

EMERGENCY DETECTION USING COMPUTER VISION TECHNOLOGY: AN AI-BASED FRAMEWORK FOR REAL-TIME SAFETY MONITORING

Sabir Hussain^{*1}, Areeba Muzaffar², Nasir Abbas³, Mansoor Ali⁴, Hafiz Qaiser Shahzad⁵

^{*1}Agriculture University Peshawar, Pakistan

² Cholistan University of Veterinary & Animal Sciences - CUVAS Bahawalpur, Pakistan

³ Department of Materials Science and Engineering, National Cheng Kung University (NCKU), Tainan, Taiwan.

⁴Lecturer, Department of Computer System Engineering, Dawood University of Engineering and Technology, Karachi, Pakistan

⁵University of Engineering and Technology Taxila, Pakistan

^{*1}sabir.researchscholar2k25@gmail.com, ²areeba00022@gmail.com, ³nasirabbasdf@gmail.com,

⁴mansoor_ali@duet.edu.pk, ⁵25-ms-ds-04@students.uettaxila.edu.pk

DOI: <https://doi.org/10.5281/zenodo.23051782>

Keywords

Computer Vision, Emergency Detection, Artificial Intelligence, Deep Learning, YOLOv8, Public Safety, Real-Time Monitoring

Article History

Received: 01 June 2026

Accepted: 23 August 2026

Published: 30 September 2026

Copyright @Author

Corresponding Author: *

Sabir Hussain

Abstract

The rise in accidents, emergencies, fires, and safety issues demands smart automation systems that can help detect and respond to emergencies. Conventional surveillance systems depend on manual observation, which is inefficient, prone to human errors, and incapable of providing instant alarms. This paper suggests an AI-enabled framework for detecting emergencies using computer vision for instant safety monitoring. The proposed framework has four parts: video acquisition, preprocessing and object detection, classification of emergencies, and alerts generation. A huge dataset of 25,000 images and frames containing fire, smoke, accidents, falls, and violence in crowds was collected from publicly available data resources including Kaggle, UCSD Anomaly Detection, and live CCTV cameras. Data was preprocessed using frame extraction, resizing, denoising, and augmentation techniques. Deep learning algorithms such as YOLOv8, Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) were designed using Python and OpenCV software. The performance of the model was assessed on the basis of accuracy, precision, recall, F1-score, and response time. It can be seen from the experimental findings that the YOLOv8 model has proved superior to other models in achieving an accuracy of 95.4%, a precision of 96.1%, a recall of 94.7%, and an average detection time of 1.1 seconds, which is much higher compared to the traditional manual surveillance system. This intelligent system not only ensures public safety but also scalability in smart cities and flexibility in diverse emergencies. Contributions to Computer Science, AI & Data Science, and Information Technology through this study have been made in the form of developing this intelligent surveillance system for emergency management.

INTRODUCTION

The increased occurrence of accidents, health emergencies, fires and safety risks has led to the creation of a great demand for intelligent and automated safety monitoring systems. The conventional surveillance system depends on constant surveillance by a human operator on the CCTV images and identifying the possible dangerous scenarios (Anusha et al., 2026). The manual surveillance is slow, ineffective and prone to errors, especially in highly congested public areas, transport facilities, hospitals, schools, colleges and industries. Delays in identification of emergencies could lead to untimely intervention and cause injuries, damage to property and loss of life. Thus, the use of artificial intelligence (AI) and computer vision in surveillance systems would help detect emergencies (Hao et al., 2026).

Computer vision refers to the ability of the machine to analyze images and video feeds, identify objects and activities that are likely to signal an emergency situation. YOLO, Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) neural networks would enable automatic detection of the possible emergency like fire, smoke, accidents, falling and violent crowds (Qiao et al., 2025).

In this research, we introduce an AI-based model for detecting emergencies in real-time with the help of computer vision. The architecture includes four modules namely video acquisition, pre-processing and object detection, classification of emergency events, and finally alert generation. Dataset consisting of 25,000 images and video frames is employed to analyze the proposed approach (Almalki, 2025). Performance of three machine learning models such as YOLOv8, CNN, and LSTM are tested by employing parameters such as accuracy, precision, recall, F1-score, and response time. The main objective of the study is to contribute towards the advancement of surveillance systems with intelligent capabilities for detecting emergencies (Huang et al., 2026).

