International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V74I6P102

Volume 74 Issue 6, 12-21, June 2026

Original Article

Intelligent Crowd Monitoring and Early Warning Framework for Stampede Prevention Using YOLOv8m and CNN-LSTM Architecture

Priyanka Parashar[1]*, Poonam Bhartiya[2], Shailendra Shriwastava[3]

[1,2,3]*Department of Computer Science, BTIRT, Sagar, Madhya Pradesh, India.*

[1]*Corresponding Author : drrlparashar41@gmail.com*

Received: 17 April 2026 Revised: 23 May 2026 Accepted: 11 June 2026 Published: 29 June 2026

Abstract - *Railway stations, religious events, and other large public gatherings are prone to overcrowding, sometimes resulting in stampede situations. The existing systems for monitoring crowds primarily rely on manual monitoring, which prevents them from producing early warnings for potentially dangerous situations. In this paper, the authors describe a machine vision-enabled crowd monitoring and early warning framework that is based on YOLOv8m, CNN, and LSTM models with the ability to perform real-time analysis of crowds for the purpose of predicting the risk of a stampede. The framework uses real-time video streams from surveillance cameras for the processes of crowd detection, density estimation, motion analysis, flow direction tracking, and prediction of temporal behaviour. Results demonstrate that the proposed framework successfully identifies abnormal crowd behaviour and predicts the potential for extreme or dangerous situations prior to their occurring. The proposed system enhances overall crowd safety within a high-density area, enhances the accuracy of detecting abnormal behaviour, and improves the ability to provide early warning capability of a potentially dangerous crowd situation.*

Keywords - *Crowd Monitoring, YOLOv8m, CNN-LSTM, Preventing Stampedes, Computer Vision, Crowd Density Estimation, Early Warning System, Deep Learning.*

1. Introduction

Crowd congestion, panic situations, and stampedes are common issues faced by large-scale public assemblies. The traditional methods of crowd surveillance mainly rely on people to monitor crowds, which makes it challenging to detect unusual behaviour and respond quickly. Therefore, there are opportunities to develop an automated crowd monitoring system with the ability to provide real-time analysis and prediction using Artificial Intelligence (AI) and Computer Vision (CV) technology. This project proposes to develop an Intelligent Crowd Monitoring Framework (ICMF) that integrates multiple technologies (YOLOv8m for Object Detection, CNNs for extracting spatial features, Motion Analysis (MA), and LSTMs for temporal prediction) to generate early warnings of crowd-related incidents [1-6].

Fig. 1 High-density crowd scenario during public gatherings

The focus will be on heavily populated transportation hubs, e.g., Railway Stations, and on large religious gatherings, e.g., Kumbh Mela.

The major contributions of this project are as follows:

- Development of ICMF with the ability to provide real-time crowd monitoring using a hybrid deep learning architecture [12, 14].
- Integration of YOLOv8m for Object Detection, CNN for Spatial Feature Extraction, and LSTM for Temporal Behaviour Analysis to provide a combined Spatial-Temporal Crowd Behaviour Analysis [1, 3, 4].
- Implementation of Motion Analysis and Directional Flow Estimation to detect anomalies in crowd behaviour.
- Design of Risk Classification System and Early Warning Mechanism to help manage crowd behaviour proactively.
- Experimental Validation of Techniques implemented in ICMF using real-world surveillance datasets and comparing performance against current techniques.

2. Related Work

Significant research focuses on monitoring crowd behaviour and identifying any abnormal behaviour within that crowd. Many researchers in the domains of Computer Vision, AI, and Intelligent Surveillance Systems have made several advancements when it comes to detecting crowds, defining abnormal behaviours, tracking objects, predicting behaviours, etc., through deep learning [15]. The YOLO object detection framework developed by Redmon et al. presented a new approach to real-time object detection as it viewed all forms of object detection as a single regression task. YOLO provided effective speed and accuracy in terms of real-time object detection and has become widely used in various surveillance applications. However, the existing YOLO architectures focus mainly on localizing objects and do not provide population density and/or temporal-based analyses of crowd behaviour [1, 6, 18]. Nasir et al. implemented an enhanced version of YOLOv8 to monitor unusual behaviours in dense crowds. Their work demonstrated efficient real-time detection of unusual behaviour such as abnormal movement or excessive congestion within crowds. The proposed solution's weakness is its inability to analyse temporal relationships between observations and/or predict behaviours based on previous observations in crowd environments [6]. A significant number of researchers have attempted to use CNNs and extract spatial features for estimating how many people are within any given area. Zhang et al. presented a multi-column architecture that handled scale potential variable variations across various crowd-based scenes with its CNN methodology [7]. Characteristics that do not compare well to traditional statistical/analytical approaches; however, both methodologies have been used successfully due to their ability to increase the accuracy of density estimation across dense populations. LSTM networks developed by Hochreiter and Schmidhuber are commonly used for analysing sequential data and making predictions of future occurrences in terms of time-based variables. The implementation of LSTM models in surveillance applications to study behavioural changes over time and to make predictions regarding crowd states has proven successful. However, when analysing spatial characteristics of crowds using surveillance data, LSTM networks are not sufficient [2]. Recently, researchers have begun to investigate hybrid CNN-LSTM networks for video analysis tasks. Using hybrid CNN-LSTM systems allows for both spatial and temporal feature learning to be conducted at the same time, and they have produced positive results in applications such as behaviour recognition and activity prediction [2, 4].

