

VisionFlow AI: Real-Time Crowd Detection and Safety Automation System

Keerthana Gigi¹, Dr Shajeni Justin²

¹Department of Computer Science, Sacred Heart College, Thevara, Kochi, India

²Assistant Professor, Department of Computer Science, Sacred Heart College, Thevara, Kochi, India,

Abstract

VisionFlow AI is an advanced system tool that provides real-time crowd behaviour analysis and tracking. The system also initiates safety actions in public places automatically when it detects large crowds gathering. Many current technologies continue to depend on human observation and CCTV monitoring, which are often slow, inaccurate and ineffectual when crowds are changing rapidly. To significantly reduce this problem, this system utilises deep learning and employs video streaming techniques, such as YOLOv8n, for rapid person detection. Additionally, it includes zone analysis, which allows it to provide information on crowd distribution across multiple areas, enabling users to identify potential congestion points. Once it has reached the specified threshold for crowd occurrence, an automatic decision module initiates the appropriate safety protocols (e.g., alarms) and notifies appropriate people to take necessary action. Performance testing indicates that it has an extremely low latency and is very reliable with a high level of accuracy, making it appropriate for real-time platforms. Due to its construct and because it is modular, affordable, flexible, and scalable, it is the ideal solution for events, sporting venues, transportation hubs, and emergencies.

Keywords: Crowd Monitoring, YoloV8, Computer vision, Automated Safety Control, People Detection, Real-Time Video Analytics

1. Introduction

1.1 Background and Motivation

As urbanisation increases and large gatherings become more frequent at public events, effective crowd management has become essential for public safety. Railway stations, temples, shopping centres, and events can experience sudden spikes in population density, and if this is not closely monitored, it could pose a danger to people in these locations. Traditionally, surveillance depends heavily on human operators to observe live CCTV feeds, leading to potential delays, fatigue, and subjective judgments. Some of the problems associated with spike population masses have been resolved using computer vision and image processing through crowd analysis. Unfortunately, these deep-learning-enabled detection module algorithms require execution on a GPU-accelerated computer for high throughput, which is not available to all users, nor is it a viable option in environments with resource constraints. As such, a solution offering practical alternatives will be addressed.[1].

1.2 Problem Statement

Numerous individuals take pictures of possible crime scene locations to investigate for signs of wrong behaviour. The visual capture framework constitutes another way to provide insight into how to act in the

aftermath of an incident, but imaging systems are unable to identify anything currently. In cases of many people being present, it is extremely difficult to manually maintain supervision of so many live video feeds that contain constantly and often randomly moving subjects in and around busy public spaces. Not being able to identify the number of people in a location, or what future entry and exit might be, will invariably lead to conditions of overcrowding, inadequate space and many significant safety issues before anything has the opportunity to be done to address these problems. The majority of new technology that relies on the monitoring of people uses not only traditional CPUs, but also the acceleration of these systems through the use of GPUs. The objective of this is to create an accurate and reliable mechanism that uses only the microprocessor to monitor the crowd and alert officials when there are problems.

1.3 Objectives and Contributions

This project aims to develop a live solution for tracking and identifying people, as well as assessing the density of a population by using visual data captured from ordinary surveillance cameras connected to regular computers, to specify the state of occupancy against pre-defined criteria of safety and to provide an alert for unsafe conditions to occur. In addition, keeping the system economically viable and computationally efficient to allow the solution to be used as reliably as possible for extended periods of time will also be a primary objective of the project. Primary outcomes include the creation of a user-friendly/existing human recognition framework, development of a web-based dashboard to continuously monitor the environment and receive automated notifications about upcoming alerts promptly, with fewer redundant notifications.

2. Literature Survey

The study by A. A. Alsabei, T. M Alsubait, and H. H. Alhakami addresses dangerous behaviours during Hajj gatherings in high-density situations [4]. Previous methodologies using optical techniques struggled due to issues like occlusion and unpredictable motion. This project implements a mixed-model approach with YOLOv9, enhancing multi-scale processing and feature extraction capabilities. The model, trained on the HAJJv2 dataset of actual low-density Hajj videos, effectively identifies actions such as pushing and rapid running, even with many individuals present.

The study by Alhawsawi, Khan, and Rehman highlights the challenges of detecting people in high-altitude drone footage due to their tiny appearances against complex backgrounds [6]. They propose a Context Enrichment Module (CEM) incorporating dilated aurous convolutions, which enhances the standard YOLOv8's spatial context without adding parameters, thereby improving the detection of small figures amidst similar patterns. Evaluating the VisDrone-CC2021 dataset shows that the enhanced YOLOv8 effectively addresses these challenges and underscores the importance of contextual reasoning in aerial crowd counting.