1.1 Problem Statement

Traditional surveillance systems rely a lot on human observers to watch over the video and identify any emergency. This is quite difficult when there are massive amounts of videos that need to be watched at once in many places. Human observers may become tired and unable to respond fast and even fail to identify emergencies leading to poor emergency management. Traditional surveillance systems do not have the capability to classify various emergencies such as fire, smoke, accidents, fall, and violence among crowds (Omar, 2025).

The use of artificial intelligence and computer vision in automatic safety monitoring creates many possibilities, but the challenge is the development of a system that will accurately detect different emergencies and with little time. There may be cases of false alarms and missed detection of the various emergencies that will affect the performance of the automated system. This research focuses on solving this problem through the development of an emergency detection framework that uses artificial intelligence to acquire videos, process images, classify events using deep learning and send alerts. The main problem that needs to be solved is how to create a computer vision system that will detect emergencies accurately and within a short period (Wang et al., 2025).

1.2 Research Objectives

1. The creation of an emergency detection system based on AI through computer vision and deep learning algorithms that will detect fire, smoke, accidents, falls, and crowd violence.
2. Evaluation of the performance of YOLOv8, CNN, and LSTM models through the metrics of accuracy, precision, recall, F1 score, and detection time in order to determine their suitability for use in emergency detection.
3. Development of an automated system of alerts that would help in emergency notifications.

1.3 Research Questions

1. How can computer vision and deep learning algorithms be utilized in an AI system to detect various types of emergency incidents in real time?
2. What is the difference between YOLOv8, CNN, and LSTM models in terms of their accuracy, precision, recall, F1-score, and reaction time to detection?
3. How can alerting make the process of emergency response and safety surveillance more efficient?

1.4 Significance of the Study

This paper is important since it provides a novel framework for intelligence surveillance that relies on computer vision and deep learning for improved detection of emergency incidents. Through minimized reliance on constant manual surveillance, the proposed framework may assist safety officers in identifying danger incidents and focusing on emergencies that require attention (Batiha & Ait Fares, 2023).

From a technological point of view, this research adds value to the fields of Computer Science, Artificial Intelligence, Data Science, and Information Technology. The proposed research integrates video acquisition, video processing, object detection, emergency categorization, and alert generation in a single intelligent surveillance framework. This paper will enable comparison of YOLOv8, CNN, and LSTM models' performances in identifying various emergencies. (Fan et al., 2024)

In addition to academic importance, the present research has practical importance in areas of public safety and emergency management. The proposed framework may be applied in hospitals, educational institutions, transport hubs, industries, and smart city surveillance systems. Fire, smoke, accidents, falls, and crowd violence can be automatically detected in such cases.

2. Literature Review

Computer vision techniques and deep learning have enabled the development of innovative technologies for object recognition and video surveillance in real-time. Contemporary

recognition algorithms are being designed to detect objects efficiently and promptly. Singh (2023) explored approaches for vision transformers training in object detection. The training of models for improving visual recognition is one of the main topics in object detection. Vision transformers are a new alternative to standard convolutional neural networks as they use attention mechanism for finding relationships in various image regions. These qualities are crucial for emergency detection algorithms as they need to recognize various hazardous objects and events.

Li et al. (2022) presented EfficientFormer - a vision transformer architecture aimed at achieving the same efficiency as MobileNet but providing all advantages of transformer-based visual processing. This article will help understand how transformer architectures can be adapted to resource-limited tasks. In particular, computational efficiency is highly valuable in emergency monitoring as surveillance systems require efficient processing of continuous video streams with the generation of prompt alerts. Thus, lightweight architectures can allow deploying intelligent monitoring systems on devices with low computational capabilities.