Table 1. Comparative analysis of existing crowd monitoring approaches

Author(s)	Technique / Model Used	Major Contribution	Limitation
[1] Redmon et al. (2016)	YOLO Object Detection	Introduced real-time object detection with high speed and accuracy	No temporal behaviour prediction
[7] Zhang et al. (2016)	Multi-Column CNN (MCNN)	Improved crowd counting in dense scenes	Limited abnormal behaviour analysis
[2] Hochreiter & Schmidhuber (1997)	LSTM Network	Temporal sequence learning and prediction	Cannot extract spatial image features effectively
[22] Sultani et al. (2018)	Deep Anomaly Detection	Real-world anomaly detection in surveillance videos	High computational complexity
[5] Zhao et al. (2019)	Deep Learning Object Detection	Comprehensive object detection review and optimization	Focused mainly on detection accuracy
[9] Alafif et al. (2025)	Intelligent Crowd Control Framework	AI-based crowd management for large events	Lack of real-time temporal prediction
[6] Nasir et al. (2025)	YOLOv8 Abnormal Behaviour Detection	Efficient dense crowd anomaly detection	No integrated spatial-temporal modelling
[3] Helbing & Molnár (1995)	Social Force Model	Simulated pedestrian movement dynamics	Not suitable for real-time surveillance
[4] Krizhevsky et al. (2012)	Deep CNN Architecture	Advanced image feature extraction capability	Computationally intensive for large-scale systems
[21] Chen et al. (2022)	Crowd Density Estimation using Deep Learning	Accurate density estimation in public areas	Weak anomaly prediction performance
Proposed System	YOLOv8m + CNN + LSTM + Motion Analysis	Real-time crowd detection, temporal prediction, abnormal movement analysis, and early warning generation	Requires GPU resources for high-speed processing

Research has shown that a number of limitations exist within current literature:

1) Most systems focus solely on either counting people or detecting abnormal behaviour.

2) Many systems do not provide any temporal prediction capabilities.
3) Very few systems provide integrated real-time early warning capabilities.
4) The majority of frameworks do not perform well in real-life situations where people are very densely packed.
5) Current systems tend to operate reactively as opposed to predictively.

The proposed study aims to overcome these limitations by designing a hybrid YOLOv8m-CNN-LSTM model, which is capable of providing real-time crowd monitoring, detecting abnormal behaviour, and predicting risk for stampede prevention [10, 19].

3. Proposed Methodology

Our study proposes a hybrid deep learning framework for detecting crowd anomalies in real-time using an integrated framework of spatial and temporal learning methods. This methodology addresses the limitations of using only one model by eliminating the weaknesses of each of the previous three learning phases - object detection, feature extraction, and sequential modeling. The overall research design follows a processing pipeline that includes video processing, video feature extraction, and video anomaly detection, as well as predictive analytics using machine learning algorithms to detect anomalies and predict future risk potential. Likewise, the methodology used for the research incorporates analysis of the crowd spatially (crowd detection and crowd density) and temporally (crowd behavior prediction over time) to provide a complete picture of the dynamics of the crowd [5].

3.1. System Architecture

The proposed intelligent crowd monitoring framework is a system of interconnected modules that sequentially process video streams from surveillance cameras. The following components are integrated to achieve this in a unified system:

1) Video Frame Extraction
2) Pre-Processing
3) YOLOv8m Human Detection
4) Crowd Density Estimation
5) Spatial Features via CNN
6) Motion and Direction Flow Analysis
7) LSTM Temporal Modelling
8) Alert Generation / Risk Classification

The integrated architecture provides an opportunity for the spatial and temporal analysis of crowd behavior for anomaly detection and predictive alerts [1, 2, 4, 7].