The study by Bilade, Fulaware and Gaikwad improves safety by detecting falls, overcrowding, fire, and weapons from real-time video streams [2]. The general problem stated in this work is that the human eye cannot persistently or effectively monitor multiple CCTV feeds and, hence, many delays and errors occur. This deficiency is overcome here, where AI-based computer vision with deep-learning models is implemented on live and recorded video. No standard benchmark dataset is used, but custom-created training data is included in the proposed system for fall and crowd detection. This study proves that incident detection and response time are much improved through automated surveillance.

The study by Siva addresses the limitations of traditional CCTV surveillance, which often fails due to human fatigue and slow responses [5]. It introduces a system employing the YOLOv8 framework for real-

time object detection, combined with deep learning-based anomaly detection. The system is trained on extensive surveillance datasets to enhance detection accuracy. This method enables effective real-time crowd analysis and threat identification, targeting issues like overcrowding and suspicious activities in public and workplace settings. The study by Gupta et al. addresses challenges in manual crowd monitoring, such as low accuracy and high labour costs [3]. It proposes an AI-based real-time system using YOLOv8, classical image processing, and centroid tracking to detect and count individuals in video frames. The method involves pre-processing, object detection, and tracking, validated using the CrowdHuman dataset with occluded samples, leading to improved detection accuracy and crowd counting over traditional approaches.

3. Definition of the Challenge and System Requirements

3.1 Problem Formulation

Recent urbanisation and increased crowd-related incidents have created a pressing need for effective safety supervision. Traditional methods relying on manual observation are slow, inefficient, and prone to human error, making it difficult for authorities to manage large crowds. Although research has put forth automated crowd analysis systems, these face significant challenges, including inaccurate head counts, tracking failures, delayed warnings, and a lack of real-time decision support. Additionally, the absence of integrated alert systems and predictive analytics compromises their effectiveness. As such, there is a clear demand for an advanced crowd monitoring solution that can intelligently and reactively count individuals in real-time using live video feeds and promptly alert authorities to potential threats.

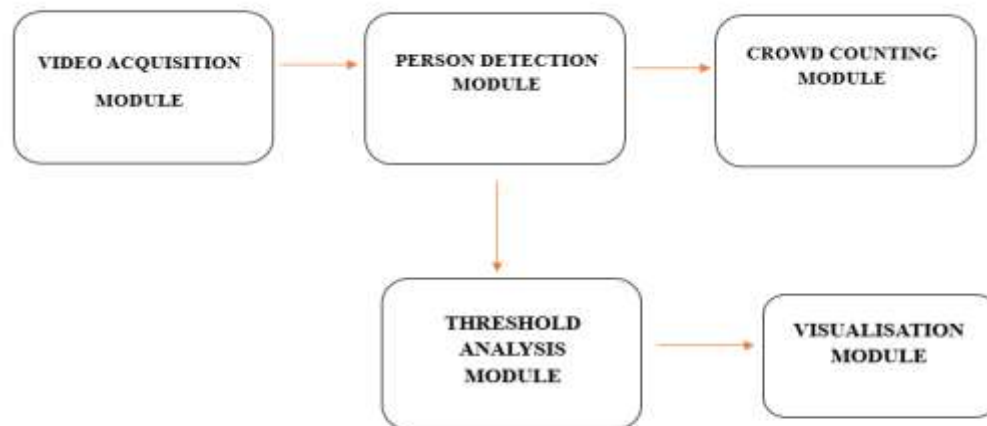
3.2 Framework Supplies and Constraints

An effective people observation system requires performance measurement criteria for real-time data collection using deep learning algorithms to accurately detect individuals and minimise counting errors, such as duplication. The system should trigger visual and audible alarms if safety limits are exceeded and feature an interactive dashboard for real-time updates on occupancy levels and event notifications. It must provide reliable identification across varying conditions, operate with low latency, and support scaling with additional video inputs for extensive coverage. Despite advances in mobile technology, limitations in video clarity and processing power hinder performance due to the lack of dedicated GPUs and potential connectivity issues. Continuous re-training of recognition systems may be necessary to ensure effectiveness in new environments.

4. Proposed System Architecture

4.1 High-Level Design Overview

The new crowd management system is a modular, real-time framework that detects individuals, measures occupancy, and triggers alerts when crowd levels exceed safe limits. It consists of four layers: data acquisition, analysis using a YOLOv8n deep learning algorithm, evaluation against pre-established limits, and notification/display. The system processes video frames to rapidly identify people, calculates occupancy, and classifies conditions as Normal, Warning, or Critical. It displays the image stream and occupant count, generating alarms for unsafe density levels, thereby ensuring reliability in dynamic environments.

Figure 1: System Architecture of the Proposed Crowd Monitoring System

4.2 Module Overview

This scheme will provide real-time crowd management by identifying individuals, tracking how many people are present in an area at any given moment, and generating visual indicators for when the number of people present in a certain area reaches a predetermined level of crowd density. To do this, each of the modular components of the overall system performs unique functions as part of a continuous processing pipeline that yields accurate results with efficient analysis of the real-time video feed.

