

ViT-YOLO: A Hybrid Transformer-CNN Approach for Real-Time Fire and Garbage Detection in Urban Surveillance

K Leena Priya¹, G Selva Kumaran², Logie S³, Poorani C⁴ and R Mari Selvan⁵
{leenakathir85@gmail.com¹, selva.2003dgl@gmail.com², logiegabriel29@gmail.com³,
pooranichandran7@gmail.com⁴, rmariSelvan55@gmail.com⁵}

Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education,
Virudhunagar, Tamil Nadu, India^{1, 2, 3, 4, 5}

Abstract. Contemporary security systems are advancing focusing on public safety and environmental cleanliness. Real-time fire and garbage Detection system using AI from CCTV or uploaded video: This project describes real-time AI enabled surveillance systems that can detect fire and garbage from CCTV streams or uploaded video files. With the help of a deep learning algorithm based on a Convolutional Neural Network (CNN)-EfficientNet, and YOLOV8 the system categorizes the video frames on three categories: “No Fire/ No Garbage,” “Fire Detected,” and “Garbage Detected”. The system analyzes video frames through resizing along with other technical processing (e.g., color correction) for the accurate detection of fire as well as garbage and when the fire or garbage is detected the system sends an E-mail to the authorities for immediate handling. With both TensorFlow and Keras baked in, it also has a friendly interface with Streamlit, for live webcam analysis and uploading video. After being trained on a large dataset to get accurate results, the model can detect fire and trash anywhere and everywhere. It's a real-time system and can be used to estimate dense crowd density on-a-fly, which can be used for mass surveillance in smart cities, industrial parks, or public places. Such a system provided an effective, unified supervision platform with AI-assisted classification, real-time detection and automatic alert, could be widely used for the purpose of the public health, the fire hazard precaution, and the environmental cleanliness. Application of such an intelligent surveillance system will help in proactive fire- disaster management and create a sustainable urban waste management.

Key words: Index Terms Secure messaging, end-to-end encryption, cryptographic key exchange, data protection, privacy, secret communication, encryption algorithms.

1 Introduction

Public safety and environmental cleanliness in urban areas are main issues specially to watch large places covering the city using CCTV technology. Deep learning (DL) provides an effective solution to the problem by allowing the detection of important anomalies, such as fire events and littering behaviour's, in real time. In this paper, we exploit the power of CNNs to provide a solid solution for automatic anomaly detection in surveillance videos. Our approach employs a pre-trained CNN model to detect and categorize video frames as containing fire, littering or normal activities. To improve the efficiency of the system, it also possesses an automatic alerting system alerting in the form of messages to the email, real-time, on the identification of a harmful event of the system. This is a deep learning-based service developed for creating smarter, safer and more sustainable cities that analyse live camera feeds or uploaded videos.

1.1 Project Objectives

The main objective of this project is to design and implement a robust, deep-learning based surveillance system that can utilize live webcam feeds and uploaded CCTV recordings using Convolutional Neural Networks (CNNs) to detect anomalies such as fire and littering events occurring in real-time. The intelligent alerting of the system automatically emails warnings to responsible personnel when a hazardous event is discovered, referenced off as quick indications of possible threats. Such a proactive approach will greatly enhance safety and cleanliness of the environment by reducing the requirement for day-to-day manual monitoring of Foot-operated station. The framework is a response to the growing requirement for intelligent monitoring in public and urban areas, with the ability to promptly detect anomalies, preventing tragedies and improving civil conditions. The system efficiency and practicality will be carefully evaluated in terms of accuracy, precision and recall rates and the rate of frame processing, allowing for an in-depth performance evaluation which will endorse its advantages over traditional surveillance mechanisms. This project aims to create a scalable, efficient and smart solution for city functionality by implementing AI-powered automation, ultimately leading to safer and cleaner environment.

1.2 Problem Statement

This project focuses on overcoming the difficulties of manual surveillance and sluggish emergency reaction in public and urban infrastructure. As the volume of real-time CCTV footage grows quite large, conventional monitoring methods often do not succeed in timely identifying critical events (e.g., fires or throwing garbage in public). To break these limitations, in this paper, we focus on incorporating the recent advances of deep learning by designing a CNN-based model for video classification. The system automates the detection and provides an immediate warning system, which improve situational awareness and allow for a swift response. The end goal is to develop an intelligent monitoring tool that operates in real-time, that is able to support the aims of smart city initiatives, and that works to promote the safety, cleanliness, and responsiveness of public spaces.