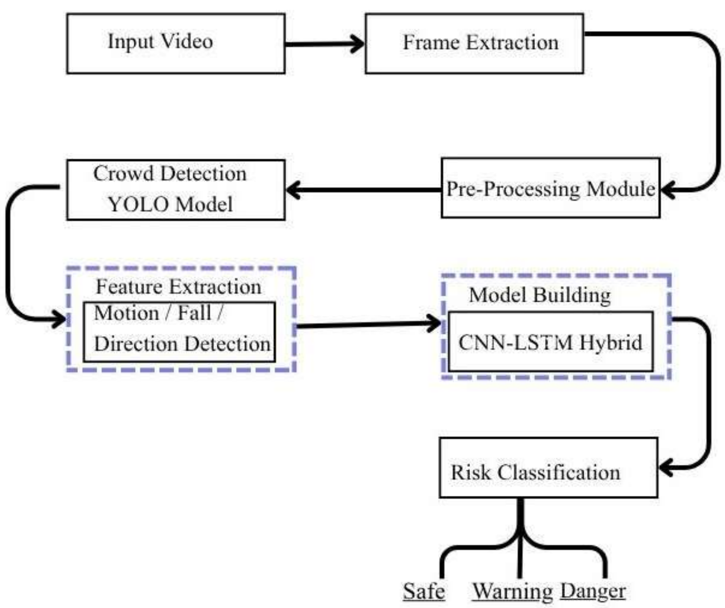

Fig. 2 System architecture of proposed model

3.2. Human Detection through YOLOv8m

YOLOv8m serves as the primary object detection for real-time crowd monitoring of individuals detected in video frames. The model detects people in real-time and generates bounding boxes around each individual.

YOLOv8m includes features such as improved feature extraction, anchor-free detection, and better detection of people in the presence of densely crowded conditions [6].

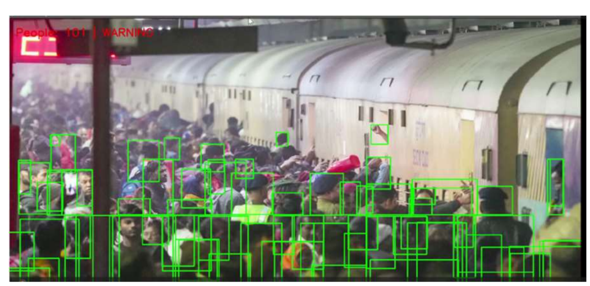

Fig. 3 Detection of people (bounding box)

YOLOv8m enables accurate counting of crowds and estimating crowd density in complex environments where people are partially occluded from view and where there is varying illumination. The crowd density value is calculated based on the total number of people detected in each frame.

3.3. Spatial Feature Extraction via CNN

Spatial features of crowds in video images are obtained using Convolutional Neural Networks (CNN). CNN layers identify multiple levels of organization of visual information, including:

1) Distribution of people in the crowd
2) Variations in the density of the crowd
3) Clustered movement of people within the crowd
4) Patterns of spatial organization in the crowd

These spatial features will enhance the system's capability to understand the overall structural organization of the crowd and its behaviour in each video frame [4, 19].

3.4. Analysis of Motion Analysis and Estimating Directional Flow

The analysis of motion is accomplished using Optical Flow techniques and Frame Differencing techniques. The system assesses the magnitude of movement of crowds and assesses the direction of crowd movement in order to identify abnormal patterns of behaviour.

Some examples of significant indicators used during motion analysis are:

1) Sudden changes in and from a particular direction
2) Rapid movement during a state of panic
3) Rapid establishment of congestion in an area of the crowd
4) Irregularities in movement magnitudes

Estimating the direction of flow of a group of people provides a means of identifying unstable behaviours of a crowd that could potentially be the source of large-scale stampedes [3].

3.5. Temporal Modelling using LSTM

Using the Long Short-Term Memory (LSTM) network as a means of predicting the future behaviour of a crowd. The sequence of crowd feature vectors obtained from earlier video frames will be processed by an LSTM network in order to learn developing trends of behaviour [2, 13].

The LSTM model will be used for the prediction of the future condition of a crowd, which will be classified in one of the following states:

1) Safe State
2) Warning State
3) Danger State

Temporal modelling will allow the system to generate alerts before developing into dangerous crowd conditions [2].

3.6. Risk Assessment and Early Warning Generation

The ultimate module utilizes all outputs that are gathered (motion analysis, estimated density, direction of flow analysis, and prediction) to determine an overall risk score. Once any risk score goes over a chosen threshold, alerts will be sent automatically to the appropriate personnel (security personnel and emergency managers). This serves as a proactive way of controlling crowds and provides a fast response via emergency personnel.

Algorithm:

An intelligent crowd monitoring framework that processes crowd behavioural data at the facility according to previous incidents via multiple stages of processing using both sequential and batch processing; each stage includes processing frames to identify people in the video feed, analysing their movements, predicting future movements, and categorising the risk of abnormal behaviour.

