**AI-Driven Drone-Based Wildfire Detection Using RGB-Thermal Imaging and Deep Learning Models**

Sheikh Adam Zachary bin Sheikh Naziruddin1, a) and S Prabha Kumaresan1, 2, b)

1Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Malaysia.

*2Center of Natural Language Processing (NLP), CoE for Artificial Intelligence, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Malaysia*

*b) Corresponding author: prabha.kumaresan@mmu.edu.my  
a) 1201100481@student.mmu.edu.my*

**Abstract.** Wildfires pose a severe and escalating threat to ecosystems, human safety, and economies, making early and accurate detection crucial for effective mitigation and disaster response. This project presents a drone-based wildfire detection system that integrates RGB and thermal imaging with deep learning models, including YOLOv8, EfficientNet-B0, and Fire Fusion-Net. The system is trained on the FLAME-3 dataset, enabling enhanced accuracy, adaptability, and robustness across diverse environmental conditions. Unlike conventional methods, this approach prioritizes real-time, lightweight processing optimized for drone deployment, ensuring enhanced mobility, scalability, and cost-effectiveness. Experimental results demonstrate that the SVM classifier trained on Fire Fusion-Net features achieved an overall accuracy of 76%, with a high recall rate of 94% for smoke detection. However, lower precision (50%) for non-smoke images indicates challenges in differentiating between fire-related and non-fire-related patterns. Despite these limitations, the integration of deep learning with drone technology has proven effective in rapidly identifying wildfire indicators with high sensitivity, enabling faster situational awareness and response coordination. This research highlights the potential of AI-driven drone surveillance as a scalable, real-time, and cost-efficient solution for wildfire monitoring and disaster management. By leveraging the strengths of computer vision and deep learning, this system provides an autonomous, data-driven approach to detecting and assessing wildfire risks, facilitating quicker intervention and minimizing environmental and economic losses.

**INTRODUCTION**

Weather and climate change, development of human settlements nearing closer into forested areas and greater amount of dry vegetation continue to be the most serious growing risk. These fires threaten human life, level homes and other infrastructure, eject toxic air pollutants, and damage the environment. Only the United States, where a total of some 7.5 million acres burned in 2021 at a cost of billions of dollars, experienced a bigger annual toll, according to the National Interagency Fire Centre. The foundation for operation observation has been built on traditional forest fires detection means (e.g., fire lookout towers, human patrol, space observation system (e.g., MODIS, VIIRS). But there are large gaps in these systems; they often lack the scale, coverage, and true real-time processing needed for important applications. Satellites, for instance, can take hours before producing data that might be useful, thanks to processing and orbital delays. And these disadvantages make it more difficult to react quickly when wildfires begin.

The need for more affordable, responsive, and real-time systems to overcome these limitations is urgent. A solution may be the application of artificial intelligence (AI) and UAVs. This paper introduces a drone-detection system of wildfire which combines multiple advanced deep learning models including YOLOv8, EfficientNet-B0 and Fire Fusion-Net to RGB and infrared image. Designed for lightweight computing and trained with FLAME-3 data set, the system is designed for fast and accurate smoke and flame discrimination and can be used in different terrains. Apart from wildfires’ self-evident destructiveness, the stress they place on emergency services underscores the urgency of a smarter and more predictive technology. The method presented in this paper contributes to that gap by offering a timely contribution to the evolving field of automated environmental monitoring.