Video Capture Module:

The system enables continuous analysis of visual data by collecting live on-screen streams from real-time surveillance devices or other monitoring equipment installed within the selected observation environment. The video input is decomposed into sequential image frames to enable automated analysis. Before advanced processing, these frames may undergo initial enhancements such as scaling, format standardisation, or quality adjustments to ensure uniform input conditions and improve detection reliability. After completing the preparation stage, the refined frames are forwarded to the human detection component for further analytical processing.

Module For Detecting People:

In order to find humans in a given video frame, the discovery part uses a pre-trained deep learning object recognition model to analyse visual patterns and spatial data from each incoming frame. Through this analysis, the unit locates people and identifies their locations with bounding boxes around them. The structured outcome data will serve as an input to the next stages of counting and evaluation.

Population Counting Module:

It sums the number of people that have been detected in frames and produces updated crowd size changes. Enumeration is done as frequently as possible to give a real-time quantitative measure of the number of persons in people density, which is essential to assess how many persons are in the crowd at any given point in time and assist with identifying crowd safety/causes. The content has not been changed structurally, but some punctuation has changed.

Decision-Making Module

The threshold analysis component evaluates crowd conditions by comparing real-time counts against a predetermined boundary for safety. This comparison will be used to group the environment being

monitored into various alert levels (normal, caution, critical). If a crowd exceeds the allowable count, the cutoff value component will activate a warning signal and initiate the proper alert responses to allow for an appropriate intervention.

Visualisation Component:

This section shows processed data in a monitoring interface via a live video stream, which includes overlays like detection boxes, current crowd counts, and alert notifications. These visual cues help the operator quickly understand crowd conditions and make informed decisions when abnormal population levels are detected.

4.3 Data Flow and Design Rationale

A system that facilitates the quick and efficient processing of video streams was developed by treating each frame as a discrete item, allowing for less delay and a quicker response time from the data. The use of a pre-trained deep-learning algorithm was incorporated into the computation pipeline to improve the accuracy and reliability of identifying people. The architecture was divided into specific functional modules such as detecting people, estimating crowd size, and managing alerts, making the scheme simple to maintain and providing flexibility in terms of future tactics updates. The backend of the solution was developed using the lightweight Flask framework due to its efficient performance and ability to work well in conjunction with computer vision applications. The strategy was also designed with an asynchronous methodology for sending out alert notifications, ensuring that there was no interruption in the normal flow of the primary analysis phase while delivering the notifications. Overall, the focus of the design process was to produce a dependable, fast-responding, and easy-to-configure solution that could provide an effective and maintainable method for monitoring crowds.

5. Operational Workflow of the System

5.1 Arrangement Roadmap

The proposed crowd monitoring solution will function in real time, as shown in the above diagram, and requires only one webcam located at a location of interest. The camera will produce a digital video feed (or camera input) and send it immediately to the system for examination. Analysing each frame of the live feed will provide information on crowd counting and the estimation of troop size. The number of people counted will then be compared against previously defined safety limits or criteria, and whether additional attention is required for the gathering based on its size or the conditions in the physical environment.

5.2 Yolo -Based Estimating Density of Crowds

Real-time people detection using a live camera feed can be achieved with deep learning object detection techniques to identify individuals. This method allows for a single pass to locate individuals in the live view, making it efficient. In contrast to older approaches that required multiple scans of each image—making continuous monitoring difficult—modern systems draw bounding boxes around identified persons and tally the valid identifications per frame. By consolidating person recognitions across frames, the system can estimate the overall crowd size and density.

5.3 Threshold Evaluation Strategy

The number of individuals, as calculated based on predefined safety limits, is evaluated to determine their status. A warning level indicates a gradual increase in the number of occupants; a critical level indicates possible dangerous, congested conditions. The estimation method gives the opportunity to intervene quickly and to proactively manage risk through one camera view.

5.4 Alert Generation Mechanism

The on-site alarm system continuously monitors crowd density and allows for prompt reactions if an area becomes too crowded. It evaluates live camera feeds of the venue and calculates the current number of people in an area, and compares that number to predetermined safety thresholds. When the number of people in that area approaches or exceeds those thresholds, the system automatically sends a series of alerts, including activating a loud alarm to alert people in the venue to take immediate action to alleviate the overcrowding, for example, by prohibiting anyone from entering the venue or redirecting people already present to leave the area. In addition, the system sends real-time information about the number of people in a given area via WhatsApp to authorised personnel to enable rapid assessment and action from outside the venue. Attendees who are registered in advance will also receive pre-departure alerts when the venue is expected to reach maximum capacity, which will help control the rate at which people are arriving at the venue, ensure that the venue is kept at a safe level of occupancy, and help to prevent congestion in the venue before it becomes significant.