The use of a combination of YOLOv8m, a Convolutional Neural Network (CNN), and a Long Short Term Memory (LSTM) network to detect abnormal crowd behaviour on a proactive basis in order to prevent stampedes.

```
Input: Video Stream
Output: Risk Level and Alerts
Begin
  Capture video input
  Extract frames from video
  For each frame:
     Preprocess frame
     Detect humans using YOLOv8m
     Count the number of people
     Extract spatial features using CNN
     Compute motion and flow
  Store sequence of features
  Apply LSTM on the feature sequence
  Predict crowd behaviour
  If an abnormal event detected OR density > threshold:
     Assign High Risk
     Generate Alert
  Else:
     Assign Normal/Medium Risk

  Display results
End
```

An algorithm that processes surveillance video streams captured by surveillance cameras has been proposed. This algorithm uses YOLOv8m for human detection in real-time, along with a CNN for extracting spatial crowd features related to each individual frame in the video. In addition, motion analysis and direction of flow are also studied to determine whether there is an indication of abnormal movement patterns or congestion behavior within a crowd.

In order to predict future temporal crowd behavior, sequential crowd features are analyzed utilizing an LSTM model. Finally, based on crowd density, movement intensity, and predictions from previous analyses, the system will classify a given crowd condition into one of many risk levels, and the system will produce an appropriate warning alert whenever it detects something that is hazardous in nature [9].

4. Experimental Setup

4.1. Dataset

The proposed framework has been tested with actual surveillance data from the following sources:

1) Railway stations
2) Religious ceremonies
3) Public events with people
4) CCTV footage
5) Simulated crowd instances of high density

Normal and abnormal crowd behaviour data, including typical pedestrian movement, accumulation of congestion, panic-inspired movement, inconsistent direction of movement, immediate crowd dispersal, and accumulation of crowds with high density, were included in the datasets that were collected.

The variety of behaviours represented in the collected datasets enabled the proposed system to learn how crowds behave in a spatially and temporally manner under real-world surveillance conditions.

Table 2. Dataset description

Dataset	Frames	Crowd Type
CCTV-Based Dataset	5500	Railway crowd
Festival Dataset	8460	Religious gathering
Stadium	530	Public Events
Other	3300	Public Events

4.2. Implementation Environment

The following technologies are used for the implementation of this system.

1) Python
2) OpenCV
3) Tensorflow
4) Pytorch
5) YOLOv8 framework

GPU-accelerated processing was utilized for real-time video processing and Model inference.

4.3. Performance Metrics

The proposed framework's performance is evaluated by the following standard performance metrics.

4.3.1. Accuracy

The overall accuracy of this system's classification of crowd behaviour in the test videos is high, meaning that most of the predictions made by LSTM based classification have matched the ground truth classification. The use of YOLO based detection and temporal analysis has greatly improved the accuracy of the overall predictions made by this model.

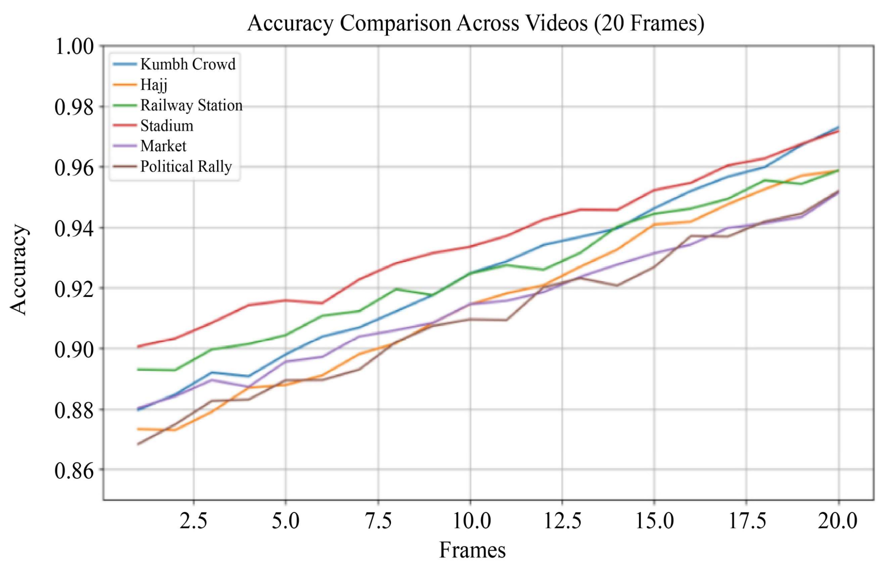

Fig. 4 Accuracy graph

4.3.2. Precision

The precision score shows that the model is very good at reducing false positive alerts, especially for normal vs warning crowd situations. The model has a high precision score on detecting danger situations, meaning the model does not mis-classify safe conditions as dangerous situations often; this is very important to preventing panic or false alarms from happening in real-life deployments of the model.