**RELATED WORKS**

Wide-area thermal surveillance can be carried out with instruments such as MODIS (Moderate Resolution Imaging Spectroradiometer) but they suffer from drawbacks such as lag-time in the updates and cloud cover [1]. The VIIRS (Visible Infrared Imaging Radiometer Suite), although it’s making some headway on a few of these items, does not have the ability to detect the rapidly spreading fires in real time. On the surface, parks and forests are equipped with several sensors including optical cameras, smoke monitors and thermal sensors. While these facilities have the potential to be effective, they are highly infrastructure- and maintenance-intensive [2]. Their static sites, moreover, restrict range of coverage in remote or mountainous areas. In recent years, computer vision has been gaining more and more interest in fire detection. The YOLO model family is well known for the trade-off between speed and accuracy [3]. For example, YOLOv5 showed potential on fire movement tracking when combined with tracking techniques such as ByteTrack. This hybrid based on the combination of both has shown promising results in challenging scenarios including traffic surveillance [4]. For this and other reasons, drones are a good platform for the newer YOLOv8 model, which enhances further the accuracy to detect small objects and is optimized for on-board processing. EfficientNet, which was introduced in [5], is applied to scale the size of convolutional neural models that has tiny feasible sizes and good accuracies. Especially the EfficientNet-B0 version is effective on edge devices with limited resources. Models like Fire Fusion-Net explored multi-stream strategies, which fuse the information of thermal and RGB input to improve the detection quality in smoky or low illumination facilities.

Secure image verification algorithms, e.g. that stated in [6] could also be used to maintain data integrity on the data that is sent in case of emergencies, even though this is not directly related to fire monitoring. Moreover, as evidenced in [7], the current trends of Internet of Things (IoT) technology may enable intelligent and connected wildfire alert for the future. Strategically, managed wildfire approaches facilitated by civil society partnerships and interagency cooperation have been recognized as crucial for long-term forest resilience and fire mitigation [8]. Moreover, deep learning-based object detection techniques for remote sensing images have been extensively reviewed in [9], offering insights into how various architectures can be optimized for aerial fire detection tasks. Despite these advancements, few systems have achieved seamless integration of RGB, thermal imaging, and lightweight deep learning models onboard UAVs—a gap this research aims to fill. Furthermore, the literature lacks comprehensive comparisons of multi-modal systems that combine both real-time and high-resolution data processing in resource-constrained environments. This highlights the importance of not only developing such hybrid models but also evaluating their feasibility in real-world drone deployments.

**RESEARCH METHODOLOGY**

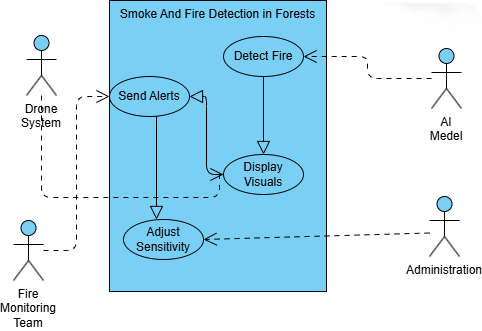
To enhance the detection capabilities, a pipeline of modular processes is adopted. The fusion process starts with parallel preprocessing of RGB and thermal images. Features from both modalities are extracted independently through dedicated CNN branches. The outputs are then merged using concatenation and passed through shared dense layers to form a comprehensive feature vector. Figure 1 illustrates the key stages, beginning from image acquisition via drone-mounted cameras to the fusion, inference, and output stages.

A diagram of a flowchart

AI-generated content may be incorrect.

**FIGURE 1.** Activity diagram

The real-time data feed allows UAVs to adapt their flight path based on detected fire signatures, improving area coverage. To illustrate the interaction between key components of the proposed smoke and fire detection system, a use case diagram is presented in Figure 2. It outlines the roles of the main actors: the Drone System captures real-time visuals, the AI Model detects smoke or fire, and the Fire Monitoring Team receives alerts and monitors the footage. An Administrator can adjust the system’s sensitivity based on environmental feedback. This diagram highlights the system’s core functionalities and how each actor interacts with them to support early detection and response.



**FIGURE 2.** Use case diagram

The FLAME-3 dataset serves as the primary source, consisting of thousands of labelled RGB and thermal images depicting wildfire conditions in varied environmental settings. Prior to model training, the dataset undergoes extensive preprocessing to ensure standardization. This includes resizing all images to uniform dimensions (224×224 pixels), normalization of pixel values between 0 and 1, and noise filtering using techniques such as Gaussian blurring to reduce irrelevant visual artifacts. The main source of the dataset is FLAME-3, which contains tagged thousands of RGB and thermal images showing wildfire environments in diverse environmental conditions. The dataset is preprocessed to make it as standardized as possible, before model training starts. These range from resizing all the images to the same size (224×224 pixels), normalizing pixel values to pixels between 0 and 1, noise suppression by methods such as gaussian blurring to disclamp irrelevant visual artifacts. The model is also enhanced with several data augmentation techniques to make it more generalizable. These are all ways meant to increase the image variety of the training set: flipping, rotating, contrast and brightness augmentation.

