

Accurate Fire and Smoke Localization Using Multi-Variant YOLO Deep Learning Frameworks

Srishti Saha¹, Dr. K. Ashesh²

¹Master's Student, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India – 522502
Email: srishtisaha2911@gmail.com

²Deputy-HOD (Academics), Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India – 522502
Email: imasheshk@kluniversity.in

Received: 5th January, 2026; Revised: 18th January 2026; Accepted: 25th January, 2026; Available Online: 28th January, 2026

Abstract: Early detection of fire and smoke is important to provide timely warnings of potentially hazardous situations and so help save human life, property and the natural environment. Traditional line-based vision-based detection methods rely on features that are manually designed and usually are not able to ensure constant performance under different challenging environments like differences in light conditions, background noise, and smoke density. In order to overcome these limitations, a DL-based fire/smoke detection system utilising cutting-edge YOLO architectures is suggested. A well selected set of photographs of some smoke/fire situations is used. YOLO format has bounding box annotations to denote the object localisation information. Preprocessing involves the arrangement and preparation of data, application of appropriate normalisation algorithms and integration of data within a common file structure to facilitate data detection and classification at later stages. In this work, we train and test several models of YOLO: YOLOv5, YOLOv8, YOLOv9, YOLOv10, YOLOv11s and a modified YOLOv11-DH3 to compare their performance. Additionally, to test the performance enhancements, the current YOLOv11x model is also presented. The model is evaluated with standard metrics including precision, recall and mAP. Based on the results obtained from the studies, it was determined that in terms of the performance of models, the YOLOv11x model could be the most appropriate one with 0.930 accuracy, 0.981 recall, and 0.967 mAP. Moreover, the system is integrated with a web application developed using Flask library, which enables them to track the smoke and fire in real time, enhancing deployment and usability.

“Index Terms: Fires, Feature extraction, Shape, YOLO, Accuracy, Detectors, Monitoring, Image color analysis, Web and internet services, Deep learning.”

How to cite this article: Srishti Saha¹ Dr. K. Ashesh². Accurate Fire and Smoke Localization Using Multi-Variant YOLO Deep Learning Frameworks. Int J Drug Deliv Technol. 2026;16(85s):1577-1579. DOI: 10.25258/ijddt.16.85s.167.

INTRODUCTION

Fire represents a major environmental and societal hazard, posing significant risks to human safety, built infrastructure, and ecological systems. The fire risk domain has shifted from the industrial domain to densely populated residential areas and natural environment like forests because of the process of urbanisation and increased frequency of large public events. Fire incidents are prone to happen from fireworks, power failure, industry activities and climatic conditions. Consequently, effective prevention strategies and continuous monitoring play a vital role in ensuring public safety [1]. It is evident from history that uncontrolled fires can result in substantial property damages, significant loss of life and long term environmental damage, and thus early detection and rapid response is of paramount importance [2].

Despite the continued development of fire-safety technology, real-world complex conditions are still a major challenge for successful early detection. The traditional fire monitoring approach is largely dependent on physical or chemical sensing systems which are sensitive to environmental disturbances and that have limited spatial coverage [3]. Such systems can generate false alarms or misclassify the

fire and non-fire phenomena, particularly in dynamic and/or visually cluttered scenes [4]. Furthermore, most existing detection systems are not capable of providing adequate performance in changing conditions, such as illumination, weather, backdrop complexity and fire size, indicating a major lack of development of an accurate and versatile fire and smoke monitoring system [5].

This is to build an intelligent visual detection system which can accurately and quickly detect fire or smoke incidents. The proposed method offers a high accuracy detection and a sufficient operation efficiency for real time applications. The framework aims to enhance robustness across different scenarios, while overcoming the limitations of conventional systems, by leveraging large-scale visual data and advanced paradigms of detection via learning. Among the major contributions are the comparison of several detection paradigms, the comprehensive performance evaluation under the same test scenario and the provision of a scalable monitoring solution for both urban and natural environments [7].

The relevance of this effort is to improve the early warning capability and assist the speedy emergency response, and reduce the casualties and economic

losses of fire disaster. Better fire and smoke detection could help safety authorities gain better situational awareness, help with evacuation planning and provide critical decision support during an emergency [8]. Moreover, the use of intelligent visual monitoring can reduce false alarm rate and realize the continuous large area surveillance [9] and reduce the maintenance cost. Finally, this contribution helps build intelligent fire prevention infrastructure, which helps to make communities safer and to strengthen environmental protection policies [10].

