# Safety-Oriented Prompt Design for Small Vision–Language Models in Early Wildfire Smoke Detection

Anonymous Authors

***Abstract*—Early wildfire smoke detection is a crucial safety task that allows for timely intervention before small-scale ignitions become major catastrophes. Current wildfire detection systems tend to depend on computationally demanding models, fine-tuning processes, or fixed inference behaviour, which limits their adaptability and deployment on limited-capacity edge platforms.**

**This paper investigates inference-time prompt design as a lightweight approach to control detection behaviour in a small vision–language model in a zero-shot setting without any fine-tuning. Using a compact 4B-parameter model, three prompt strategies—Baseline, Recall-Boost, and Balanced—are evaluated on the FIgLib-Test wildfire smoke dataset. The proposed approach enables explicit control over safety-oriented trade-offs between missed smoke detections and false alarms under identical inference conditions by modifying only the textual prompt.**

**Quantitative results and qualitative visual analysis illustrate a significant reduction in missed smoke detections while maintaining practical operational behaviour when applying recall-oriented and balanced prompt strategies. The findings highlight prompt-controlled inference as an efficient and flexible solution for safety-critical multimodal perception in real-world wildfire monitoring.**

***Index Terms*—Vision–language models, Small Language Model, multimodal intelligence, safety-critical systems, prompt-based inference, edge AI, wildfire smoke detection**

## I. Introduction

Wildfire smoke detection is a critical safety early-warning task that enables rapid intervention before small-scale fires escalate into large-scale events. In this paper, we investigate how inference-time behavior can be controlled in a small vision–language model using prompt strategies, without any fine-tuning.

### *A. Wildfire Smoke as a Safety-Critical Problem*

Due to climate change, the world has witnessed critical environmental problems. one of the most dangerous problems is wildfires. The increase in drought and heatwaves made wildfires a global problem. The Mediterranean and North African regions are considered the most vulnerable regions. In the past seven years, Tunisia has seen severe wildfires. In the summer of 2022, hundreds of hectares of pine forest were burned in the Al Kef governorate in the middle western region of Tunisia [1]. Statistical evidence suggests expanding burned areas and an increase in the number of forest fires over recent seasons; during the summer of 2023, Civil Defense teams recorded 693 forest fires due to heat wave conditions [2]. More recently, in mid-2025, Tunisian authorities recorded more than 550 hectares of agricultural land burned by wildfires [3]. Significant efforts have been made to contain this global problem, with most solutions leverage deep learning techniques to detect flame or smoke, to enable early wildfires containment.

### *B. Limitations of Existing Detection Systems*

The advancement of deep learning techniques coupled with the availability of satellite datasets introduces a ground opportunity to develop detection systems able to detect wildfires and prevent disasters. A wide area coverage along with long term fire statistics can be achieved by sensors such as MODIS. Satellite based wildfire detection system play a crucial role in monitoring large scale fire and perform post fire assessment. However, there are many limitations of existing systems that limit their efficiency in detecting smoke at an early stage since their revisiting time is measured in hours. The system's real-time applicability for quick response is diminished because of the delay in data collecting and processing, which requires fires to be substantially ignited before they can be spotted from space [4]. This allows wildfire to spread and reduce the controlling chances. Furthermore, their sensitivity to thin or low-altitude smoke plumes is further diminished by cloud cover interference and spatial resolution limitations. Because of this, satellite-based systems are mainly reactive and cannot deliver timely notifications for situations requiring early response or immediate intervention [5].

To overcome the limitations of timely notifications from satellite systems, many studies have proposed camera and drone-based wildfire and smoke detection systems using Convolutional Neural Network (CNN) models and YOLO as an object detection framework. These systems have achieved promising accuracy under controlled conditions. However, their real-life implementation faces several technical limitations.