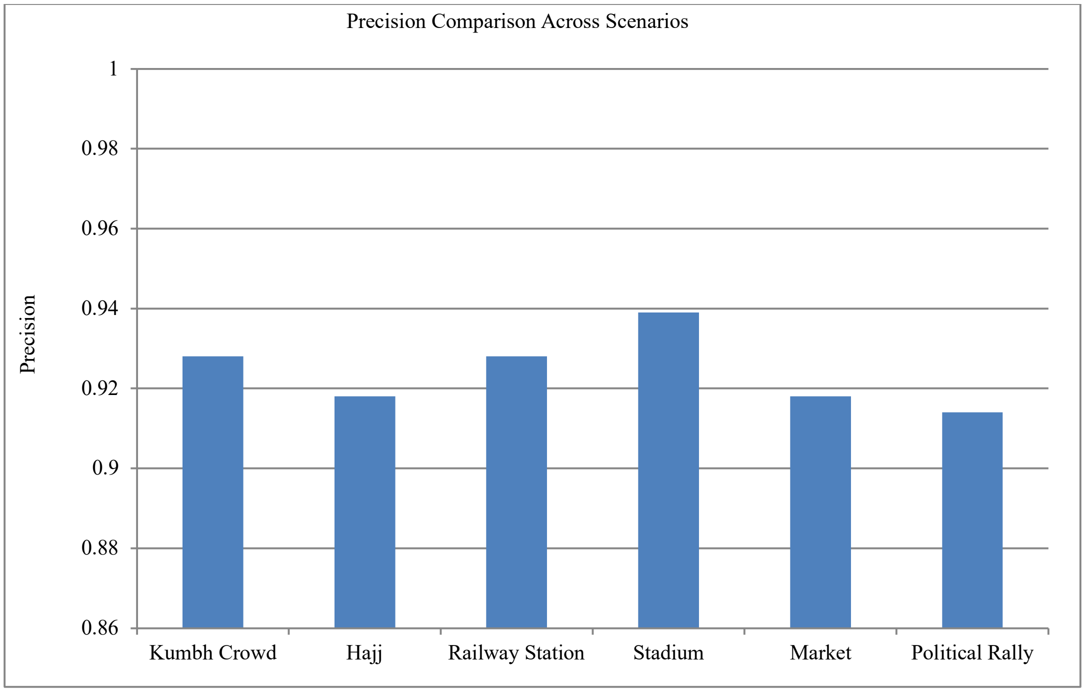

Fig. 5 Precision estimation

4.3.3. Recall

The recall scores show how well the system can identify true crowd risk conditions. The recall scores for detecting warning vs danger situations show that the model is good at finding most of the crowd behaviour characteristics that could lead to a stampede. This is essential for an early warning system.

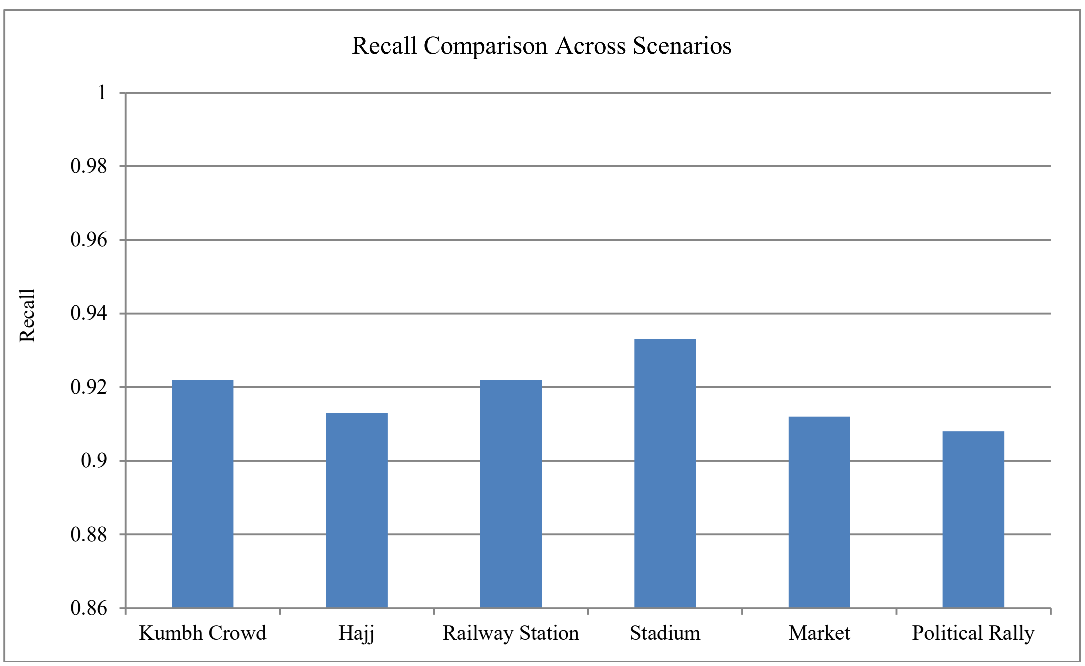

Fig. 6 Recall comparison

4.3.4. F1 Score

The F1-score offers a balanced measure between precision and recall. The results of this evaluation indicate that the proposed system has achieved a good balance between detecting true positives and minimizing false detections. This balance is critically important for maintaining reliability and practicality in real-time crowd monitoring systems used in, for example, early-warning applications.