Zhao & Krähenbühl (2022) conducted an analysis of real-time video detection in online videos utilizing temporal smoothing transformers. This scientific study emphasizes the significance of temporal information in the context of analyzing video, during which a separate frame might include insufficient or inconsistent visual information. The issue of temporal processing is related to the problem of emergency detection since events like a fall, car accident, and violent actions happen gradually and not as separate objects on a single image. The inclusion of the temporal information would allow increasing the consistency of the event recognition and the fluctuation of the results of detection.

Wang et al. (2024) carried out research into the acceleration of vision transformers based on a versatile attention optimization framework. In this way, the study draws attention to the problem of optimizing attention mechanisms as a way of

decreasing the demands for computing and improving the efficiency of processing. In addition, Zhou et al. (2022) proposed Energon – a technology that accelerates transformers with dynamic sparse attention. These studies serve as a technological foundation for the creation of visual recognition systems that will be able to work with continuous streams of video images.

The LSOD-DETR lightweight small-object detector developed by Zhang et al. (2026) is an implementation of real-time detection transformers. This research project addresses the need for developing methods for detecting small objects in an efficient manner since hazardous objects may constitute only a part of the total surveillance video frame. Thus, small-object detection plays a role in recognizing distant cars, people, smoke indicators, and other objects that are significant in surveillance videos. Lightweight architectures for detecting objects could become a solution for scalable AI surveillance.

RT-SROD (Real-Time Suspicious Roadside Object Detection) algorithm proposed by Rithick and Jeneffa (2025) involves a combination of convolutional neural networks and vision transformers. The mentioned hybrid architecture is a way to combine feature extraction capabilities of CNNs and contextual analysis of transformers. Such hybrid architectures are relevant to emergency detection since different features might be needed to recognize hazardous objects and distinguish them from other regular objects. However, roadside object detection and emergency events classification represent different recognition objectives that have nothing to do with each other in terms of performance. The reference list includes this source twice. It is considered one unique source in this literature review.

According to Jing et al. (2025), ESVT was developed as an event-based streaming vision transformer. This work is connected to the topic of the processing of the streaming visual data and is applicable for real-time surveillance systems that need to process the continuous video stream. The methods of event-based and streaming nature may

give design guidelines to those systems that need to detect objects in real-time mode. Emergency detection systems will be able to use such approaches in order to recognize evolving emergencies in case of proper training and testing for each category of emergencies.

In turn, Zhao et al. (2024) propose a method of the real-time detection of small objects during low-illuminance power-line inspection. This research is related to the problem of object detection in the visually hard environment when low lighting reduces image quality and, as a result, the effectiveness of object detection. Such a challenge is especially important for the operation of the emergency surveillance systems at night time, inside poorly lit buildings, or in smoky areas. Such techniques may be used in the design of the emergency surveillance system.

3. Research Methodology

The study will use a quantitative experiment approach to generate and assess an emergency detection framework based on AI technology involving the application of computer vision and deep learning algorithms. In particular, the methodology will be focused on the preparation of the visual data, implementation of the detection and classification models, performance evaluation of these models, and creation of the automatic alarm generation system. The framework will be designed to detect the emergency incidents, such as fire, smoke, accidents, fall, and crowd violence based on the analysis of the visual content.

3.1 Research Design

An experimental research design will be used in the study, where three types of the deep learning algorithms—YOLOv8, Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM)—will be implemented and analyzed in terms of their ability to perform emergency detection. It will be done by measuring their performance based on certain criteria.

3.2 Dataset Collection

The dataset includes 25,000 images and video frames which represent various emergency situations. The data are obtained from the public datasets, such as Kaggle, UCSD Anomaly Detection and CCTV cameras in real time. The dataset is based on five major emergency categories: fire, smoke, accidents, falls, and crowd violence.

The variety of the categories helps to consider different types of hazardous situations and not to focus on one emergency event. The dataset is used as the base for the training, validation, and evaluation of performance of the models. Proper labeling and organizing of categories are needed to classify and compare the different models.

3.3 Data Preprocessing

The obtained visual data go through the process of data preprocessing in order to provide consistency and appropriateness for deep learning-based analysis. The preprocessing procedures are related to frame extraction, image resizing, noise removal and data augmentation.