6. Implementation Details

This system was built using Python and utilises OpenCV to capture and process video frames in real time, along with a pre-trained deep learning model for accurately detecting humans from both moving and still images. Confidence, and IoU thresholds along with Non-Maximum Suppression were set as optimum levels to ensure that detection accuracy and speed were balanced and duplicate detections are not produced. This system is deployed via a lightweight Flask application which provides a real-time web dashboard containing live video, continuous estimates of the size of a crowd (value-crowd, normal, absolutely, warning and critical) and visual cues that indicate how many people are in the crowd. Lastly, a multi-core CPU provides the robust performance that the application needs to function smoothly and respond promptly.

Figure 2. Configuration interface for real-time AI-driven crowd density monitoring and alert generation.

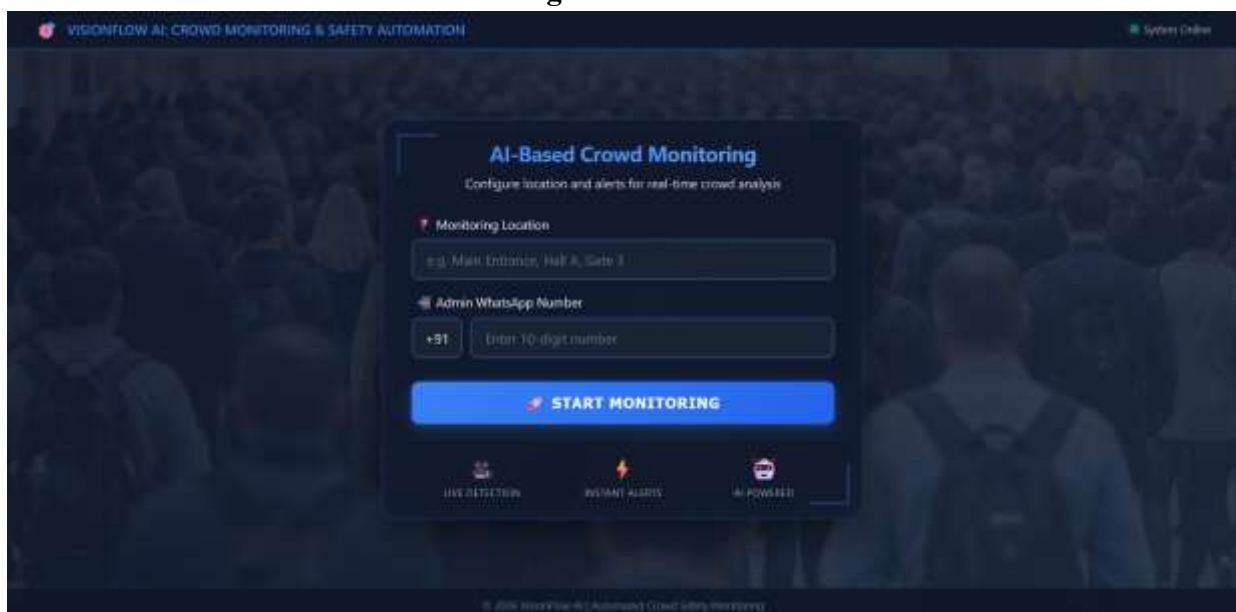
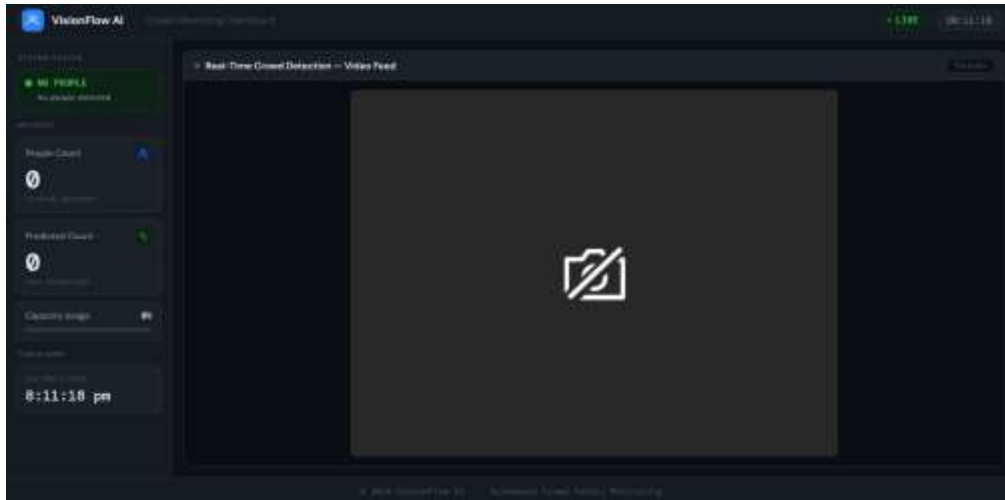


Figure 3: Real-time dashboard with video feed, counts and timestamps.