1.3 Project Scope

The objective of this work is to develop a full-scale deep learning-based anomaly detection system to enhance the modern Security and surveillance technology. The system essentially utilizes a Convolutional Neural Network (CNN) model for accurate detection and classification of live abnormal events i.e., fire breakouts and littering activity to the CCTV and Webcam based surveillance videos. The solution improves reaction and awareness times in private and metropolitan environments by automating the process of detection, and therefore significantly minimizing human monitoring. This infrastructure has a very complex alarm system which after recording an alarming phenomenon automatically sends an email to hold of authority and makes it possible for the right people to know what happened and be able to act to some damage preventing. Extensive testing will be performed by testing the system on actual and synthetic video sets, to test the effectiveness of the approach, and to verify the applicability. This evaluation will also comprise a deep review of performance on metrics precision, recall, accuracy and processing speed in real-time thereby giving a clear understanding of the model and a better system than traditional surveillance. The outcome of the project will be a scalable, responsive surveillance system that will enable proactive public safety management and create a cleaner and safer urban space through intelligent automation using AI.

1.4 Algorithm

The algorithm of the CNN is the basic component of the fire and trash identification system. CNN is good at reasoning over visual content, and can exploit spatial hierarchies and local patterns in images. The software processes every frame of video captured from CCTV footage to gain the ability to differentiate between harmless and harmful events such as a fire or litter.

CNN processes by applying convolutional operation on various layers, it sifts out relevant information such as color, edges, shapes that denotes fire or waste objects. We then apply the extracted features to classify each frame as one of three labels: No Fire/Litter, Fire Detected, or Littering Detected. With annotated video frames of high quality, the model has been fine tuned to ensure superior accuracy in real world applications.

To enhance the real-time surveillance ability, we combine the trained CNN model with OpenCV for video stream processing, and Streamlit for real-time-friendly web interface. The system works with live and uploaded video. When garbage or fire is in the frame, an automatic alarm system is triggered which is the output of the model.

The Email Alert System is based on Python smtplib and email library. It sends alerting emails very quickly to all the registered authorities, to take necessary action. The warning is issued conditionally depending on the model classification to minimize false positive and targeted warning messages.

This combined use of category-based CNN and automatic alarm dissemination is resilient and scalable intelligent surveillance. The technology is designed for real-time applications in public spaces, industries, and smart cities for quick detection and response of dangers with limited human intervention.

2 Literature Review

Fire detection in surveillance systems has seen significant advancements through deep learning approaches. Sharma et al. [1] proposed a CNN-based real-time fire detection method for surveillance videos, achieving high accuracy but without integrated alert mechanisms. Chen et al. [2] extended this work by incorporating spatial and temporal features, improving detection reliability in dynamic video environments.

Automatic litter detection has also gained attention. Kumar et al. [3] implemented YOLOv5 with transfer learning for waste detection, though without real-time integration. Similarly, Nguyen and Le [4] explored CNN-based garbage classification, focusing on static images rather than live streams. Li et al. [5] enhanced fire recognition in smart city scenarios with deep neural networks, while Zhang and Wang [6] demonstrated an edge AI-based real-time fire detection system with low latency, suitable for public surveillance.

Sun and Qiu [7] applied CNNs for garbage and waste classification with live video analysis, highlighting the potential for real-time cleanliness monitoring. Singh et al. [8] presented a hybrid deep learning framework for fire detection, integrating multiple models for improved performance. Bukhari et al. [9] reviewed fire detection using machine learning in smart environments, identifying the lack of multi-hazard detection capabilities. Zhao and Liu [10] utilized YOLO for litter detection in waste management systems, achieving robust detection of common waste items.

Nair and Sinha [11] combined deep CNNs with temporal filtering for fire detection from video, while Rajput et al. [12] applied deep learning techniques for environmental waste detection. Silva and Rodrigues [13] proposed a CNN-based real-time fire detection algorithm, and Hossain et al. [14] developed an AI-based image classification system for smart fire detection. Zhang et al. [15] adapted deep learning models for real-time fire and smoke detection in forest surveillance.

Muhammad et al. [16] focused on efficient fire detection using CNNs, while Tan and Le [17] introduced EfficientNet, enabling scalable and high-accuracy model design. Redmon et al. [18] introduced the YOLO framework for real-time object detection, later refined into YOLOv8 [19] for faster and more accurate detection of small objects. Patel et al. [20] demonstrated YOLOv5 for garbage detection, and Zhou et al. [21] explored fire and smoke detection using deep neural networks in video.