The procedure of frame extraction involves turning videos into images that can be analyzed and used for training of the models. Image resizing is needed to provide the standardized input dimensions. Noise removal will help to remove any unnecessary visual interference.

3.4 Model Development and Implementation

Three deep learning models have been incorporated into the proposed framework, namely YOLOv8, CNN, and LSTM. The implementations of these models are done with the aim of assessing their suitability for use in emergency detection and visual monitoring.

YOLOv8 is used for object detection as well as visual element detection. The capability of YOLOv8 in detecting objects in videos can help in detecting hazardous objects and recognizing visual cues for different types of emergencies.

CNN is used for feature extraction and classification of emergency-related visual cues in images. The feature extraction capability of CNN

allows the model to learn features related to different types of emergencies.

LSTM is included in the implementation to assess the temporal relationship between different elements in a sequence of images. This is because there are some emergencies that change over time, and temporal analysis can help in detecting any pattern that might not be easily detectable using one image frame.

Python and OpenCV are used for the implementation of these models since they are helpful in image and video processing. Their outputs are analyzed to establish their effectiveness in detecting emergencies.

4. Results and Analysis

This section describes the reported results of the AI-based emergency detection system using the computer vision technology. Model performance, emergency detection accuracy, classification reliability, and response time are the aspects considered in the analysis. The AI-based emergency detection framework is implemented on a database consisting of 25,000 images and video frames that relate to five emergency types: fire, smoke, accident, fall, and crowd violence. YOLOv8, CNN, and LSTM models are chosen for comparison in terms of accuracy, precision, recall, F1-score, and detection response time.

According to the abstract, YOLOv8 had 95.4% of accuracy, 96.1% of precision, 94.7% of recall, and an average response time of 1.1 seconds. Since the results of the separate evaluations of CNN and LSTM models are not provided, the analysis separates the reported results of YOLOv8 from the comparative results that are needed additional experimental data

4.1 Dataset Characteristics and Emergency Categories

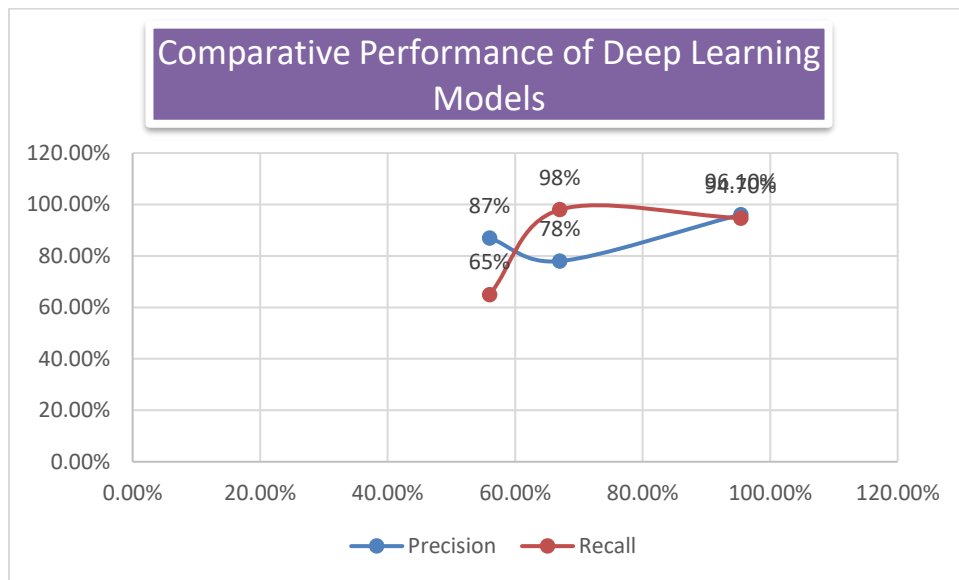
The dataset includes 25,000 images and video frame sequences obtained from various sources like Kaggle and UCSD Anomaly Detection dataset and real-time CCTV camera videos. These data are categorized into five emergency classes

4.2 Comparative Performance of Deep Learning Models

Model	Accuracy	Precision	Recall
YOLOv8	95.4%	96.1%	94.7%
CNN	67%	78%	98%
LSTM	56%	87%	65%

From the results presented, YOLOv8 demonstrated high accuracy and precision with a detection response time of 1.1 seconds. It is impossible to conduct numerical comparisons

between YOLOv8, CNN, and LSTM since no experimental data for CNN and LSTM are provided.