2. LITERATURE REVIEW

Lighting is a fundamental and important aspect of fire detection as well as smoke detection. Using learning-based visual representations, recent developments in computer vision have greatly improved the performance of smoke and fire detection systems. The method of smoke detection based on 3D convolutional neural network proposed by Lin et al. [11] has good temporal feature extraction ability, and has been proved to improve the smoke detection accuracy in dynamic scenes. Their expensive computing requirements prevent them from being easily used in real time for 3D models, however. In forest fire images, Li and Li [12] attempted to employ the generative learning technique to produce clearer images in complex background, but did not directly detect the smoke. Jeong et al. addressed efficiency by reducing the complexity of the model and speeding up the model inference to create a lightweight real-time wildfire smoke detection system [13]. Recurrent models are effective at detecting smoke early but they usually cannot accurately localise the smoke to a location. Zhao and Ban had investigated the potential of using time-series satellite data to identify fires at an early stage and showed the potential of remote sensing for wide area monitoring [14]. The application of satellite-based detection is limited in the temporal precision because of cloud and lagged reaction, and in the scope of application due to limited temporal coverage.

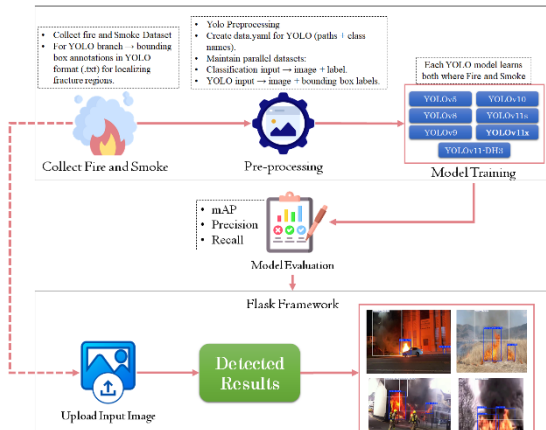
There are standard techniques, such as feature-driven, used in previous publications. To overcome these limitations, Jamali et al. [15] introduced a fire detection method based on saliency and texture and colour information which is interpretable and with low computation cost. The handcrafted properties, however, are vulnerable to environmental variations and are not robust when facing complex situations. Recent works have shifted towards end-to-end detection paradigms. Huang et al. [16] developed a forest fire smoke detection algorithm which improves the algorithm's sensitivity in detecting remote or weak smoke patterns. Such approaches work for modest goals, but frequently need complex model architecture and large amount of training data.

Single stage object recognition systems are often used because of the speed vs. accuracy trade-off. Later, Lian et al. [17] came up with a novel detection system for fire and smoke in real-time which performed better than previous detectors. The first work introducing real-time unified object detection, which is beneficial for visual identification, is done by Redmon et al. [18]. Later architectural progress, like the deformable convolution methods, enhanced the feature alignment and spatial flexibility in complex visual situations [19]. Some work on cross-network learning approaches for improving the feature interaction and generalisation performance in large-scale learning tasks has been conducted [20].

Despite the current techniques, there are still challenges to achieve high detection accuracy for different scenarios, real-time performance and efficiency of deployment. Most of the methods are time and resource intensive, sensitive to environmental noise or poorly generalisable for fire conditions. These gaps necessitate the requirement for a strong and scalable framework for detection which balances accuracy, speed and adaptability. This paper proposes a detection framework based on recent detection paradigms to tackle these limitations, aiming to improve the reliability and feasibility of fire and smoke detection for practical monitoring applications.

3. MATERIALS AND METHODS

We present in this paper a robust approach to early detection and localisation of fire and smoke using deep-learning. The detection models are trained on a carefully crafted image dataset in the YOLO format, with bounding boxes. The entire approach combines data set organization, data preprocessing, data normalization, model training and model evaluation in one detection pipeline. YOLOv5, YOLOv8, YOLOv9, YOLOv10 and YOLOv11x architectures are analysed for fire detection and smoke identification under various scenarios. Furthermore, an improved YOLOv11-DH3 model is presented by adapting the architecture to promote features and robustness to detect in challenging situations. Success of the model is evaluated by several performance metrics including precision, recall and mAP. To perform real-time inference and interaction with the user, the proposed detection system is implemented in a real-world scenario of a Flask based web application. The entire system is designed to enable fast hazard identification, accurate localisation and solution for intelligent fire safety monitoring applications.