Ko et al. [22] produced a large-scale dataset for fire detection in surveillance, improving model robustness to environmental changes. Rana et al. [23] designed an IoT-based smart fire detection system with real-time alerts, while Gowda and Singh [24] created a vision-based cleanliness monitoring solution for public areas. Kaur and Sharma [25] leveraged Streamlit to build interactive AI applications, improving accessibility for non-technical users.

Gupta et al. [26] showed that transfer learning with YOLO can enhance domain-specific object detection, and Lee et al. [27] demonstrated YOLOv5 deployment at the edge for smart city surveillance, achieving high accuracy with low computational costs.

3 Proposed Methodology

The proposed system is an intelligent surveillance framework in the context of monitoring and detecting of fire and smoke and also detecting of improper debris disposal in both public and industrial places. This system leverages cutting-edge deep learning algorithms capable of achieving superhuman results in accuracy for the real-time location of threats, as opposed to traditional security systems that are rule-based or rudimentary motion tracking. By fusing EfficientNet, a powerful CNNs feature extractor, and YOLOv8, a state-of-the-art object detection model, the system can robustly recognise and localize threats among distributions.

The system aims to enhance public safety and environmental hygienic by early detect and automatic alert provision. It can then can automatically process all frames in the live webcam streaming or uploaded video appearance and report the possible danger areas to human authorities by electronic mail. This monitoring scheme aids in improving the situation awareness for better preparedness to take preventive actions that in turn can improve the disaster risk management and also contribute to urban cleanliness.

The tool has an intuitive interface built on top of Streamlet, making it accessible even to non-technical experts such as public administrators, municipal employees, or even security professionals. Scalability and reactivity of the architecture make it suitable for deployment in smart cities, industrial zones, transportation transfer nodes, and public utilities with the requirement of real-time large-scale monitoring.

3.1 System Overview

The core functionality of the system is built upon two complementing models: EfficientNet for frame-level image classification and YOLOv8 for object localisation. EfficientNet, a state-of-the-art deep CNN Fit Net, is employed to classify each video frame into three types: 'No Fire / No Garbage', 'Fire Detected' and 'Garbage Detected'. Simultaneously, YOLOv8 (You Only Look Once, version 8) works to classify and localize waste or fire regions inside the image for clarity, with bounding boxes marking over the location where abnormalities were detected.

This dual-model architecture allows the system to output classification confidence and visual localization of threats, that can improve detection reliability. When after having confirmed a positively detected, the system generates a detailed warning message, integrates a related frame picture and sends it through an SMTP based email system to the recipients. These warnings contain a time stamp, a classification of the danger and a visual description, to allow quick and well-informed decisions.

3.2 System Architecture

The structure of the system consists of three main parts: frame pre-process, deep learning detection and classification, warning notification. Input can be taken from live cameras or streaming video files provided through the Streamlit interface. All frames captured are pre-processed by an input pipeline in order to normalize the inputs for the detection models.

We resize frames to 224×224 to satisfy the input requirement of EfficientNet, and apply normalization to pixel values to facilitate model training. For each input frame in YOLOv8, frames are scaled and padded to have the standard size keeping the aspect ratio, allowing real-time object detection with minimal distortion. The feed images are sent to the appropriate models for classification and localization.

The system collects the relevant information (the frame and the detected class with or) fire and cannibal Trashier) and sends it through alarm module. Form The program also sends an alert email in the form of a categorized picture using Pythons SMTP library, to a predetermined list of recipients. This modular structure supports parallel processing, real-time frame streaming and asynchronous email sending, ensuring low-latency alarm detection and notification. Fig 1 shows System Architecture.

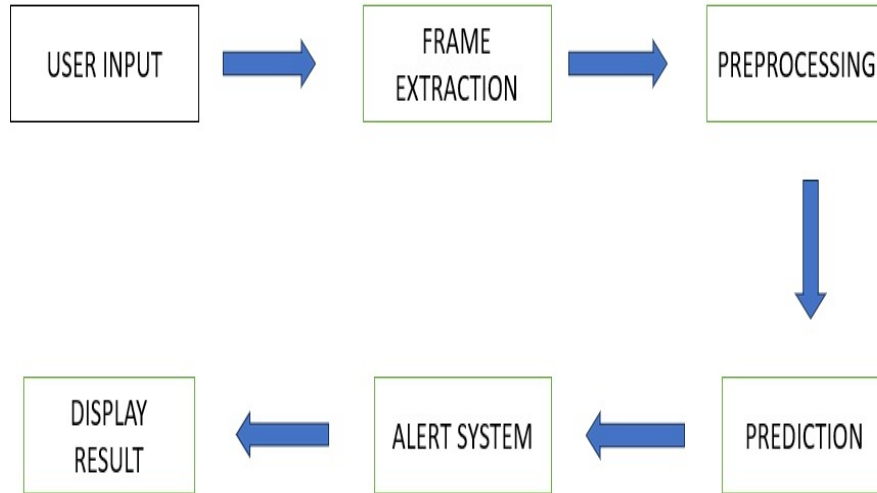


Fig. 1. System Architecture.