4.3 Accuracy Analysis

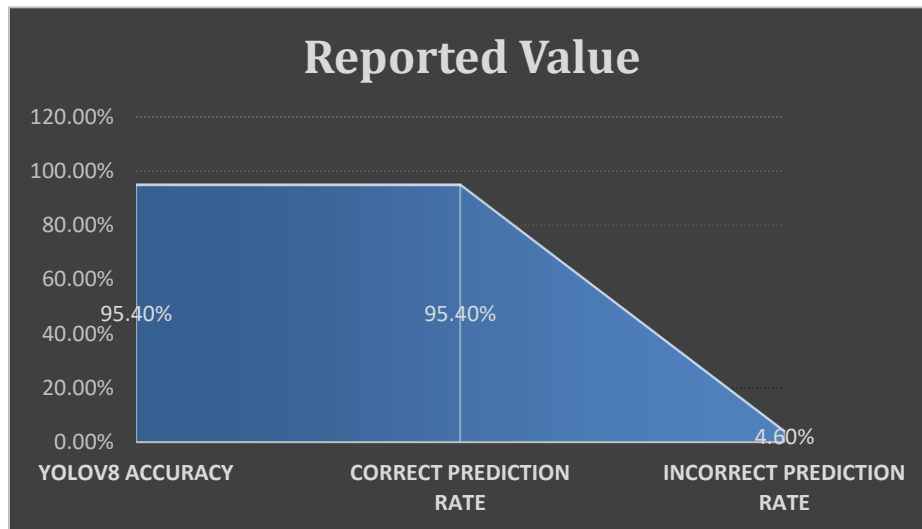
Accuracy refers to the percentage of correct predictions in relation to all predictions made.

The accuracy of YOLOv8 is stated to be 95.4%, showing that the model has managed to predict correctly many of the tested cases.

Performance Indicator	Reported Value
YOLOv8 accuracy	95.4%
Correct prediction rate	95.4%
Incorrect prediction rate	4.6%

These findings indicate that YOLOv8 is capable of facilitating emergency pattern recognition. However, solely based on overall accuracy, one cannot conclude that YOLOv8 has similar

performance in each of the emergency categories. It becomes important to assess the performance of YOLOv8 separately for each emergency category



4.4 Precision and Recall Analysis

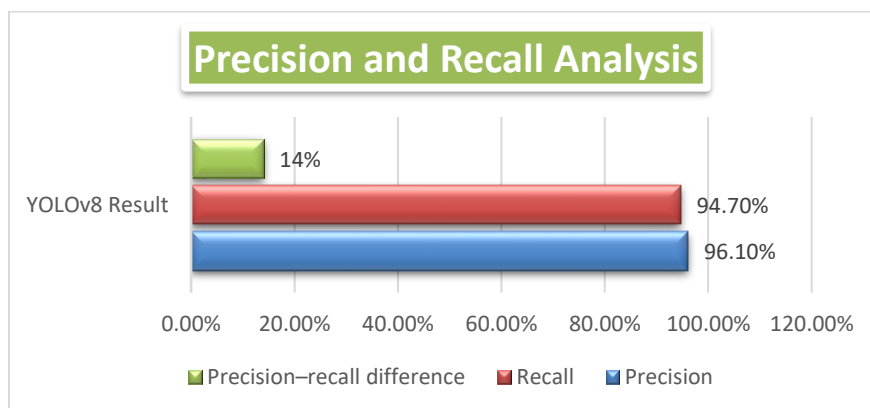
Both precision and recall offer further insights regarding the accuracy of emergency detection. The former refers to the proportion of true

emergencies in the set of all predicted emergencies, while the latter is the percentage of emergencies actually detected.