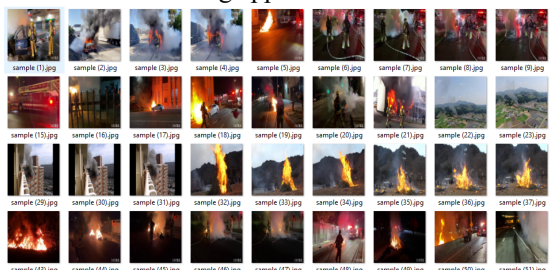


“Fig.1 Proposed Architecture”

The proposed system is based on four levels, information layer, experimental layer, model processing layer and application layer. The data layer is responsible for acquiring datasets, YOLO-based annotation and setup. The experiment layer and the training layer are used to carry out a comparison test of the multiple versions of YOLO. The output of the main model layer after the image or video input undergoes real-time inference is the bounding box, class label and confidence score. Lastly, the trained model is implemented in the application layer using a web interface to enable real-time fire/smoke detection, user interaction, and visualisation.

a) Dataset Collection:

A set of fire and smoke images is selected, and the tests are conducted on this set of images, which are available from a public internet repository. Many of the visual examples have examples of both fire and smoke and are variously set in different locations and with different backgrounds and lighting conditions. Many item classes related to fire and smoke are associated with the images and marked with bounding boxes in the YOLO format. It is data with a very high diversity of complexity in the scenes and in the scale of the items within it, and is an ideal set for training and evaluating powerful object detection models. This diversity provides good generalisation and good performance to real-world fire monitoring applications.



“Fig.2 Dataset”

b) Pre-Processing:

Pre-processing of data set ensures uniformity, well-annotations, and structure which is key for successful learning. It generates appropriately processed images and corresponding annotations for

classification and detection tasks, thereby improving training stability, predictive accuracy, and the model’s ability to generalise to unseen data.

c) Data Preprocessing: This step is concerned with making the dataset more consistent and efficient for training the model for many learning pipelines. We normalise images to remove variations due to illumination and scale discrepancies and align class labels properly with annotation files. We produce one configuration file which specifies the path to the datasets and classes for detection. We maintain parallel dataset structures for classification and detection, which could facilitate compatibility among different models and enhance the stability, convergence, and generalisation capabilities in actual situations when training.

d) Algorithms:

YOLOv5: YOLOv5 is an efficient single-stage object detector that can predict a high-quality fast single-stage localisation by predicting both class probability and bounding-boxes. Combined with a balanced design, it has a good real-time inference, and can be used to achieve reliable detection performance in various visual scenarios.

YOLOv8: YOLOv8 strengthens the robustness of detection by utilizing the technique of feature aggregation and anchor-free prediction. It can efficiently extract multi-scale spatial information to increase the localisation accuracy and generalisation, especially in the complex and dynamic changing visual situations.

YOLOv9: YOLOv9 is designed to enhance the performance of detecting objects, with the hope of improving the feature extraction and improving the prediction of the object bounding region. Multi-scale detection capability enables accurate detection of objects in complicated background, different sizes of objects and different environmental conditions.

YOLOv10: YOLOv10 focuses on architectural optimisation for faster inference and higher accuracy, YOLOv10 aims to boost the inference speed and accuracy. Efficient feature fusion and head detection facilitates accurate localisation and robust performance in real-time monitoring and safety-critical applications.

YOLOv11s: YOLOv11s is a lightweight detection architecture designed for rapid inference that has a minimal impact on performance. It offers a tradeoff between computational efficiency and accuracy of detection and ensures robust object recognition in more complex scenarios with changes in lighting conditions and when objects are partially occluded.

YOLOv11-DH3: YOLOv11-DH3 includes architectural changes for enhanced feature extraction and robustness of detection. The AI’s greater ability to fuse features helps improve resilience and localization accuracy in challenging, visually complex environments, allowing the system

to accurately identify objects even in difficult surroundings.

YOLOv11x: YOLOv11x is a next-generation detection architecture, featuring more sophisticated feature extraction and multi-scale representations. High accuracy and excellent generalization, and the ability to accurately and stably recognize objects during real-time operation in different challenging environments.