Firstly, when implemented in new locations with varying camera viewpoints, lighting conditions, or weather conditions, the majority of CNN-based systems frequently require retraining or fine-tuning and substantial amounts of labeled training data. Secondly, when detecting smoke which is the earliest sign of a wildfire these models tend to generalize poorly to thin or distant smoke and often fail to differentiate between smoke and fog or clouds. Moreover, these models are primarily designed to maximize accuracy, with limited control over safety-related trade-offs. CNN-based approaches also require

high computational resources, reducing their suitability for real-time deployment on edge platforms [6].

### C. Safety-Oriented Trade-Off Perspective

A variety of detection errors have different impacts in safety-critical wildfire monitoring systems. False negative errors, in particular, can result in a delayed response and allow fires to spread uncontrollably, causing serious harm to the environment, agriculture and the economy. On the other hand, false positive detections, which triggers alarms when there is no real fire or smoke, usually result in controllable operational overhead like temporary system alerts or manual verification. Because of this, there are inherent asymmetric error costs in wildfire smoke detection, with missed detections being much more important than false alarms.

Despite this imbalance, the majority of current wildfire detection systems are created and assessed using global performance metrics like accuracy, precision, or F1-score, which essentially assume that various error types are equally important. The conservative detection behaviour that prioritises lowering false alarms at at the cost of early smoke sensitivity is frequently the result of optimising such metrics. Strict decision thresholds raise the possibility of missed detections, and this design decision is especially hazardous for prompt warning scenarios where smoke may appear faint, far away, or partially obscured.

From an operational point of view, various safety requirements must be supported by real-world wildfire monitoring systems based on deployment contexts, risk levels, and environmental conditions. For instance, while lower-risk situations might allow for more cautious operation, times of intense heat or drought might necessitate aggressive detection behaviour that prioritises early smoke recognition. However, because their decision boundaries are set during training and fine-tuning, traditional deep learning-based techniques lack mechanisms to dynamically modify their behaviour at inference time.

Motivated by these limitations, this paper follows a safety-focused approach and specifically addresses wildfire smoke detection as a trade-off control issue. We investigate inference-time behaviour modulation through prompt design in a small vision-language model instead of optimising a single operating point during training. Without changing any parameters or retraining, the model can be directed towards conservative, safety-first, or balanced detection behaviour by modifying prompt semantics. This approach is especially appropriate for wildfire monitoring applications that are restricted in resources and safety-critical because it allows for adaptability and lightweight control over detection sensitivity.

### D. Research Gap and Contributions

There are still a number of significant gaps in the vast amount of research on smoke and wildfire detection. While camera-based deep learning techniques heavily rely on large labelled datasets and fixed decision limits defined during training, current satellite-based systems are missing the temporal accuracy necessary for early smoke detection. While vision-language models for smoke and fire detection have been explored recently, these methods usually rely on prompt or fine-tuning procedures and do not specifically address safety-oriented trade-offs between false alarms and ignored detections. Specifically, inference-time control of behaviour for small vision-language models used in safety-critical wildfire monitoring scenarios has not been thoroughly studied.

In order to overcome these limitations, this work reframes wildfire smoke detection as a safety-oriented trade-off control problem and emphasises inference-time model behaviour over training-time optimisation. We examine how prompt design alone can be used to modulate detection sensitivity in a compact vision-language model, allowing flexible control over conservative and safety-first operating modes without additional computational cost, as opposed to changing model parameters. The following is a summary of this paper's primary contributions:

- We present a safety-focused approach to early wildfire smoke detection that takes into consideration the asymmetric error costs between false alerts and missed detections.
- we present several inference-time prompt strategies, in order to direct a small vision-language model towards conservative, safety-first, and balanced detection patterns without any fine-tuning.
- We use both quantitative and qualitative analysis to examine the effect of prompt design on detection behaviour through an experimental evaluation on a wildfire smoke dataset.

## II. Related Work

### A. Vision-Based Wildfire and Smoke Detection

The majority of early wildfire detection research relied on vision-based techniques that used manually created features taken from pictures or video streams. These methods used colour, texture, and motion cues to detect flames or smoke patterns, but their robustness in actual outdoor settings was limited because they were extremely sensitive to environmental changes like lighting, background complexity, and weather effects. In addition, most datasets in deep learning are imbalaced which is considered an obstacles to achieving a safety critical task [7].