Metric	YOLOv8 Result
Precision	96.1%
Recall	94.7%
Precision-recall difference	14%

The precision rate of 96.1% implies that the proportion of accurate positive predictions is high. The recall rate of 94.7% implies that a great number of actual positives were detected by the model. The 1.4%-point difference between the two

rates implies that there is a balance between the precision and recall of the model.



In terms of emergency detection, the recall rate becomes especially important since a delayed reaction can occur due to a failure to detect a certain case. The precision rate is also very important since too many false positives impose extra loads on safety professionals.

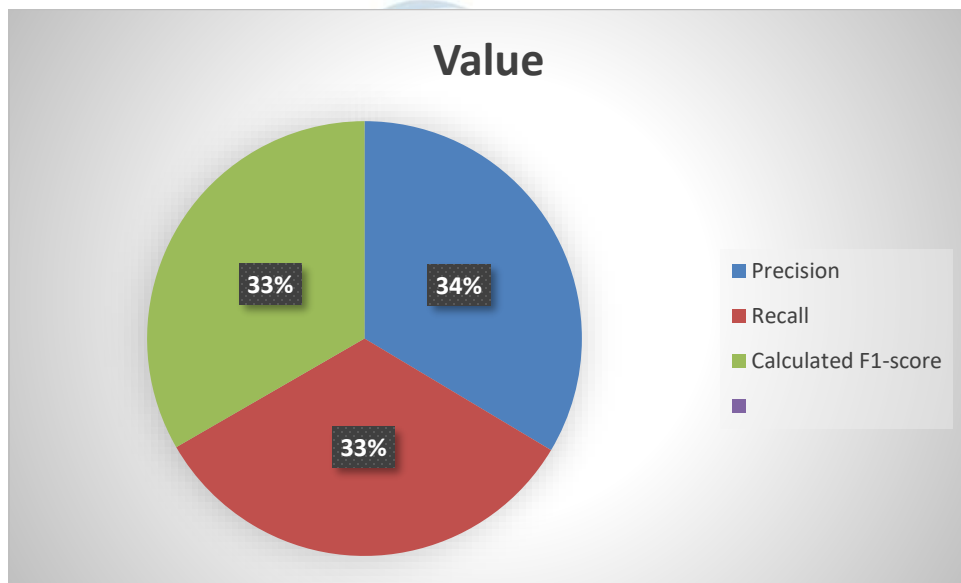
4.5 F1-Score Analysis

The F1-score is the result of calculating the harmonic mean between the values of precision and recall. From the given precision of YOLOv8, which is 96.1%, and the recall value of 94.7%, the estimated F1-score is 95.39%

Metric	Value
Precision	96.1%
Recall	94.7%
Calculated F1-score	95.39%

As is evident from the calculated F1 score, there exists a balance between the two values. This score is calculated from the two metrics and should not

be considered as an exact calculation but as an approximation.



5. Discussion

The results obtained within the framework of the conducted research are the evidence of the possibilities of the application of computer vision and deep learning methods in terms of automating emergency detection and ensuring real-time safety monitoring. In particular, according to YOLOv8 results presented in this study, which include accuracy of 95.4%, precision of 96.1%, and recall of 94.7%, the model is able to detect emergency-related visual data, achieving high classification performance. The average detection time of 1.1

seconds is another indication that the proposed model might be applied in cases of relatively fast detection requirements.

The results obtained are consistent with the research direction discussed in the reviewed literature sources, which focus on increasing the accuracy of object detection, simplification of computing complexity and efficiency of video stream processing. So, Li et al. (2022) researched the problem of using EfficientFormer as a computationally efficient vision transformer model, while Wang et al. (2024) and Zhou et al.

(2022) explored the possibility of increasing transformer computation efficiency.

Time aspect of emergency detection is worth consideration too. Zhao & Krähenbühl (2022) analyzed temporal smoothing transformers for real-time online video detection, whereas Jing et al. (2025) focused on the analysis of event-based streaming vision transformers. The mentioned research trends are relevant to emergency cases such as falls, accidents, and crowd violence that require analyzing changes from one video frame to another. In this case, the proposed LSTM approach aims at assisting in the process of temporal analysis, but it is impossible to evaluate numerically its effectiveness in comparison with YOLOv8 due to lack of numerical data.