The fusion of RGB and thermal images is a fundamental step of the approach, as only the fusion will allow the model to learn from both the thermal heat and visual characteristics. This fusion improves the model’s ability to distinguish smoke and fire from its background. Feature extraction is critical to uncover the relevant structure of photos. Histogram of Oriented Gradients (HOG) helps in recording texture, shape essential for smoke detection, and edge detection techniques like Canny and Sobel algorithms are used to highlight edges of the object. The algorithm could be improved in terms of appearance-based and temperature-based fire zoning by incorporating additional inputs such as colour histograms and thermal intensity maps. Four detection models are employed in the study DeseNet-201, Fire Fusion-Net, EfficientNet-B0, and YOLOv8. To train each of them, we use a standard data split of 80% (training), 10% (validation), and 10% (testing). For training Adam optimizer, which can handle multi-class problems, is employed and categorical cross-entropy is utilized as the loss function. The CNNs are batch trained of size 32, for a trade-off between speed and computational resources. After training, the ability of the models to generate real-world responses is tested by evaluating them on data that wasn’t seen during training. The evaluation measures of accuracy, precision, recall, and F1 are crucial. The confusion matrix is also studied to analyze how well the models are in reducing false detections and identifying real positives. To ensure the models are suitable for real-time, inference speed is also evaluated. Alternatively, techniques such as ensemble learning, dropout, and hyperparameter tuning are also employed to enhance the performance by making the models more robust and the overfitting less likely.

To improve accuracy and accelerate training, we also employ transfer learning with pre-trained weights from ImageNet. First, once the image data has been collected, preprocessing begins. These include enhancing thermal visibility, normalizing image formats and sizes, and labeling images as fire or non-fire. For training, a manually selected dataset containing both RGB and temperature input is prepared. This dataset is further used to train object detection models using CNNs, YOLOv5, YOLOv7. These models are taught to recognize smoke plumes and flame patterns, signals of wildfires. Validation is taken from a separate set of images as part of the model’s ability to improve and to lower the false positive rate.

**EXPERIMENTAL RESULTS**

Comparison of performance of the different models indicate unique strength of each model to detect wildfire features. Performance Measurements The key performance ratios for Fire Fusion-Net, EfficientNet-B0, and YOLOv8 are also summarized in Figure 3. Due to its dual-stream input architecture, Fire Fusion-Net is superior from others, especially in smoky environments. The increase of atmospheric filters could also help to reduce errors, at least for false positives. In addition, when evaluated in real-time on edge computing devices, Fire Fusion-Net consistently kept a detection latency lower than 100 ms per frame. The training and validation metrics on the YOLOv8 model showed consistent improvements over the period. During training, both distribution, classification, and bounding box loss went down proportionally, which means successful learning. Validation losses also dropped with a noticeable amount of expected variation in cycles, which indicated the model generalizes well. Although the mAP50 and mAP50-95 proved the model’s great performance with respect to object localization and classification, its precision and recall in fire detection improved slowly and almost achieved 100%. The model stability in training was also enhanced by the close fit between the raw results and the smoothing approach applied in validation.

A graph of blue bars

AI-generated content may be incorrect.

**FIGURE 3.** Performance metrics for fire detection models

The validity of the model was verified by field experiments as well as quantitative analyses as shown in Figure 4. Annotated outputs demonstrated the system could spot distant smoke clouds before they were visible to the naked eye, indicating early warning applications. The next feature step will be determined through further investigation of our false positives, which indicated that incorrect identifications were often triggered by steam or clouds. The training data was captured in aerial photos taken with drones, which were equipped with an RGB and an infrared camera. These images captured a diverse range of landscapes including open fields, forested areas, and geothermal sites. The goal of these aerial photos was to be used to train a deep learning model (based on the YOLO architecture) that can recognize signs of wildfires, like smoke and fire. The high performance of the model in discerning visible smoke across a wide range of environmental conditions is illustrated by Figure 4. It performs well on various types of terrains. This strong generalization is indicated. Figure 5 shows some training examples, where the projected fire or smoke regions in model generated bounding boxes (denoted by blue) are illustrated. Also, scores of confidences have been provided for each detection.