6.1 Detection and Alert Module Integration

The recognition and notification components are integrated for automatic crowd monitoring. The system continually screens and compares the current crowd count with the predefined threshold. Once the measured level reaches the warning or critical range, an alarm is created automatically. Because all three actions occur in parallel (detection, visual display, and generating signal), the overall performance will not be affected.

6.2 Alert Rate Control and Error Handling

A cooldown mechanism will limit the number of repeated updates and/or notifications when there are long periods of overcrowding. Additionally, for temporary interruptions of the video gallery stream, fault handling will allow for continued operation of the application and resumption of normal operations once the stream has resumed.

7. Experimental Setup

A real-time video feed from a laptop camera was evaluated without prerecorded data, enabling assessment under normal conditions. The system monitored varying population densities by analysing live video frames, classifying individuals as low, medium, or high density with assigned notification thresholds. Tested with three controlled individuals, the framework aimed to address issues like occlusion in single-camera setups while measuring system performance regarding recognition accuracy, crowd categorisation, and notification activation. An alarm was triggered when monitored individuals exceeded a predefined limit, with this occupancy threshold adjustable based on practical needs.

Table 1: Summary of the test cases that were evaluated as part of the system validation process

Persons	Level
1 – 35	Normal
36 – 50	Warning
≥ 51	Critical

Figure 4: Normal operation with crowd count below the threshold (limit=50).

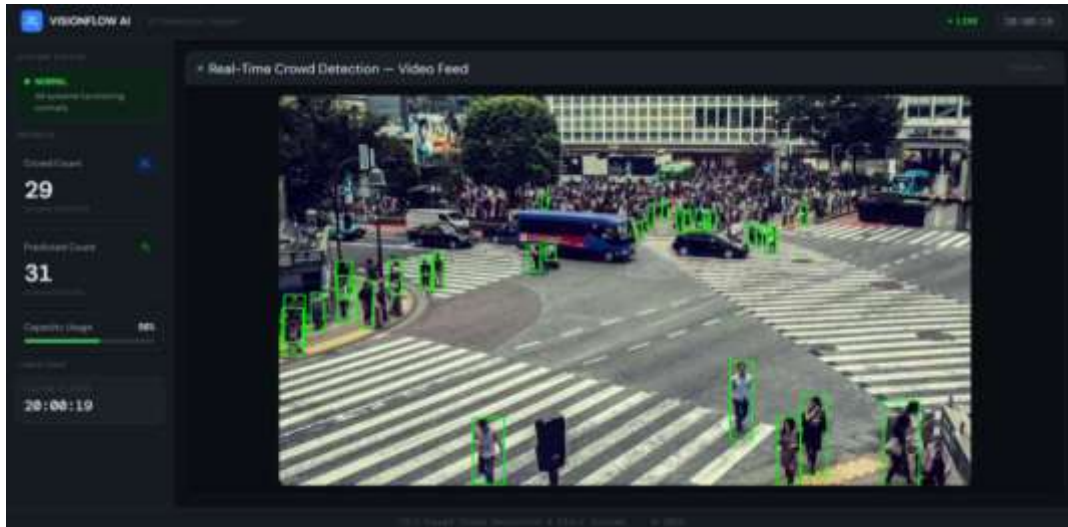


Figure 5: Warning state as detected occupancy approaches the threshold (limit=50).