With the advancement of deep learning, Convolutional neural networks (CNNs) are now the most common approach for detecting smoke and wildfires. CNN architectures and object detection frameworks, such as YOLO-based models, have been used in numerous studies to identify smoke or fire in surveillance photos and videos. [8], [9]. Under controlled circumstances and benchmark datasets, these techniques have shown encouraging accuracy. However, their performance has not been tested using different dataset partitioning [7]. However, current CNN-based systems typically require a lot of labelled training data and a lot of fine-tuning to adjust to different environments. [10]. likewise, in early wildfire scenarios, their performance typically decreases when dealing with

thin, distant, or partially occluded smoke [11]. Generalisation, robustness, and early smoke sensitivity are frequently identified as the main unsolved issues in comprehensive reviews of vision-based fire detection [7].

### B. Vision–Language Models in Safety-Critical Tasks

In the past few years, vision–language models (VLMs) have become a potent paradigm for visual understanding by simultaneously merging image content and natural language processing. By utilising semantic alignment between visual features and textual prompts, models like CLIP followed by multimodal architectures allow zero-shot or few-shot reasoning. [12], [13]. Their application to safety-critical tasks, such as hazard recognition, environmental monitoring, and disaster response, has been driven by this capability [14].

Several studies have investigated the application of vision-language models to enhance generalisation across various scenes and visual conditions in the context of wildfire detection [14].

To adapt pretrained models for fire or smoke detection tasks, these methods typically rely on prompt tuning or fine-tuning techniques. Although encouraging, most existing vision–language and large language model–based wildfire detection approaches depend on large model variants and additional training stages, which significantly increase computational and memory requirements. As a result, their deployment on resource-constrained edge platforms is often impractical due to high inference latency, elevated power consumption, and substantial hardware demands [15]. Furthermore, rather than specifically examining safety-oriented trade-offs between missed detections and false alarms, previous work has primarily focused on increasing detection accuracy [14], [16], [17].

### C. Test-Time Prompt Adaptation and Compact Models

In order to control model behaviour at inference, test-time prompt adaptation has been considered as a lightweight substitute for parameter fine-tuning. Prompt design has been demonstrated to improve model predictions in both language-only and vision-language settings by highlighting particular semantic concepts or decision preferences without affecting model weights. [18]–[20].When compared to retraining-based techniques, these strategies provide flexibility and lower computational cost.

Although prompt-based adaptation is gaining popularity, current research focuses more on accuracy, domain generalisation, or performance adjustment than safety-critical trade-off control. Moreover, lightweight or small vision-language models have received little attention, especially when it comes to wildfire smoke detection. [21]. Therefore, there is still much to learn about how inference-time prompt strategies may impact safety behaviour and modulate detection sensitivity in SLM models.

Unlike previous work, this paper explicitly emphasises safety-oriented trade-offs related to early wildfire smoke detection while focussing on prompt-driven inference-time behaviour control in a small vision–language model. The suggested method seeks to enable flexible and effective deployment in real-world, resource-constrained monitoring systems by skipping any kind of fine-tuning.

## III. Problem Formulation

### A. Binary Smoke Detection Task

The task is formulated as a binary smoke classification problem. Two classes are defined for this purpose: smoke present and smoke absent. Given an input RGB image captured by fixed-view ground cameras, the objective is to determine whether visible smoke is present in the image. Let $x \in \mathbb{R}^{H\times W\times 3}$ denote an input image of height $H$, width $W$, and three color channels. The corresponding output label $y \in \{0, 1\}$ indicates the absence ($y = 0$) or presence ($y = 1$) of wildfire smoke.

Unlike flame detection, which can often be classified as flame or no flame with limited ambiguity, smoke detection presents additional challenges due to the semi-transparent and visually ambiguous nature of smoke, especially during the early stages of ignition. Smoke may appear thin, distant, or partially occluded, and can be visually similar to clouds, fog, or atmospheric haze. This visual ambiguity can confuse detection models and lead to delayed or missed predictions. As a result, early smoke detection is particularly sensitive to decision thresholds and model behavior, as subtle visual cues must be interpreted under uncertain conditions.