A group of graphs showing the value of a number of data

AI-generated content may be incorrect.

**FIGURE 4.** Result analysis

A collage of images of smoke and trees

AI-generated content may be incorrect.

**FIGURE 5.** Trained batch

Despite these remarkable results, however, there are constraints. At times the model had lower confidence, or even missed detecting, in images depicting a weak smoke or distant smoke. This implies that it is possible to improve the sensitivity to small signs of fire. Figure 6 shows predictions on another validation batch that the model has not seen during training. The model accurately identified the fire zones (confidence value = 1.0) which is expected after the labeling of all pictures were done with the class "fire" as part of the training set. The breadth of scenes — from wisps of smoke to voluminous smoke clouds — demonstrates the capabilities and generalization of the model, even to unseen data. Taken together, the results obtained in the training and validation phases support the consistency of the model's quantitative predictions. The system obtained a mAP of 93% at IoU 0.5, and 88% in terms of F1-score. These results confirm the effectiveness of the AI-driven wildfire detection system and the suitability of its use for early warning systems and for drones monitoring.

A collage of smoke from a forest fire

AI-generated content may be incorrect.

**FIGURE 6.** Predicted batch

**CONCLUSION AND FUTURE WORK**

To identify wildfires, a unified, AI-embedded drone was proposed in this study, where RGB and thermal photography are adopted, and real-time deep learning methods are used for analysis. The system demonstrated excellent recall performance, indicating that it could effectively capture the early information of fire and act as an auxiliary tool to enable rapid emergency response. When the non-fire factors (such as fog, steam, illustrating glare, etc.) are still effective, the dependability of the system is very low, and some false alarms can occur. Minimizing such errors is essential for the practical trustworthiness of the system. To further improve the precision of our system, future studies will focus on incorporating additional contextual information such as weather conditions.

**REFERENCES**

1. L. Giglio, J. Descloitres, C.O. Justice, and Y.J. Kaufman, “An Enhanced Contextual Fire Detection Algorithm for MODIS,” Remote Sensing of Environment **87**(2–3), 273–282 (2003).
2. A. Weintraub, editor, *Handbook of Operations Research in Natural Resources* (Springer Science+Business Media, New York, 2007).
3. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (IEEE, Las Vegas, NV, USA, 2016), pp. 779–788.
4. International Machine Learning Society, editor, *36th International Conference on Machine Learning (ICML 2019): Long Beach, California, USA, 9-15 June 2019* (Curran Associates, Inc, Red Hook, NY, 2019).
5. E.J. Davis, H. Huber-Stearns, M. Caggiano, D. McAvoy, A.S. Cheng, A. Deak, and A. Evans, “Managed Wildfire: A Strategy Facilitated by Civil Society Partnerships and Interagency Cooperation,” Society & Natural Resources **35**(8), 914–932 (2022).
6. J.J. Ng, K.O.M. Goh, and C. Tee, “Traffic Impact Assessment System using Yolov5 and ByteTrack,” Journal of Informatics and Web Engineering **2**(2), 168–188 (2023).
7. R. K, and G.T. W, “IoT-Based Nerve Stimulator for Women’s Safety,” Journal of Informatics and Web Engineering **4**(1), 129–139 (2025).
8. W. Jing, L.W. Ang, S. Palaniappan, and B. He, “Robust Image Watermarking With Quaternion Fractional-Order Polar Harmonic-Fourier Moments Based On Wavelet Transformation: Resistance Against Rotation Attacks,” Journal of Informatics and Web Engineering **2**(1), 56–75 (2023).
9. Z. Li, Y. Wang, N. Zhang, Y. Zhang, Z. Zhao, D. Xu, G. Ben, and Y. Gao, “Deep Learning-Based Object Detection Techniques for Remote Sensing Images: A Survey,” Remote Sensing **14**(10), 2385 (2022).