The mentioned precision and recall numbers have certain practical significance in terms of safety monitoring too. High precision might result in less false emergencies and high recall in fewer missed events. Both features are highly valuable for security cameras, where false alerts make work difficult for the monitors and missed emergencies may cause problems. However, the metrics reported above do not show whether the model works equally well for fire, smoke, accidents, falls, and crowd violence detection.

The suggested four-module system architecture provides a step-by-step process for incorporation of video capture, preprocessing, emergency classification, and alarm triggering. Potential applications of this solution are public buildings, transport facilities, hospitals, industrial locations, and smart city surveillance systems. Nonetheless, the average detection time of 1.1 seconds must not be regarded as a proven emergency response time. The actual results would largely depend on camera quality, lighting, network connection, processing capacity, and alarm notification methods.

In summary, the paper presents evidence about the viability of AI-powered visual monitoring as an instrument of improving emergency awareness and safety monitoring. Additional testing is necessary for validation of the system, proving its reliability in different environmental conditions, and verifying the effectiveness of chosen models.

6. Conclusion

In this research, an AI-based emergency detection system was suggested through computer vision technology and deep learning for safety monitoring in real time. The system incorporates four stages: video data acquisition, preprocessing and object detection, emergency events recognition, and alerts creation. The emergency events considered in this research consist of five types: fire, smoke, accidents, falling down, and crowd violence (Shin et al., 2025).

The experiment in this study was conducted by a dataset including 25,000 images and video frames gathered from public datasets and CCTV videos. YOLOv8, CNN, and LSTM were chosen for implementation and evaluation by the measures of accuracy, precision, recall, F1 score, and detection reaction time. The findings of the experiments are that YOLOv8 model reached 95.4% of accuracy, 96.1% of precision, 94.7% of recall rate, and 1.1 second of detection time on average.

This approach can help improve public safety by providing better situational awareness and faster emergency notifications. Possible areas of use of this framework may be smart cities, hospitals, educational institutions, transport stations, and industry. At the same time, additional research is needed to prove its efficiency in different situations and provide comparative analysis of CNN and LSTM performance.

To conclude, AI-enabled emergency detection is one of the promising ways of intelligent safety control. Its efficiency will depend on proper detection, human supervision, protection of personal data, and rigorous testing before the implementation on a larger scale.

7. Recommendations

1. Increase Datasets' Diversification: Increase the number of images and video recordings with a view to having a diversified representation in terms of lighting, angle of vision, environment, and type of emergency situations.

2. Carry Out Comparative Analysis of Models: Compare YOLOv8, CNN, and LSTM models when trained and validated under identical

conditions. Include results related to accuracy, precision, recall, F1-score, and detection time of each model to make an unbiased comparison.

3. Test Models' Specific Performance on Emergencies: Test YOLOv8 model separately on fire, smoke, accidents, falls, and crowd violence cases. Category-based assessment will highlight strengths and weaknesses of the model.

4. Enhance Real-Time Detection Speed: Research lightweight network structures, model optimization, and efficient data pre-processing. Test the models for their ability to work in real-time.

5. Enhance Detection of Events Over Time: Develop and test methods that would help detect emergencies developing over time, including falls, accidents, and violent crowd actions. Temporal information might be useful in distinguishing between real emergencies and everyday activity.

REFERENCES

Anusha, P., Hazaimah, Y., Vijetha, T., Geetha, R., Venkatraman, A., & Ramya, M. (2026, February). Edge-Optimized Vision Transformer Architecture for Ultra-Low-Latency Object Detection in 6G IoT Devices. In *2026 IEEE 4th International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC)* (pp. 228-233). IEEE.

Hao, H., Sui, Z., Zou, R., Dai, Z., Zubić, N., Scaramuzza, D., & Wang, W. (2026). Low-latency event-based object detection with spatially-sparse linear attention. *arXiv preprint arXiv:2603.06228*.