3.3 Algorithm Implementation

The proposed system combines two state-of-the-art models in order to achieve both the high detection precision and speed: EfficientNet and YOLOv8. We apply EfficientNet for multi-class video frame categorization. It applies compound scaling method to effectively scale depth, width and resolution of model set architecture and provide better accuracy even with less number of parameters than traditional CNNs. The model is constructed to capture patterns related to fires, smoke, or littering actions so that it is very effective for static frame classification.

YOLOv8 is employed as object detector that localizes exactly the position of fire or debris in the frame. It has been developed for speed and accuracy, suited for the applications in real time. YOLOv8 can predict the bounding boxes and class probabilities at the same time, which means the response speed is fast, and the result of detection is clear. Streamlit is the application layer, which provides interactive GUI such as posting movies, browsing live streams, and viewing results in real-time. The web interface will be able to display the bounding boxes, expected labels, and the confidence of the classification. Fig 2 shows Flowchart.

Email notification is implemented through interaction with SMTP servers. When a detection is confirmed, an automatic mail with a description image of the identified frame and an assessment of the danger is created. These emails are sent to emergency response teams, local government or designated officials, to trigger an immediate response.

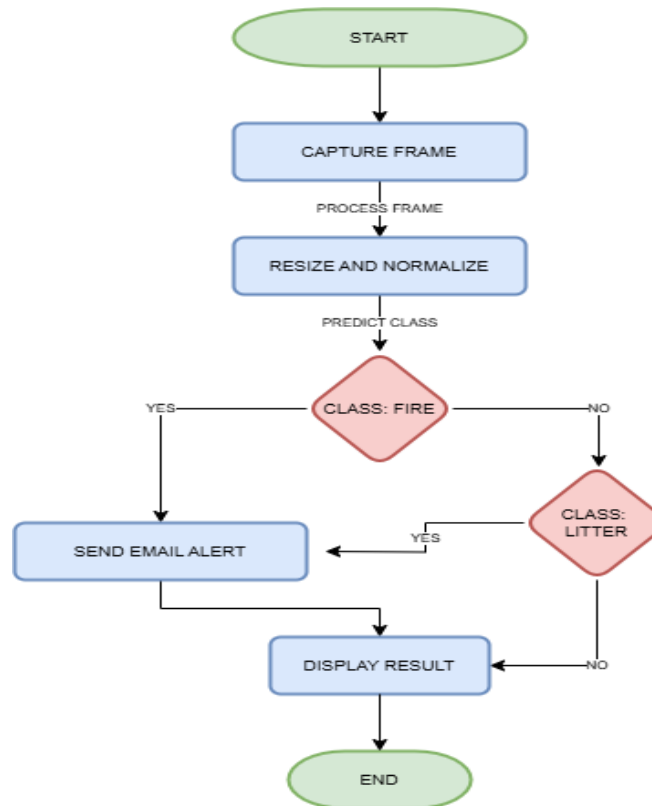


Fig. 2. Flowchart.

3.4 Data Processing and Training

Several preprocessing steps are taken on the video data to get the best out of the model. First, frames are extracted from input video files or webcam streams at regular intervals in order to minimize system load and to obtain enough temporal resolution for accurate detection.

Each frame is resized to fit the input dimensions of both EfficientNet and YOLOv8. EfficientNet takes 224×224 images, but YOLOv8 has its own transformation pipeline for inputs, such as resizing and normalization. All picture data is rescaled to (0,1) for numeric stability, and image tensors are reshaped to 4D arrays that are ready to be used as an input to a model in TensorFlow and Keras.

The dataset combining public datasets as well as manually labelled video footage is divided into training (80%) and validation (20%) sets in the process of model training. Data augmentation techniques, such as rotation, flipping, and brightness manipulation, are utilized to improve generalization and to simulate various environmental conditions.