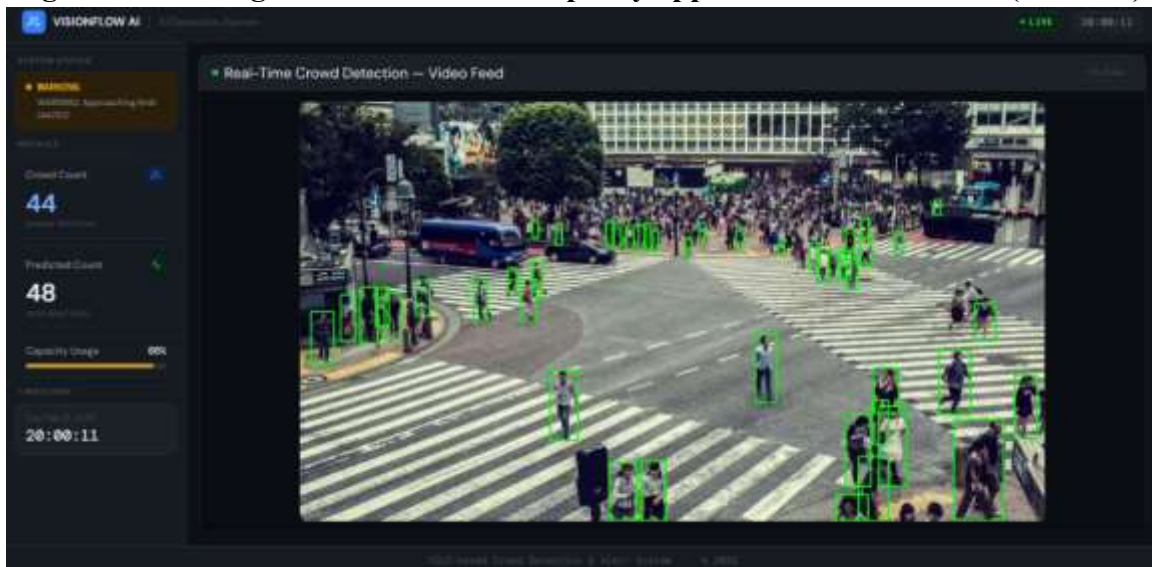
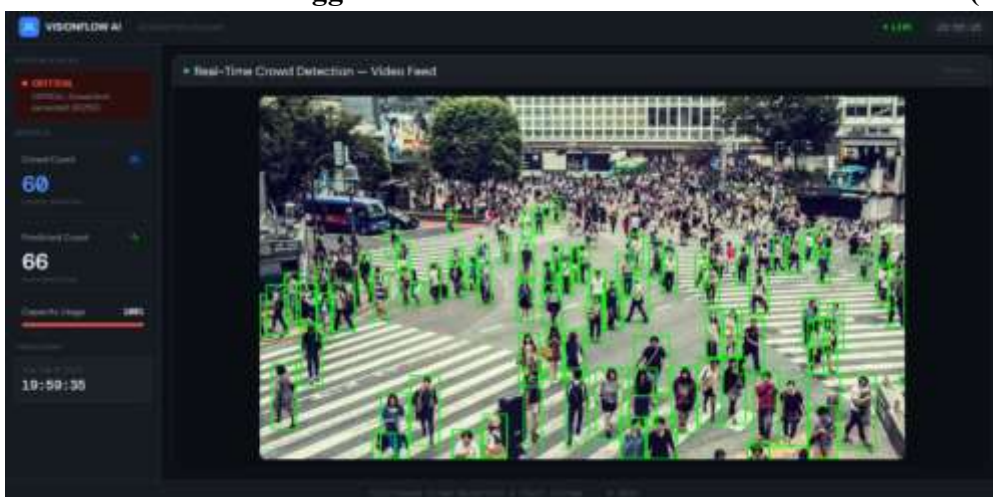


Figure 6: Critical condition triggered when crowd count exceeds the threshold (limit=50).



8. Performance Evaluation and Metrics

8.1 Assessment Measures

The study of how well this crowd monitoring solution works included two areas of examination: recognition parameters and a system of measurement for overall system efficiency. The accuracy of recognition was assessed by counting the number of errors made for each person recognised in a one-hour event (live-streamed) using a laptop camera for capture. By evaluating correctness through precision and recall, and real-time efficiency by assessing the frames processed per second (FPS). To assess how long it takes to process each frame (to evaluate processing delays) as well as to refresh the system (to evaluate responsiveness).

8.2 Detection Accuracy and Alert Capability

The evaluation demonstrated that humans could be recognised successfully regardless of the occupancy of people present in an area. At low- and moderate-density occupancies, the identification of persons was relatively stable and yielded consistent results. As density became higher, small variations occurred in the identification process due to partial blockage of persons being identified; however, when the number of individuals exceeded the defined parameters of occupancy, either a cautionary or critical warning was sent from the notification system based on the severity of the event. The notification system provided immediate updates when the monitored number of people exceeded the configured occupancy limits. The number of notifications sent to users was kept to a minimum by performing validation checks and establishing specified time intervals to align with threshold level criteria.

Figure 7: Advanced crowd advisory alert generated by the system through an instant messaging service