In this work, a small vision–language model is employed to address binary smoke detection by performing semantic interpretation of visual content guided by a textual prompt. The model output is influenced by the prompt formulation at inference time, rather than relying on fixed decision boundaries learned through fine-tuning.

### B. Safety-Oriented Operating Modes

Wildfire smoke detection systems are considered safety-critical tasks in which different types of classification errors have asymmetric consequences. In particular, a false negative error, where early smoke is missed, can allow wildfires to spread uncontrollably by delaying intervention, causing severe environmental, agricultural, human, and economic damage and loss. On the other hand, false positive errors incur limited operational costs, such as human or drone-based verification [22]. Therefore, early wildfire monitoring systems must prioritize minimizing false negative errors, as the cost of failing to detect smoke is significantly higher than that of issuing false alarms.

Despite this asymmetry, most existing wildfire detection systems implicitly assume equal importance of all error types by focusing on global performance metrics such as accuracy, recall, or F1-score. As a result, these approaches tend to favor detection behaviors that reduce false positives at the expense of missing early or subtle smoke. In addition, CNN-based and vision–language-based detection models typically operate with fixed decision boundaries determined during training or fine-tuning. Adapting their behavior to different safety

TABLE I
CLASS DISTRIBUTION OF WILDFIRE SMOKE LABELS IN THE FIGLIB-TEST DATASET.

| Class | Label | Number of Images |
|---|---|---|
| No Smoke | No | 2472 |
| Smoke | Yes | 2408 |

requirements therefore requires retraining, threshold tuning, or additional calibration stages [23].

In this paper, we introduce the concept of a safety-oriented detection model that adjusts detection behavior according to the operational risk level. A safety-oriented mode prioritizes recall by favoring smoke detection under uncertainty and tolerating higher false alarm rates, whereas a conservative mode prioritizes precision by requiring strong visual evidence before declaring the presence of smoke [18], [20]. The proposed approach enables inference-time behavior control through prompt design in a small language model operating in a zero-shot setting, without any parameter fine-tuning or retraining. This lightweight mechanism is well suited for real-time wildfire monitoring on resource-constrained edge platforms, as it provides flexible control over safety-related trade-offs.

## IV. METHODOLOGY

### A. Dataset Description

The FIgLib-Test dataset is used for the experimental evaluation in this work. FIgLib-Test is a publicly available dataset for wildfire smoke detection, consisting of sequences of RGB images of wildland fires captured by fixed-view cameras deployed on distant mountain tops in Southern California as part of the High Performance Wireless Research and Education Network (HPWREN) [24]. The images are associated with annotations indicating the presence or absence of visible wildfire smoke. For each sample, a ground-truth dictionary is provided in which the attribute *forest_fire_smoke_visible* is labeled as either "Yes" or "No". The class distribution of the FIgLib-Test dataset is reported in Table I, illustrating a near-balanced split between smoke and no-smoke images.

The characteristics of the FIgLib-Test dataset make it particularly suitable for evaluating safety-oriented detection behavior, where subtle visual cues must be interpreted under uncertainty. These characteristics include visually challenging smoke scenarios associated with early wildfire stages, such as thin, distant, and partially occluded smoke plumes, as well as negative samples containing clouds, haze, or atmospheric artifacts that may visually resemble smoke. To ensure a fair comparison across different prompt strategies, all experiments are conducted using the same dataset split, and only the smoke detection attribute is utilized in this study.

### B. Small Vision–Language Model

A small vision–language model, Gemma-3-4B-IT, is employed to perform binary wildfire smoke detection in a zero-shot inference setting. To achieve the objective of minimizing computational overhead while retaining semantic reasoning capabilities suitable for edge-oriented deployment scenarios, a small language model (SLM) is selected.