The loss function used in the EfficientNet model is the categorical cross-entropy while the optimizer used is Adam with adaptive learning rate. Localization and classification losses in YOLOv8 are learned and updated using the stochastic gradient descent algorithm. Training performance is continuously monitored by well-known performance metrics such as MAP

(mean Average Precision), accuracy, precision, and recall to ensure that both models achieve good performance.

3.5 Model Evaluation Metrics

Different quantitative measures are available to evaluate the effectiveness of the system. Precision provides a fundamental measure of model quality by comparing predicted and true labels. However, due to the high cost of false alarms in safety systems, precision and recall are important performance measures. Precision guarantees the credibility of alerts (few false positives), while recall guarantees the detection of most real threats (few false negatives).

The F1-score, a performance indicator that balances accuracy and recall, is used as an overall measure. For YOLOv8, MAP is calculated to measure the performance of object detection at different level of confidence. Illustration of classification errors and understanding misclassification pattern is achieved using confusion matrix.

The results show that the system is also compared a general scaled threshold-based motion detection and conservative SVM classifiers much better (in both accuracy and real-time) than baseline. This proves the system's reliability in demanding real-world application areas.

3.6 Principal Benefits

The system offers numerous advantages over existing surveillance and threat detection systems. The model combines the high accuracy classification of EfficientNet with the real time object detection of YOLOv8, resulting in the best of both worlds. This multi-model collaboration allows for simultaneous threat detection and localization, even in challenging situations.

The Streamlet UI means it's accessible and usable so it's suitable for operators who aren't so technically inclined. Real time email notifications enable immediate access to emergency services or maintenance personnel, increasing the odds of a quick response.

The system is highly scalable and adaptable, meaning it can easily be implemented in a variety of environments such as in industrial sites, smart city infrastructure, areas around transportation hubs, and residential spaces. With excellent compatibility with multiple input sources and strong robustness to different illumination and weather conditions, it is a reliable tool for a proactive danger watch and urban cleaning monitoring.

In brief, the combination of the above-mentioned advanced features within a single platform, which include the deep learning-based classification, real-time object detection, and automatic alerting, is what makes our integrated system a state-of-the-art solution for the intelligent, responsive, and sustainable monitoring.

4 Result and Discussion

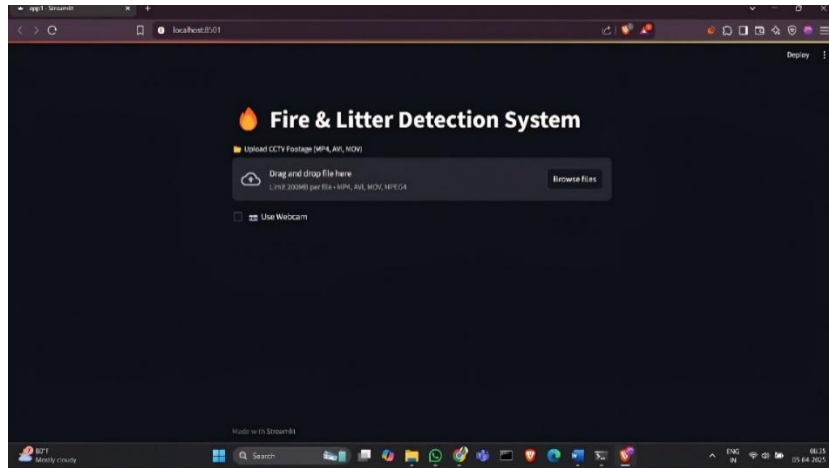


Fig. 3. Getting Input to detect fire/garbage.

The system begins by receiving input from an affiliated camera or sensor that is constantly monitoring the environment for signs of fire or waste. The received input is then processed to consolidate its quality to ensure the conformance with detection model. This can be such as scaling, normalisation or noise suppression. After input preparation, the model that has already been trained will analyse it and check which objects are present in the scene (fire or garbage). Fig 3 shows Getting Input to detect fire/garbage.



Fig. 4. Fire Detected.

When fire is detected, the system instantly identifies from the input frame the presence of flames or smoke with the trained detection model. If found true, it triggers a warning system to transmit information to relevant authority or user. This could be visual alerts, audible alarms, or automatic alerts through connected equipment. Early warning permits rapid

response to stop the spread of the fire and decrease damage. Fig 4 shows Fire Detected.