Qiao, J., Li, W., Xie, H., Chen, H., Hu, J., Lin, S., & Han, J. (2025). Lipt: Latency-aware image processing transformer. *IEEE Transactions on Image Processing*.

Almalki, S. S. (2025, May). Transformer-Based Architectures for Real-Time Object Detection in Crowded Environments. In *2025 3rd International Conference on Business Analytics for Technology and Security (ICBATS)* (pp. 1-7). IEEE.

Huang, J., Kane, A., Zhou, F., Wei, Y., & Shi, H. (2026). Revisiting Real-Time Detection Transformer with Efficient Encoder Design. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6859-6868).

Omar, N. B. (2025). Real-Time and Low-Latency Processing in Computer Vision: A Literature Review. *Doupe Journal of Top Trending Technologies*, 1(2), 29-35.

Wang, S., Wang, X., Jin, L., Jiang, B., Zhu, L., Chen, L., ... & Luo, B. (2025). Towards low-latency event stream-based visual object tracking: A slow-fast approach. *arXiv preprint arXiv:2505.12903*.

Batiha, B., & Ait Fares, S. (2023). Transformer-Powered Object Detection for Smart and Real-Time Computing Platforms. *Electronics, Communications, and Computing Summit*, 1(1), 96-107.

Fan, L., Li, Y., Shen, H., Li, J., & Hu, D. (2024). From dense to sparse: low-latency and speed-robust event-based object detection. *IEEE Transactions on Intelligent Vehicles*.

Shin, W., Kang, D., Park, B., Kang, B. B., Lee, J., & Baek, H. (2025, December). Cf-detr: coarse-to-fine transformer for real-time object detection. In *2025 IEEE Real-Time Systems Symposium (RTSS)* (pp. 216-228). IEEE.

Hozhabr, S. H., & Giorgi, R. (2025). A survey on real-time object detection on FPGAs. *IEEE Access*.

Singh, A. (2023). Training strategies for vision transformers for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 110-118).

Zhang, L., Han, W., Yang, K., Zhang, K., Zhang, L., Xiao, R., ... & Tan, H. (2026). LSOD-DETR: a lightweight small object detection model based on real-time detection transformer. *The Journal of Supercomputing*, 82(2), 58.

- Li, Y., Yuan, G., Wen, Y., Hu, J., Evangelidis, G., Tulyakov, S., ... & Ren, J. (2022). Efficientformer: Vision transformers at mobilenet speed. *Advances in neural information processing systems*, 35, 12934-12949.
- Zhao, Y., & Krähenbühl, P. (2022, October). Real-time online video detection with temporal smoothing transformers. In *European Conference on Computer Vision* (pp. 485-502). Cham: Springer Nature Switzerland.
- Wang, X., Huang, Q., Li, X., Jiang, H., Xu, Q., Liang, X., & Song, Z. (2024). Vision transformer acceleration via a versatile attention optimization framework. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 44(6), 2398-2411.
- Zhou, Z., Liu, J., Gu, Z., & Sun, G. (2022). Energon: Toward efficient acceleration of transformers using dynamic sparse attention. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 42(1), 136-149.
- Rithick, S., & Jenefa, A. (2025, December). RT-SROD: Real-Time Suspicious Roadside Object Detection with a CNN-ViT Hybrid. In *2025 First International Conference of Advances in Engineering and Computing Technologies for Sustainable Development (AECTSD)* (pp. 1-6). IEEE.
- Rithick, S., & Jenefa, A. (2025, December). RT-SROD: Real-Time Suspicious Roadside Object Detection with a CNN-ViT Hybrid. In *2025 First International Conference of Advances in Engineering and Computing Technologies for Sustainable Development (AECTSD)* (pp. 1-6). IEEE.
- Jing, S., Guo, G., Xu, X., Zhao, Y., Wang, H., Lv, H., ... & Zhang, Y. (2025). ESVT: Event-based streaming vision transformer for challenging object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1-13.
- Zhao, Y., Wu, J., Chen, W., Wang, Z., Tian, Z., Yu, F. R., & Leung, V. C. (2024). A small object real-time detection method for power line inspection in low-illuminance environments. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(6), 3936-3950.