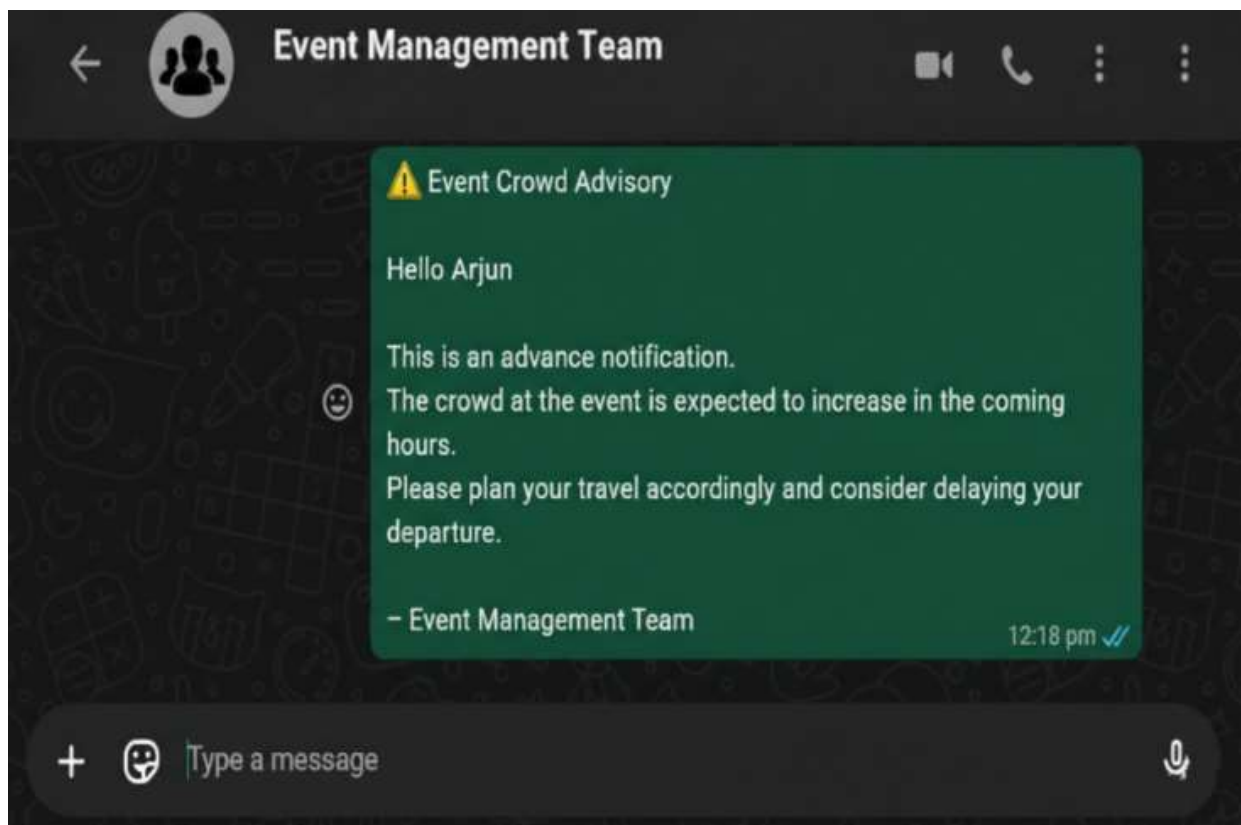
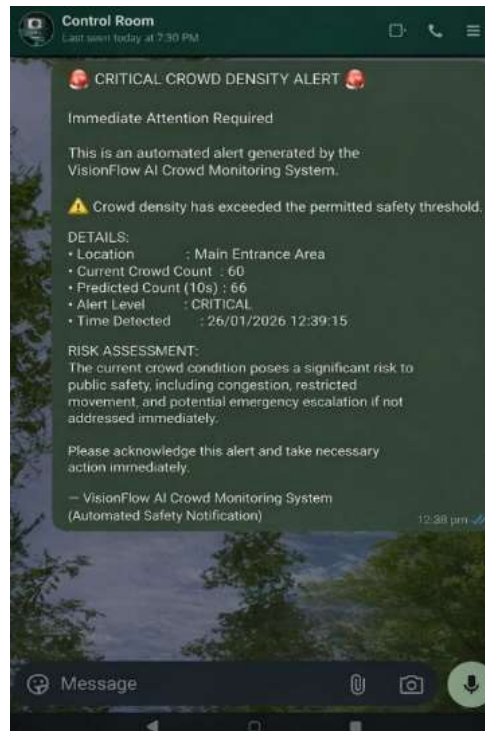


Figure 8: Automated critical-level alert issued when the crowd density exceeds the safety threshold

8.3 Dashboard Resilience and System Latency

The results indicate high dashboard stability and minimal system response time due to acceptable processing delay, as well as adequate refreshing for both population estimation and notification from the interface. Provided a smooth display on wall-mounted (Flask-created) monitors without appreciable lag time. Demonstrated appropriate use of the proposed model for real-time indoor crowd monitoring with a laptop camera in controlled conditions.

9. Discussion

This document outlines a system for archiving live video from a computer's webcam that can identify individuals in near real-time. It emphasises reliable operation under normal conditions, ensuring accurate information over extended periods. In emergencies, the system automatically notifies when population density exceeds thresholds, offering timely alerts and managing crowd situations effectively, including support for partially obstructed views and varying lighting. The solution boasts low hardware costs and enables real-time group monitoring, though it has limitations such as reliance on a single camera and potential visibility issues in dense crowds or poor lighting. Future enhancements may involve additional sensing technologies and behaviour pattern analysis to improve effectiveness.