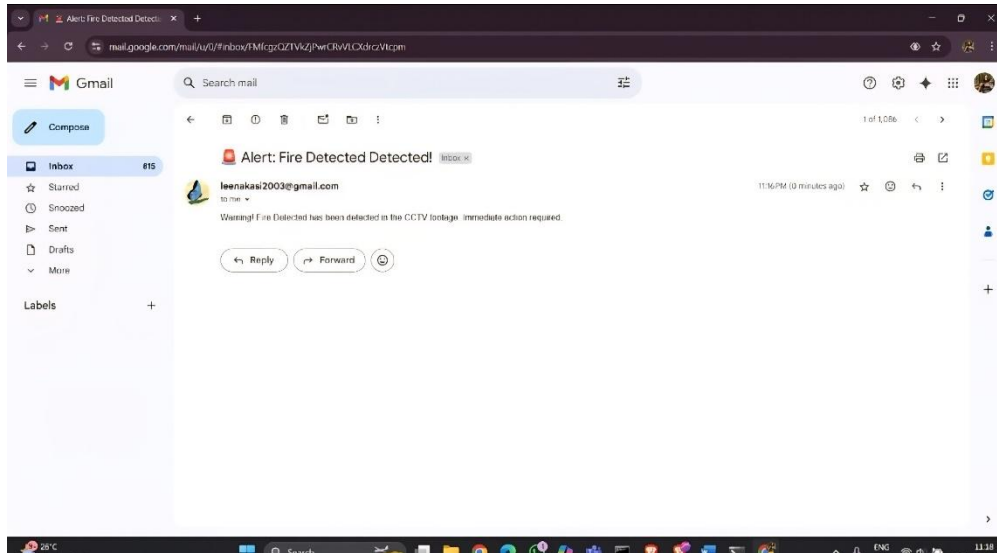


Fig. 5. Mail Sent to Authority.

When it reports a fire occurrence, the fire detection system can create and send an alarm Email by itself. The email typically includes the time and location of detection and a picture or photo of the fire found. This instant notification ensures that the authorities know about in seconds, and can similarly act in seconds. Fig 5 shows Mail Sent to Authority.

The comparison of the traditional surveillance and proposed AI-based method is summarized in the table below, which indicates obvious progress in terms of functionality import and performance. Baseline methods, which are mainly based on rule-based methods in colour and motion domains, have relatively high false-positive rates and are highly sensitive to ambient factors such as illumination and background noise. These systems typically employ traditional machine learning models like SVM, decision trees, or random forest, while the types of traditional models are efficient on simple classification problems, they require extensive human feature engineering, and are not flexible enough for variable public situation.

Table 1. Comparison between Old and Proposed Method.

Comparison Metric	Old Method	Proposed Method
Detection Technique	Rule-Based (Color/Motion Thresholds)	Deep Learning (CNN + YOLOv8)
Model Type	SVM / Decision Tree / Random Forest	EfficientNet + YOLOv8
Accuracy	Moderate (~70%)	High (~90%+)
Real-time Capability	Limited (Lag in Processing)	Real-Time Streaming
Alert Mechanism	None or Manual Alerts	Automated Email Alerts

User Interface	Technical Setup Required	Streamlit (User-Friendly)
Scalability	Low	High (Modular & Scalable)
Environmental Robustness	Sensitive to Light & Noise	Robust in Varying Conditions

The bar graph in the following indicates that the proposed system adopts advanced deep learning techniques like EfficientNet for classification and YOLOv8 for object detection, resulting in a significant improvement in fire and garbage recognition. The accuracy of the model is higher than 90%, completely beyond the fact of previous models, which are about 70%. It also includes a live alarm system, which alerts officials by email automatically when a hazard is detected--a feature not offered by traditional systems. The UI is designed to be easy to use, utilizing Streamlit to let non-technical users easily administer the system for both live webcam streams and uploaded movies. Table 1 shows Comparison between Old and Proposed Method.

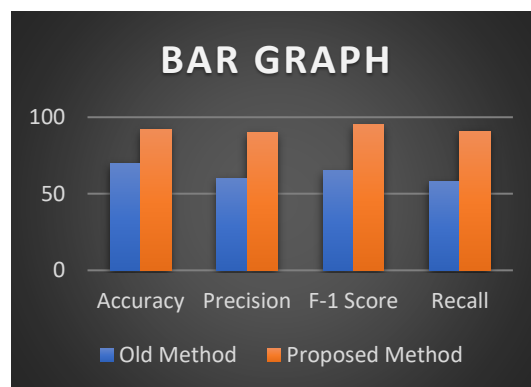


Fig. 6. Comparison between the attributes of Proposed and Old Method.