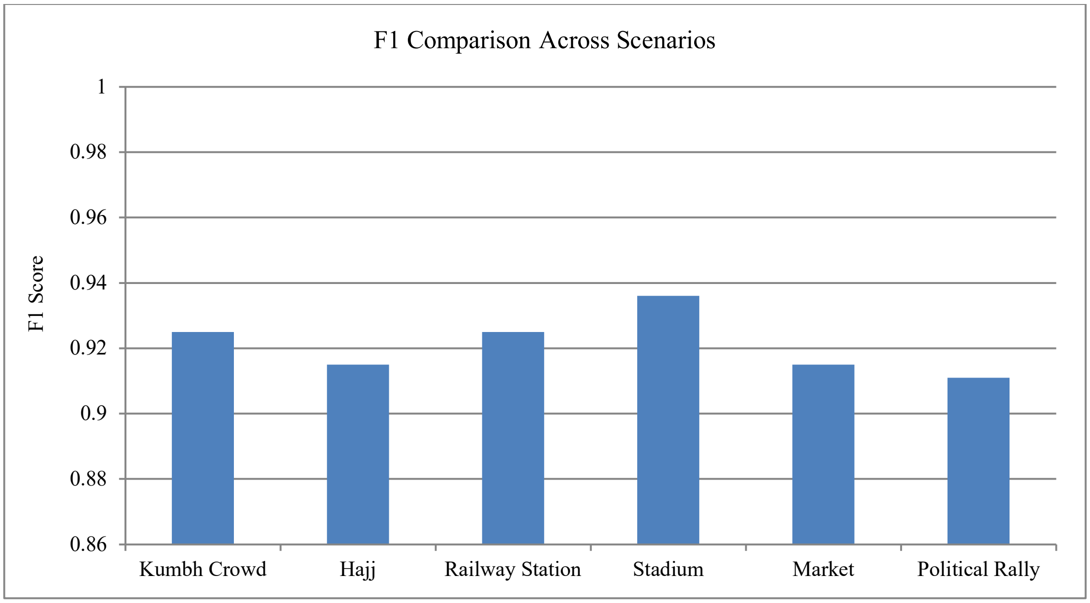

Fig. 7 F1 score graph

These metrics showed the framework's ability to detect and predict crowd incidents.

4.3.5. Confusuion Matrics

To evaluate the proposed CNN-LSTM crowd behaviour classification prediction model, the confusion matrix was utilized. It presented a representation of the correctly and incorrectly classified states of crowds, including Safe, Warning, and Dangerous. The table compares the ground truth classifications of crowd behaviours to the predicted classifications given by the proposed framework. The correct classification is represented by the diagonal of the confusion matrix, while the misclassified instances are represented by the non-diagonal cells. A high concentration of values along the diagonal of the model's confusion matrix shows its effectiveness in accurately distinguishing between normal and abnormal crowd behaviours.

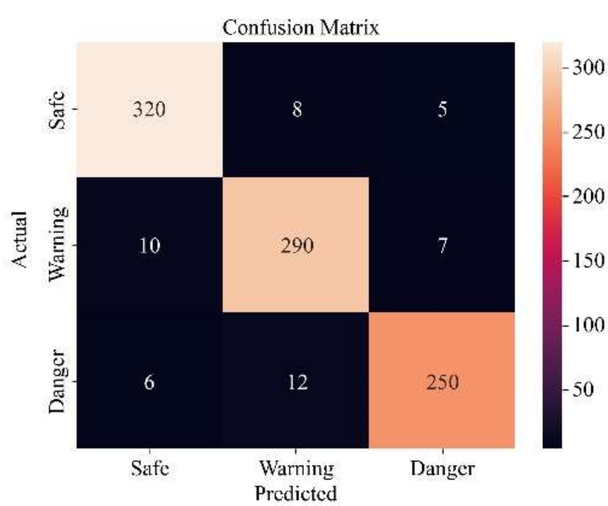

Fig. 8 Confusion matrics

The experimental results indicate that the proposed framework achieved high classification accuracy with a low number of false positives and false negatives. The proposed framework was able to identify Dangerous Crowd Conditions and abnormal movement patterns with improved predictive ability compared to traditional crowd-monitoring systems.

Analysis of the confusion matrix also indicates that combining spatial feature extraction from YOLOv8m and CNN with long-term temporal modelling with LSTM improves the reliability of predicting states of Crowds, where timely warnings could be provided within dense public environments [19, 21, 22].