4. EXPERIMENTAL RESULTS

Precision: The percentage of correctly categorized instances/samples that are classified as positive. So, the formula for precision is:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} (2)$$

Recall: Recall is a ML performance measure which represents a model's ability to correctly identify the

“Table.1 Performance Evaluation Table”

ML Model	Precision	Recall	mAP
Yolo v5	0.925	0.983	0.964
Yolo v8	0.925	0.979	0.962
Yolo v9	0.929	0.973	0.961
Yolo v10	0.909	0.959	0.960
Yolo v11s	0.927	0.980	0.963
Proposed Yolo 11-DH3	0.792	0.822	0.852
Extension Yolo v11x	0.930	0.981	0.967

The model is tested using the metrics of precision, recall and mAP across a few different YOLO architectures. YOLOv11x among them has better detection capabilities, resilience and generalisation, achieving better overall performance.

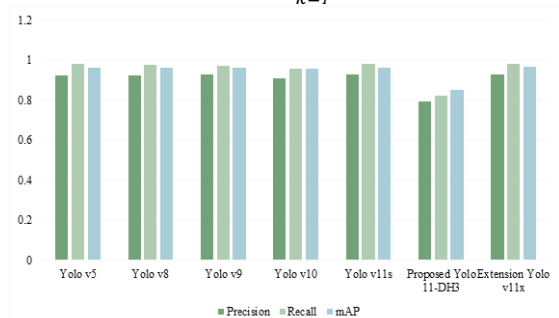
“Graph.1 Comparison Graph”

instances of a given class. It measures how well the system is able to identify relevant positive examples, based on the ratio of true positives to the number of correct prediction of positive examples made by the model. It's the number that measures how well a model can spot all the positive observations.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} (3)$$

mAP: One statistic used to evaluate the quality of a ranking is MAP. The context of the recommended items in the list and the quantity of items are taken into account. The mAP@K is the average of Average Precision at K per user (or per query) for all users (or queries) in the evaluation set.

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k$$



The comparison graph presents precision, recall and mAP of YOLO variations and YOLOv11x is the most precise model while YOLOv11-DH3 is less precise than other architectures evaluated.

FIRE & SMOKE DETECTION

Upload an image to detect fire and smoke using our advanced YOLOv11 model. Our AI-powered system analyzes your images to identify potential fire hazards and smoke patterns with high accuracy.

The detection system processes images through a trained deep learning model that can recognize fire and smoke in various environments, including residential areas, industrial facilities, and outdoor settings.

IMAGE FILE REQUIREMENTS

- ✓ **Supported Formats:** JPG, JPEG, PNG
- ✓ **Recommended Size:** Up to 10 MB per file
- ✓ **Image Quality:** Clear, well-lit images work best
- ✓ **Resolution:** Minimum 320x240 pixels recommended
- ✓ **Content:** Images should contain visible fire or smoke for detection

The system will analyze your image and provide detection results with bounding boxes and confidence scores for identified fire and smoke regions.

UPLOAD YOUR IMAGE

DRAG & DROP YOUR IMAGE HERE

or

Browse Files

JPG, PNG, JPEG formats only

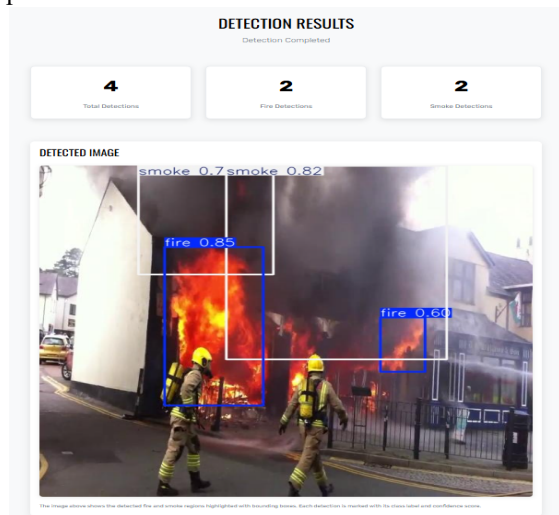
sample (21).jpg

0.04 MB

Detect Fire & Smoke

“Fig.3 Upload a Image”

Furthermore, the system has an image upload function at the web interface that allows the user to upload photos of a fire or smoke and automatically analyze the image in real-time for hazard detection.