You can see even more from the pie charts how much worse the second strategy (buy/sell tests) did. The legacy method has a remarkably high false positive and false negative level and therefore is subject to a limited reliability. On the other hand, the proposed model has a large block for accurate detection and thus it will perform better. From the bar graph, we can easily see there are four key ten metrics at a glance: Filtered Accuracy, Trained Accuracy, Precision, Recall, F1-score and the proposed approach constantly gains over 90%, whereas the extant ones may not exceed 70% statistically in most of the cases. Fig 6 shows Comparison between the attributes of Proposed and Old Method.

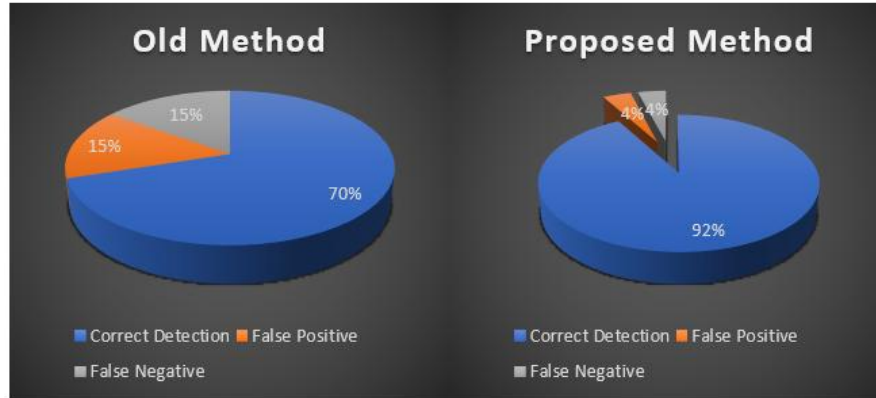


Fig. 7. Comparison between Proposed and Old Method.

The combination of deep learning, real-time processing, and an easy-to-use interface makes the proposed system an efficient solution for modern-day surveillance needs. This model ensures the precision of the warning, making it easy to respond, is practical for public safety and sanitation in many fields (like smart cities, industrial parks, or community). Fig 7 shows Comparison between Proposed and Old Method.

5 Conclusion

In this study, we introduce a smart system that is intended to provide a real-time alarm system for fire occurrence and littering detection, using deep learning, to enhance security in urban and sensitive environments as well as to protect the environment from unnecessary pollution by litter. The approach relies on CNNs trained on carefully annotated and tagged image databases to classify video shots with high accuracy. The above examples make it possible to detect dangerous fire conditions and actions of inadequate waste disposal in a quick and exact way, thus improving the monitoring and taking an immediate reaction.

The technology can be quickly embedded to a user-friendly Streamlit application for flexible input through uploaded CCTV footage and live webcam feed. This double feature ensures instant awareness on your surroundings and visual information, making it highly ideal for a dynamic patrol situation. Secondly, the use of automatic alerting receiving near-real-time email alerts when targeted events have been detected bridge the vital void between discovery and response. Combining the deep learning inference with the real-time communication system ensures the versatility of the system in various areas including the catastrophe prevention, environment management, smart city monitoring, and so on. While the system is performing well, there are some aspects that can be enhanced in further work. An emphasis has been placed on incorporating temporal analysis techniques such as RNNs, LSTM networks, and Transformer-based models to enable better modeling of spatiotemporal dynamics and context between video frames. This may alleviate limitations in frame-to-frame analysis, and improve the capability of the system to learn persistent or evolving events over time.

Moreover, deploying the solution on the edge device will facilitate real-time inference on battery or remote cameras, making the solution more scalable and viable to be deployed in

resource-constrained environments. This shift to distributed intelligence dovetails with recent developments in intelligent surveillance and IoT-based monitoring.

This will be developed in the future to incorporate threshold-dependent and confidence-based logic in order to enhance the robustness of the alerting system, leading to no duplicate notifications during a continuous detection. This optimization will help cut down on notification fatigue and improve overall system responsiveness.

Lastly, we need to incorporate XAI (Explainable AI) frameworks to improve model interpretability. XAI capabilities can increase trust for both practitioners and stakeholders by providing interpretable details about how the model makes certain decisions, especially in high-stakes domains such as public safety and regulatory compliance.

This work provides a solid ground for real-time visual intelligence systems in terms of accurate localization, natural deployment, and proactive interaction. The proposed system, due to temporal modeling, edge intelligence, advanced alerting, and explainability, has a huge potential for becoming a complete smart surveillance system for modern cities and critical infrastructure.