Given an input image and a textual prompt, the model generates a short textual response. In this study, the output is constrained to a structured JSON format containing a single key, *forest_fire_smoke_visible*, with categorical values "Yes" or "No". To guarantee that any distinction in detection behaviour is exclusively due to modifications in the prompt design, the same model weights, and inference configuration are employed in every experiment.

### C. Prompt Strategy Design

The design of prompt strategies that control model behavior at inference time without modifying model parameters constitutes the core contribution of this work. The small vision–language model (SLM–VLM) is guided toward a safety-oriented operating mode through prompt semantics, enabling explicit trade-off control between missed smoke detections and false alerts.

Three prompt strategies are evaluated:

- **Baseline Prompt (Conservative Mode):** The baseline prompt requires clear and strong visual evidence before declaring the presence of smoke. Although this approach favors accuracy and reduces false alarms, it may fail to identify subtle or ambiguous smoke patterns.
- **Recall-Boost Prompt (Safety-First Mode):** The recall-boost prompt explicitly emphasizes early smoke detection and prioritizes sensitivity under uncertainty. To reduce false negatives at the expense of increased false positives, the model is instructed to prioritize smoke detection when moderate visual cues are present.
- **Balanced Prompt (Trade-Off Mode):** The balanced prompt represents an intermediate strategy that encourages cautious detection while still responding to moderate smoke cues. The objective of this prompt is to achieve a practical real-world compromise between conservative and safety-first behavior.

All prompt strategies are applied to the same model under identical inference conditions. No prompt tuning loops, optimization steps, or fine-tuning procedures are performed.

### D. Experimental Setup

To ensure efficient processing, all experiments are conducted in an inference-only setup under identical conditions using GPU A100 acceleration. To maintain memory safety and adhere to the model's input constraints, images are resized to a maximum spatial dimension prior to inference. For each image, the binary smoke decision is extracted by post-processing the model's textual response to parse the structured JSON output. Predicted labels are compared against ground-truth annotations using standard classification metrics, including accuracy, precision, recall, F1-score, and confusion matrices, to perform quantitative evaluation.

TABLE II
QUANTITATIVE PERFORMANCE COMPARISON OF DIFFERENT PROMPT STRATEGIES ON THE FIGLIB-TEST DATASET.

| Prompt Strategy | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| Baseline Prompt | 68.4 | 94.8 | 38.1 | 54.3 |
| Recall-Boost Prompt | 49.4 | 49.4 | 99.6 | 66.0 |
| Balanced Prompt | 67.4 | 68.9 | 61.7 | 65.1 |

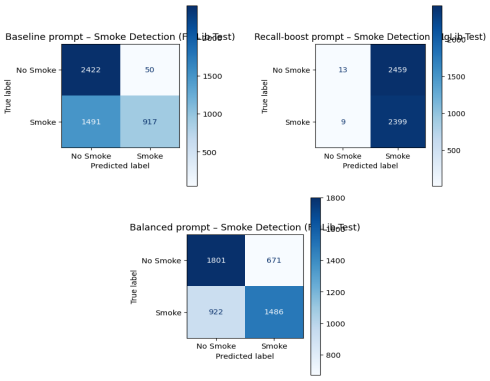


Fig. 1. Confusion matrix heatmaps for smoke detection using (left) Baseline, (middle) Recall-Boost, and (right) Balanced prompt strategies on the FIgLib-Test dataset.

## V. RESULTS AND DISCUSSION

### A. Quantitative Results and Comparison with State of the Art

This subsection reports the quantitative evaluation of the three prompt strategies (Baseline, Recall-Boost, and Balanced) on the FIgLib-Test dataset. Performance is summarized in Table II using accuracy, precision, recall, and F1-score under identical inference settings.

The Baseline prompt is more conservative, typically reducing false alarms but increasing missed smoke cases. The Recall-Boost prompt prioritizes sensitivity, reducing false negatives at the cost of more false positives, while the Balanced prompt provides an intermediate trade-off.