“Fig.4 Predicted Result”

The output obtained shows the identified fire or smoke regions with the bounding boxes and the confidence value for the proper emergency management and fire safety monitoring.

5. CONCLUSION

Overall, the aim of this endeavor is to develop a reliable and effective smoke and fire detection system to ensure safety and timely warnings in complex scenarios. The approach proposed here relies on a well-designed set of fire and smoke images as a means to accurately localise the fire and

classify the smoke using deep learning object detection. Various YOLO architectures such as YOLOv5, YOLOv8, YOLOv9, YOLOv10, YOLOv11s, and an enhanced variant of YOLOv11-DH3 are trained and analysed to thoroughly evaluate the performance of the detection framework. The system ensures stability and dependability among models by adhering to dependable and consistent training procedures, data setups, and preprocessing techniques. Moreover, we use the latest model, YOLOv11x for better detection performance and generalization. Moreover, a user-friendly web interface (implemented with Flask) was provided for practical implementation, enabling the user to upload pictures to the handy web platform and get the real-time detection results. The results of the experiments show that the proposed system has a precision of 0.930, recall of 0.981 and mAP of 0.967, which demonstrates good detection performance. To briefly summarize, the proposed approach is an efficient, robust and scalable automatic fire and smoke detection system that could be used in building monitoring and decision making systems.

The long term goal is to fuse data from several sensor modalities such as thermal and infrared sensors to provide better detection capability and to improve the durability of the system in a variety of tough environmental circumstances. Optimisation techniques such as lightweight model designs can help in deploying models to edge devices for applications that require real-time and low latency performance. Interpretability is an important property of an explainable AI technique which instills confidence in the user and helps to understand the model's decisions. Furthermore, the framework's capacity to cope with continuous surveillance or video streams can help with proactive danger management. The cloud deployment and integration with mobile devices further enhance the accessibility and scalability of the system, encouraging its adoption in industrial application, metropolitan areas and environmental safety monitoring applications.

REFERENCES

- [1] Alkhamash, E. H. (2025). A Comparative Analysis of YOLOv9, YOLOv10, YOLOv11 for Smoke and Fire Detection. *Fire* (2571-6255), 8(1).
- [2] Li, Y., Nie, L., Zhou, F., Liu, Y., Fu, H., Chen, N., ... & Wang, L. (2025). Improving Fire and Smoke Detection with You Only Look Once 11 and Multi-Scale Convolutional Attention. *Fire*, 8(5), 165.
- [3] Hu, J., He, Y., Zeng, M., Qian, Y., & Zhang, R. (2024). Smoke and fire detection based on yolov7 with convolutional structure reparameterization and lightweighting. *IEEE Sensors Letters*, 8(8), 1-4.
- [4] Gao, P. (2024). A Fire and Smoke Detection Model Based on YOLOv8 Improvement. *International Journal of Advanced Computer Science & Applications*, 15(3).
- [5] Geng, X., Su, Y., Cao, X., Li, H., & Liu, L. (2024). YOLOFM: An improved fire and smoke object detection algorithm based on YOLOv5n. *Scientific Reports*, 14(1), 4543.
- [6] Government Set to Revise Total Number of Hectares Destroyed During Bushfire Season, 9News, Sydney, NSW, Australia, Feb. 2020.
- [7] 58 People Died in the High-Rise Residential Fire in Jing'an District, Shanghai, China News Service, Beijing, China, 2010.
- [8] Theon-Site Rescue Work of the Fire and Explosion Accident at Jilin Dehui Poultry Industry Company Has Basically Ended, Xinhua News Agency, Beijing, China, 2013.
- [9] V. K. R. Kodur and A. M. Shakyia, "Factors governing the shear response of prestressed concrete hollowcore slabs under fire conditions," *Fire Saf. J.*, vol. 88, pp. 67–88, Mar. 2017, doi: 10.1016/j.firesaf.2017.01.003.
- [10] N. Zheng, X. Guo, Y. Tie, N. Dong, L. Qi, and L. Guan, "Incremental generalized multiple maximum scatter difference with applications to feature extraction," *J. Vis. Commun. Image Represent.*, vol. 55, pp. 67–79, Aug. 2018, doi: 10.1016/j.jvcir.2018.04.012.
- [11] G. Lin et al., "Smokedetectiononvideosequences using3Dconvolutional neural networks," *Fire Technol.*, vol. 55, pp. 1827–1847, 2019, doi: 10.1007/s10694-019-00832-w.
- [12] H. Li and X. Li, "Research on smoke filtering algorithm for forest fire images based on improved CycleGAN," *Fire Sci. Technol.*, vol. 43, no. 11, pp. 1596–1602, 2024, doi: 10.20168/j.1009-0029.2024.11.1596.07.
- [13] M. Jeong, M. Park, J. Nam, and B. C. Ko, "Light-weight student LSTM for real-time wildfire smoke detection," *Sensors*, vol. 20, no. 19, p. 5508, Sep. 2020, doi: 10.3390/s20195508.
- [14] Y. Zhao and Y. Ban, "GOES-R time series for early detection of wildfires with deep GRU-network," *Remote Sens.*, vol. 14, no. 17, p. 4347, Sep. 2022, doi: 10.3390/rs14174347.
- [15] M. Jamali, N. Karimi, and S. Samavi, "Saliency based fire detection using texture and color features," in *Proc. 28th Iranian Conf. Electr. Eng. (ICEE)*, Aug. 2020, pp. 1–5.
- [16] J. Huang, J. Zhou, H. Yang, Y. Liu, and H. Liu, "A small-target forest fire smoke detection model based on deformable transformer for end-to-end object detection," *Forests*, vol. 14, no. 1, p. 162, Jan. 2023, doi: 10.3390/f14010162.
- [17] J. Lian, X. Pan, and J. Guo, "An improved fire and smoke detection method based on YOLOv7," in *Proc. 32nd Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2023, pp. 1–7.
- [18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis.*

Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[19] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, “Deformable convolutional networks,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 764–773, doi: 10.1109/ICCV.2017.89.

[20] R. Wang, R. Shivanna, D. Cheng, S. Jain, D. Lin, L. Hong, and E. Chi, “DCN v2: Improved deep & cross network and practical lessons for web scale learning to rank systems,” in Proc. Web Conf. (WWW), Ljubljana, Slovenia, Apr. 2021, pp. 1785–1797.

[21] H. Li, Y. Zhang, Y. Zhang, H. Li, L. Sang, and J. Zhu, “DCNv3: Towards next generation deep cross network for CTR prediction,” 2024, arXiv:2407.13349.

[22] P. Foggia, A. Saggese, and M. Vento, “Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion,” IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 9, pp. 1545–1556, Sep. 2015, doi: 10.1109/TCSVT.2015.2392531.

[23] S. Yang, Q. Huang, and M. Yu, “Advancements in remote sensing for active fire detection: A review of datasets and methods,” Sci. Total Environ., vol. 943, May 2024, Art. no. 173273, doi: 10.1016/j.scitotenv.2024.173273.

[24] R. Khanam and M. Hussain, “YOLOv11: An overview of the key architectural enhancements,” 2024, arXiv:2410.17725.

[25] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” IEEE Trans. Syst. Man, Cybern., vol. SMC-3, no. 6, pp. 610–621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.

[26] B. Chen, M. Wei, J. Liu, H. Li, C. Dai, J. Liu, and Z. Ji, “EFS-YOLO: A lightweight network based on steel strip surface defect detection,” Meas. Sci. Technol., vol. 35, no. 11, Nov. 2024, Art. no. 116003.

[27] S. Hao, X. Li, W. Peng, Z. Fan, Z. Ji, and I. Ganchev, “YOLO-CXR: A novel detection network for locating multiple small lesions in chest X-ray images,” IEEE Access, vol. 12, pp. 156003–156019, 2024.

[28] M. Tan, R. Pang, and Q. V. Le, “EfficientDet: Scalable and efficient object detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 10778–10787.

[29] K. Gajendiran, S. Kandasamy, and M. Narayanan, “Influences of wildfire on the forest ecosystem and climate change: A comprehensive study,” Environ. Res., vol. 240, Jan. 2024, Art. no. 117537, doi: 10.1016/j.envres.2023.117537.

[30] K. Muhammad, J. Ahmad, Z. Lv, P. Bellavista, P. Yang, and S. W. Baik, “Efficient deep CNN-based fire detection and localization in video surveillance applications,” IEEE Trans. Syst. Man, Cybern. Syst., vol. 49, no. 7, pp. 1419–1434, Jul. 2019.