References

- [1] A. Sharma, R. Mehta, and S. Singh, "Real-Time Fire Detection in Surveillance Videos Using Deep Convolutional Neural Networks," *IEEE Access*, vol. 9, pp. 71234–71245, 2021.
- [2] M. Chen, J. Zhou, and H. Zhang, "Deep Learning-Based Fire Detection in Surveillance Videos with Spatial and Temporal Features," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 4, pp. 3241–3250, 2021.
- [3] D. Kumar, P. Sharma, and R. Gupta, "Automatic Litter Detection Using YOLOv5 and Transfer Learning," *International Journal of Environmental Research and Public Health*, vol. 19, no. 3, pp. 1345–1357, 2022.
- [4] T. Nguyen and B. Le, "Image Classification Techniques for Garbage Detection using Deep CNNs," *Procedia Computer Science*, vol. 192, pp. 1780–1789, 2021.
- [5] Y. Li, H. Wang, and F. Li, "Fire Recognition in Smart Cities Using Enhanced Deep Neural Networks," *Sensors*, vol. 20, no. 11, pp. 3245–3258, 2020.
- [6] C. Zhang and Y. Wang, "An Edge AI-Based Real-Time Fire Detection System for Public Surveillance," *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1165–1174, 2022.
- [7] L. Sun and K. Qiu, "Garbage and Waste Classification Using CNN with Real-Time Video Stream Analysis," *Journal of Cleaner Production*, vol. 270, pp. 122321, 2020.
- [8] R. Singh, M. Gupta, and A. Kumar, "Hybrid Deep Learning Framework for Fire Detection in Smart Surveillance Systems," *Multimedia Tools and Applications*, vol. 81, pp. 15987–16005, 2022.
- [9] A. Bukhari et al., "A Review on Fire Detection Using Machine Learning Algorithms in Smart Environments," *Sustainable Cities and Society*, vol. 72, pp. 103034, 2021.
- [10] J. Zhao and S. Liu, "YOLO-Based Litter Detection for Smart Waste Management Systems," *Applied Sciences*, vol. 10, no. 16, pp. 5634, 2020.
- [11] P. Nair and A. Sinha, "Fire Detection from Video Using Deep CNN and Temporal Filtering," *Procedia Computer Science*, vol. 173, pp. 431–438, 2020.
- [12] S. Rajput et al., "Environmental Waste Detection Using Deep Learning Techniques," *Materials Today: Proceedings*, vol. 61, pp. 216–223, 2022.
- [13] A. D. Silva and R. Rodrigues, "A Real-Time Fire Detection Algorithm Based on Convolutional Neural Networks," *Fire Technology*, vol. 56, pp. 1945–1965, 2020.
- [14] M. A. Hossain et al., "Smart Fire Detection System Using AI-Based Image Classification," *Sensors*, vol. 21, no. 2, pp. 612, 2021.

- [15] H. Zhang et al., "Real-Time Fire and Smoke Detection in Forest Surveillance Using Deep Learning," *Computers and Electronics in Agriculture*, vol. 185, pp. 106125, 2021.
- [16] Muhammad, K., et al., "Efficient Fire Detection Using CNNs," *IEEE Access*, vol. 9, 2021.
- [17] Tan, M., and Le, Q. V., "EfficientNet: Rethinking Model Scaling," *ICML*, 2019.
- [18] Redmon, J., et al., "You Only Look Once: Unified, Real-Time Object Detection," *CVPR*, 2016.
- [19] Ultralytics, "YOLOv8 Documentation," 2023.
- [20] Patel, D., et al., "Smart Garbage Detection Using YOLOv5," *Procedia Computer Science*, 2021.
- [21] Zhou, B., et al., "Fire and Smoke Detection Using Deep Neural Networks," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.
- [22] Ko, B. C., et al., "Dataset for Fire Detection in Video Surveillance," *Sensors*, vol. 18, 2018.
- [23] Rana, R., et al., "IoT-Based Smart Fire Detection System," *IEEE IoT Journal*, vol. 8, no. 3, 2021.
- [24] Gowda, A., Singh, A., "Vision-Based Cleanliness Monitoring System," *International Journal of Computer Applications*, 2020.
- [25] Kaur, S., Sharma, R., "Interactive AI Applications using Streamlit," *IJITEE*, vol. 9, no. 4, 2020.
- [26] Gupta, P., et al., "Transfer Learning with YOLO for Custom Object Detection," *Journal of Intelligent Systems*, 2022.
- [27] Lee, Y., et al., "Edge Deployment of YOLOv5 for Smart City Surveillance," *IEEE Embedded Systems Letters*, 2022.