The corresponding confusion matrices are shown as heatmaps in Fig. 1, which illustrates how errors change depending on the operating mode. Specifically, compared to the Baseline approach, Recall-Boost and Balanced reduce false negatives by reallocating errors towards false positives, which are typically more acceptable in safety-critical wildfire system.

We compare our findings with recent state-of-the-art vision-language model (VLM) based smoke detection techniques evaluated on the FIgLib-Test dataset in order to position the proposed approach in context. Strong performance has been reported in previous work using refined models like ForestFireVLM-3B and ForestFireVLM-7B, where enhancements are made through task-specific training, fine-tuning, and increased model capacity. On the other hand, a number of large language models (LLMs) using static inference settings show high precision but relatively low recall, which is not ideal for early smoke detection in wildfires [25].

TABLE III
COMPARISON WITH STATE-OF-THE-ART SMOKE DETECTION METHODS ON THE FIGLIB-TEST DATASET.

| Model | Training | Acc. | Prec. | Rec. | F1 |
|---|---|---|---|---|---|
| ForestFireVLM-7B | Fine-tuned | 78.5 | 98.6 | 57.2 | 72.4 |
| ForestFireVLM-3B | Fine-tuned | 76.3 | 98.8 | 52.6 | 68.6 |
| Gemini 1.5 Pro | – | 70.0 | 100.0 | 39.2 | 56.3 |
| GPT-4o | – | 74.5 | 95.2 | 50.6 | 66.1 |
| Qwen2.5-VL-7B | – | 60.3 | 100.0 | 19.5 | 32.7 |
| Gemma-3-4B-it (Baseline) | Zero-shot | 68.4 | 94.8 | 38.1 | 54.3 |
| Gemma-3-4B-it (Recall) | ZS + Prompt | 49.4 | 49.4 | 99.6 | 66.0 |
| Gemma-3-4B-it (Balanced) | ZS + Prompt | 67.4 | 68.9 | 61.7 | 65.1 |

In contrast to these methods, our approach uses a strict zero-shot setting with a lightweight 4B-parameter model. By modifying only the inference-time prompt, the same model can be configured to operate in conservative, safety-first, or balanced detection modes. As shown in Table III, the recall-boost and balanced prompt strategies achieve competitive recall and F1-score values without any fine-tuning, while offering explicit control over safety-related trade-offs. This indicates that prompt-controlled inference offers a useful, economical, and resource-efficient substitute for training-dependent state-of-the-art methods in early wildfire monitoring.

### B. Qualitative Visual Analysis

In addition to the quantitative assessment, a qualitative visual analysis is conducted to investigate the behaviour of various prompt strategies in challenging wildfire situations. Fig. 2 presents sample images from the FIgLib-Test dataset in which wildfire smoke is present in the ground truth but detected only by the Balanced prompt strategy.

These samples correspond to early-stage wildfire conditions, where smoke appears thin, distant, or partially blended with the background. While the Recall-Boost prompt may still reject detection when smoke cues are visually subtle or spatially sparse, the Baseline prompt fails to detect smoke due to its demand for strong visual evidence. The Balanced prompt, on the other hand, uses moderate semantic cues, such as faint haze, or plume-like textures over mountainous terrain, to successfully identify smoke.

This behaviour is consistent with the safety-oriented goal of minimising missed detections while avoiding excessive false alarms. These visual results demonstrate that inference-time prompt design enables effective control over detection behaviour without modifying model parameters or retraining, validating the proposed approach for practical wildfire monitoring applications.

### C. Safety-Oriented Trade-Off and Deployment Discussion

Wildfire smoke detection is by definition a safety-critical task where different types of classification errors have asymmetric consequences. False alarms (false positives) typically incur limited operational costs, such as additional human inspection or drone-based verification. In contrast, missing early smoke detection (false negatives) can delay intervention and allow wildfires to spread uncontrollably, resulting in severe


Fig. 2. Qualitative examples from the FIgLib-Test dataset where wildfire smoke is present in the ground truth and detected only by the Balanced and Recall prompt strategy.

environmental, economic, and human losses. Therefore, practical wildfire monitoring systems should prioritise minimising false negatives, even at the expense of increased false positives.