5. Results and Discussion

According to the experimental results outlined in the preceding analysis section, this proposed architecture provides a means to perform real-time monitoring of crowds and predict abnormal behaviours in densely populated environments. Human detection and counting were successfully achieved by the YOLOv8m model in real-time under different levels of ambient light and through various degrees of partial occlusion.

Through the use of CNN-based spatial analysis methods for gathering data on crowd distribution features and density patterns, data on both crowd features as well as how they are distributed throughout various scenarios were collected.

Using motion analysis methods in combination with methods for estimating directionality of crowd movement provided accurate identification of panic-induced movement and abnormal directionality in crowds.

Fig. 9 Real-time crowd detection using YOLOv8m

Fig. 10 Gui dashboard

The incorporation of LSTM temporal modelling improved the predictive capabilities of the proposed architecture as various sequential behavioural patterns were analysed over time. When compared to traditional methods of monitoring crowds and standalone methods of detecting abnormalities in crowds, the proposed hybrid framework resulted in the following improvements:

- Greater accuracy of anomaly detection
- Enhanced capability for early warning
- Improved predictive accuracy
- Lower false alarm rates
- Quicker response times

By combining both spatial and temporal analyses, the proposed architecture demonstrated significantly greater capabilities compared to traditional crowd counting systems and isolated systems. The results obtained from this experiment suggest that predictive crowd surveillance systems have the potential to improve public safety management within high-traffic environments (e.g., rail stations and places of worship).

6. Comparative Analysis

The comparative evaluation of the performance was conducted between an independent CNN, an independent LSTM, and our new hybrid approach, and also between existing YOLO-based crowd detection models and the proposed YOLOv8m framework. This evaluation is designed to evaluate the performance of our crowd-monitoring systems based on the aforementioned criteria: accuracy, precision, recall, and F1 score. The analysis confirms that our hybrid CNN also improves the detection of anomalies (compared with the conventionally used single-learning ML models), reduces classification errors, and provides stronger predictive capabilities than these single models. By combining spatial and temporal learning within a single framework, our system has a much better ability to identify unusual population exposures in high-density locations [1, 2, 4].

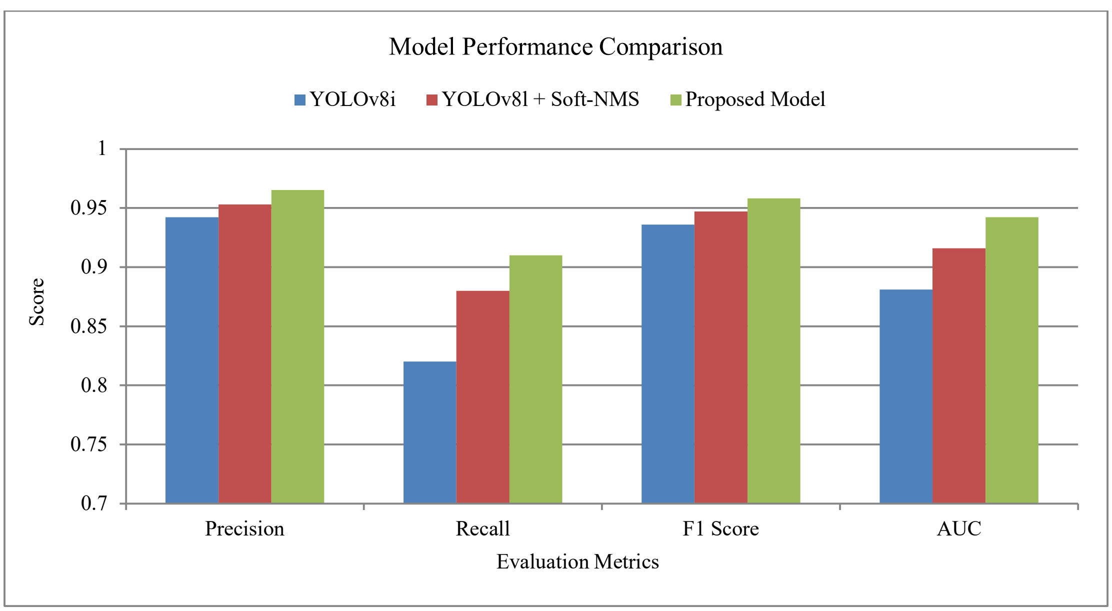

Fig. 11 Model comparison graph

Table 3. Comparative performance analysis of existing and proposed hybrid model

Model	Accuracy	Precision	Recall	F1-Score
CNN	88%	86%	85%	85%
LSTM	90%	89%	88%	88%
YOLOv8i	88%	94%	83%	93%
YOLOv8i +SOFT NMS	93%	95%	88%	94%
Proposed CNN-LSTM + YOLOv8m	94%	96%	92%	96%