10. Real-World Applications and Use Cases

The proposed technology enables ongoing video surveillance of public infrastructure to enhance safety management. It provides real-time video feeds for government agencies to monitor population density in specific areas, aiding law enforcement in preventing overcrowding. Registered participants at events can receive alerts about optimal departure times based on expected attendance, helping to manage traffic and crowd distribution. However, challenges such as environmental conditions, camera placement,

obstructions, and hardware limitations hinder implementation. The main objective is to optimise real-time functionality while reducing false alarms, particularly in low-light and low-processing-power situations.

11. Security, Privacy and Ethical Issues

There is no retention of recorded data in the system, as crowd density is estimated in real-time through image analysis from a laptop webcam, with data processed instantly and not stored. The live monitoring interface is accessible only to authorised personnel through secure authentication, preventing unauthorised access. The framework aims to enhance public safety and effective crowd management while simultaneously protecting personal privacy and complying with local data protection laws. Appropriate limits on notifications and response times must be set to ensure proportional interventions.

12. Conclusion and Future Works

This paper presents a framework for a real-time monitoring system utilising deep-learning techniques for facial image analysis to assess crowd levels at events. This system enhances safety by providing immediate information, enabling timely responses to crowd control issues. Future enhancements may include integrating additional sensors for visual monitoring, behaviour pattern analysis to detect anomalies, and improving system reliability in adverse conditions. Overall, the aim is to advance the system's robustness and decision-making capabilities regarding public safety.

13. Acknowledgement

Sincere gratitude is extended to Dr. Shajeni Justin, Assistant Professor, Department of Computer Science, for her invaluable guidance, continuous support, and encouragement throughout the course of this work. Her insightful suggestions and expert advice greatly contributed to the successful completion of this work. Appreciation is also expressed to the faculty members of the department for their support and for providing the necessary resources and academic environment to carry out this work effectively.

The author would like to thank friends and peers for their constant encouragement, constructive feedback, and assistance during various stages of this work. Their support played an important role in overcoming challenges and enhancing the quality of this research.

Gratitude is further extended to everyone who directly or indirectly contributed to the successful completion of this work.

References

1. R. Bandewar and D. Khandelwal, "AI in crowd management and safety in stadium management," *Indian Journal of Computer Science*, vol. 10, no. 4, 2025, Art. no. 175748, doi: 10.17010/ijcs/2025/v10/i4/175748.
2. V. Bilade, C. Fulaware, and A. Gaikwad, "Crowd management surveillance using artificial intelligence and deep learning," Dept. of Computer Engineering, Shatabdi Institute of Engineering and Research, Nashik, Maharashtra, India, 2024. [Online]. Available: <https://www.ijirmeps.org/papers/2024/2/230615.pdf>
3. A. Gupta, S. Raj, S. K. Upadhyay, V. Kandu, and S. Singh, "An artificial intelligence mechanism to monitor and manage crowd," *The Journal of JPT*, vol. 45, no. 2, 2025, doi: 10.52783/tjjpt.v45.i02.6290.

4. A. A. Alsabei, T. M. Alsubait, and H. H. Alhakami, “Enhancing crowd safety at Hajj: Real-time detection of abnormal behaviour using YOLOv9,” *IEEE Access*, 2024. [Online]. Available: <https://ieeexplore.ieee.org/document/10902052>
5. P. Siva *et al.*, “Smart surveillance systems using YOLOv8: A scalable approach for crowd and threat detection,” 2024. [Online]. Available: <https://www.researchgate.net/publication/390920968>
6. A. N. Alhawsaw, S. D. Khan, and F. U. Rehman, “Enhanced YOLOv8-based model with context enrichment module for crowd counting in complex drone imagery,” *Remote Sensing*, vol. 16, no. 22, Art. no. 4175, 2024, doi: 10.3390/rs16224175.
7. S. D. Sudharson *et al.*, “Proactive headcount and suspicious activity detection using YOLOv8,” *Procedia Computer Science*, vol. 231, pp. 423–430, 2023, doi: 10.1016/j.procs.2023.12.061.
8. S. S. Maharajpet and A. V. Hegde, “Intelligent real-time crowd density estimation for proactive event safety: A machine learning approach,” *Research Review: International Journal of Multidisciplinary*, 2025, doi: 10.36548/rrrj.2025.2.005.
9. S. Choudhari, B. Rajani, M. Sahajramani, N. Nanwani, S. Nawani, and S. Bhojwani, “Crowd monitoring with alerting system,” *International Journal of Computer Applications in AI*, vol. 6, no. 1, 2025, doi: 10.33545/27076571.2025.v6.i1b.140.
10. K. Fatima and M. R. Ahmed, “Real-time crowd crime and violence detection using deep learning-based face recognition and object detection,” *International Journal of Scientific Research in Engineering and Technology*, 2024. [Online]. Available: <https://www.ijreat.com>