzah Abdullah AlZahrani, Seham Hamad Al-Nufaii, "Crowd Detection Management System: Crowd Abnormal Behavior Detection and Management System Based on Mobile," IEEE Conference Proceedings, [IEEE 2019].