From a deployment viewpoint, the proposed method is particularly well suited for real-time wildfire monitoring on edge platforms with constrained computational resources. By using a compact vision–language model and modifying behaviour solely through prompt design, the system can adapt to changing safety requirements and environmental conditions at inference time. Overall, prompt-controlled inference provides a simple yet effective approach for safety-focused wildfire smoke detection in real-world scenarios.

## VI. Conclusion

This paper examined inference-time prompt design as a lightweight method for controlling safety-oriented trade-offs in wildfire smoke detection using a small vision–language model. By testing three prompt strategies—Baseline, Recall-Boost, and Balanced—on the FIgLib-Test dataset, we concluded that model behaviour can be effectively adjusted at inference time without fine-tuning or retraining. This reduces resource consumption and allows less computation.

The experimental results showed that conservative prompts reduce false alarms but increase missed smoke detections, while recall-oriented prompts prioritise early smoke detection at the cost of higher false positives. The Balanced prompt achieved a practical compromise, significantly reducing false negatives while maintaining a controlled false alarm rate. Qualitative visual analysis further confirmed that the proposed approach successfully detects subtle early-stage smoke that is often missed by conservative detection approaches.

## References

[1] I. F. of Red Cross and R. C. Societies, "Tunisia forest wildfires 2022: Final report," Sep. 2024, accessed: 2025-03. [Online]. Available: https://reliefweb.int/report/tunisia/tunisia-forest-wildfires-2022-final-report-dref-operation-ndeg-mdrtn010

[2] Nawaat Editorial Team, "Tunisian forests: Going up in flames," *Nawaat*, Aug. 2024, accessed: 2025-03. [Online]. Available: https://nawaat.org/2024/08/30/tunisian-forests-going-up-in-flames/

[3] Tunisie Numérique, "Tunisia: 550 hectares destroyed by wildfires, authorities say 95% caused by human activity," *Tunisie Numérique*, Jul. 2025, accessed: 2025-03. [Online]. Available: https://news-tunisia.tunisienumerique.com/tunisia-550-hectares-destroyed-by-wildfires-authorities-say-95-caused-by-human-activity/

[4] A. W. Ali and S. Kurnaz, "Optimizing deep learning models for fire detection, classification, and segmentation using satellite images," *Fire*, vol. 8, no. 2, p. 36, 2025.

[5] R. Ghali and M. A. Akhloufi, "Deep learning approaches for wildland fires using satellite remote sensing data: Detection, mapping, and prediction," *Fire*, vol. 6, no. 5, p. 192, 2023.

[6] A. Elhanashi, S. Essahraui, P. Dini, and S. Saponara, "Early fire and smoke detection using deep learning: A comprehensive review of models, datasets, and challenges," *Applied Sciences*, vol. 15, no. 18, p. 10255, 2025.

[7] A. Saleh, M. A. Zulkifley, H. H. Harun, F. Gaudreault, I. Davison, and M. Spraggon, "Forest fire surveillance systems: A review of deep learning methods," *Heliyon*, vol. 10, no. 1, 2024.

[8] V. E. Sathishkumar, J. Cho, M. Subramanian, and O. S. Naren, "Forest fire and smoke detection using deep learning-based learning without forgetting," *Fire ecology*, vol. 19, no. 1, p. 9, 2023.

[9] S. Wu and Y. Xia, "Enhanced yolov7-tiny for small-scale fire detection via multi-scale channel spatial attention and dynamic upsampling," *IEEE Access*, 2025.

[10] D. Spiller, A. Carbone, S. Amici, K. Thangavel, R. Sabatini, and G. Laneve, "Wildfire detection using convolutional neural networks and prisma hyperspectral imagery: A spatial-spectral analysis," *Remote Sensing*, vol. 15, no. 19, p. 4855, 2023.