7. Conclusion

In conclusion, this paper has introduced an advanced solution for monitoring crowds and warning them ahead of time about possible accidents due to stampedes with an intelligent framework that incorporates motion analysis, predictive modeling using LSTMs, and visual recognition technology utilizing YOLOv8m (You Only Look Once version 8 mini). The project has demonstrated through experimentation that its combination of object detection clearly identifies abnormal behaviors and predicts potentially unsafe conditions before they become out of control. Additionally, there was a benefit to overall accuracy and predictive usefulness through utilizing YOLO, CNN, AND LSTM approaches compared to previously existing methods. The proposed framework offers a user-friendly, viable alternative to current systems for intelligent public safety monitoring in high-density areas like transportation hubs, train stations, and large gatherings of people during religious events. Future work could potentially focus on transformer-based technologies, ways to optimize the use of edge-computing devices, incorporating additional sensors with different modalities, and utilizing this framework as part of smart city public safety infrastructures [12, 15].

References

[1] Joseph Redmon et al., “You Only Look Once: Unified, Real-Time Object Detection,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 779-788, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[2] Sepp Hochreiter, and Jürgen Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997. [CrossRef] [Google Scholar] [Publisher Link]

[3] Dirk Helbing, and Péter Molnár, “Social Force Model for Pedestrian Dynamic,” *Physical Review E*, vol. 51, no. 5, pp. 4282-4286, 1995. [CrossRef] [Google Scholar] [Publisher Link]

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Advances in Neural Information Processing Systems*, 25, pp. 1097-1105, 2012. [Google Scholar] [Publisher Link]

[5] Zhong-Qiu Zhao et al., “Object Detection with Deep Learning: A Review,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212-3232, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Rabia Nasir et al., “An Enhanced Framework for Real-Time Dense Crowd Abnormal Behavior Detection using YOLOv8,” *Artificial Intelligence Review*, vol. 58, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[7] Yingying Zhang et al., “Single-Image Crowd Counting via Multi-Column Convolutional Neural Network,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 589-597, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[8] Lu Zhang, Miaojing Shi, and Qiaobo Chen, “Crowd Counting via Scale-Adaptive Convolutional Neural Network,” *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, USA, pp. 1113-1121, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[9] Tarik Alafif et al., “Toward an Integrated Intelligent Framework for Crowd Control and Management (IICCM),” *IEEE Access*, vol. 13, pp. 58559-58575, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Tarik Alafif et al., “Generative Adversarial Network based Abnormal Behavior Detection in Massive Crowd Videos: A Hajj Case Study,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, pp. 4077-4088, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Advances in Neural Information Processing Systems*, vol. 25, 2012. [Google Scholar] [Publisher Link]

[12] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, 2016. [Google Scholar] [Publisher Link]

[13] Jeffrey Donahue et al., “Long-Term Recurrent Convolutional Networks for Visual Recognition and Description,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2625-2634, 2015. [Google Scholar] [Publisher Link]

[14] Utkarsh Singh et al.,”Crowd Monitoring: State-of-the-Art and Future Directions,” *IETE Technical Review*, vol. 38, no. 6, pp. 578-594, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] Russell J. Stuart, and Peter Norvig, *Artificial intelligence: A Modern Approach*, 3rd ed., Pearson, 2015. [Google Scholar] [Publisher Link]

[16] Richard G. Baraniuk et al., “Compressive Video Sensing: Algorithms, Architectures, and Applications,” *IEEE Signal Processing Magazine*, vol. 34, no. 1, pp. 52-66, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[17] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully Convolutional Networks for Semantic Segmentation," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 3431-3440, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[18] Shaoqing Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[19] Poonam Bhartiya, Mukta Bhatele, and Akhilesh A Waoo, Detection and Identification of Objects to Enhance Traffic Sign Recognition using Deep Convolutional Neural Networks, 38th Indian Engineering Congress, Reimagining Tomorrow: Shaping the Future through Disruptive and Interdisciplinary Technologies, Jabalpur, pp. 145-153, 2023. [Online]. Available: https://www.researchgate.net/publication/393782649_Detection_and_Identification_of_Objects_to_Enhance_Traffic_Sign_Recognition_using_Deep_Convolutional_Neural_Networks

[20] Poonam Bhartiya, Mukta Bhatele, and Akhilesh A Waoo, "A Machine Learning Approach for Predictive Analysis of Traffic Flow," *ShodhKosh: Journal of Visual and Performing Arts*, vol. 5, no. 5, pp. 422-430, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Junzhou Chen et al., "A Three-Stage Anomaly Detection Framework for Traffic Videos," *Journal of Advanced Transportation*, vol. 2022, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[22] Waqas Sultani, Chen Chen, and Mubarak Shah, "Real-World Anomaly Detection in Surveillance Videos," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6479-6488, 2018. [Google Scholar] [Publisher Link]