[11] I. El-Madafri, M. Peña, and N. Olmedo-Torre, "Real-time forest fire detection with lightweight cnn using hierarchical multi-task knowledge distillation," *Fire*, vol. 7, no. 11, p. 392, 2024.

[12] P. Gao, S. Geng, R. Zhang, T. Ma, R. Fang, Y. Zhang, H. Li, and Y. Qiao, "Clip-adapter: Better vision-language models with feature adapters," *International Journal of Computer Vision*, vol. 132, no. 2, pp. 581–595, 2024.

[13] S. P. H. Boroujeni, N. Mehrabi, F. Afghah, C. P. McGrath, D. Bhatkar, M. A. Biradar, and A. Razi, "Toward ai-driven fire imagery: Attributes, challenges, comparisons, and the promise of vlms and llms," *Machine Learning with Applications*, p. 100763, 2025.

[14] S. Wu, Y. Qiao, S. He, J. Zhou, Z. Wang, X. Li, and F. Wang, "Fireclip: Enhancing forest fire detection with multimodal prompt tuning and vision-language understanding," *Fire*, vol. 8, no. 6, p. 237, 2025.

[15] E. H. Alkhammash, "Leveraging large language models for enhanced classification and analysis: Fire incidents case study," *Fire*, vol. 8, no. 1, p. 7, 2024.

[16] L. Catalan, A. Garacochea, A. Casi, M. Araiz, P. Aranguren, and D. Astrain, "Experimental evidence of the viability of thermoelectric generators to power volcanic monitoring stations," *Sensors*, vol. 20, no. 17, p. 4839, 2020.

[17] B. Kim, S. Kang, and S. Lee, "A weighted pagerank-based bug report summarization method using bug report relationships," *Applied Sciences*, vol. 9, no. 24, p. 5427, 2019.

[18] H. S. Yoon, E. Yoon, J. T. J. Tee, M. Hasegawa-Johnson, Y. Li, and C. D. Yoo, "C-tpt: Calibrated test-time prompt tuning for vision-language models via text feature dispersion," *arXiv preprint arXiv:2403.14119*, 2024.

[19] Y. Zhu, G. Zhang, C. Xu, H. Shen, X. Chen, G. Wu, and L. Wang, "Efficient test-time prompt tuning for vision-language models," *arXiv preprint arXiv:2408.05775*, 2024.

[20] M. Shu, W. Nie, D.-A. Huang, Z. Yu, T. Goldstein, A. Anandkumar, and C. Xiao, "Test-time prompt tuning for zero-shot generalization in vision-language models," *Advances in Neural Information Processing Systems*, vol. 35, pp. 14 274–14 289, 2022.

[21] A. Marafioti, O. Zohar, M. Farré, M. Noyan, E. Bakouch, P. Cuenca, C. Zakka, L. B. Allal, A. Lozhkov, N. Tazi *et al.*, "Smolvlm: Redefining small and efficient multimodal models," *arXiv preprint arXiv:2504.05299*, 2025.

[22] X. Sun, L. Sun, and Y. Huang, "Forest fire smoke recognition based on convolutional neural network," *Journal of Forestry Research*, vol. 32, no. 5, pp. 1921–1927, 2021.

[23] Z. Wu, R. Xue, and H. Li, "Real-time video fire detection via modified yolov5 network model," *Fire Technology*, vol. 58, no. 4, pp. 2377–2403, 2022.

[24] A. Dewangan, Y. Pande, H.-W. Braun, F. Vernon, I. Perez, I. Altintas, G. W. Cottrell, and M. H. Nguyen, "Figlib & smokeynet: Dataset and deep learning model for real-time wildland fire smoke detection," *Remote Sensing*, vol. 14, no. 4, p. 1007, 2022.

[25] L. Seidel, S. Gehringer, T. Raczok, S.-N. Ivens, B. Eckardt, and M. Maerz, "Advancing early wildfire detection: Integration of vision language models with unmanned aerial vehicle remote sensing for enhanced situational awareness," *Drones*, vol. 9, no. 5, p. 347, 2025.