Human Factors in Simulation, Software and Systems Engineering, Vol. 220, 2026, 60–66
https://doi.org/10.54941/ahfe1007694

From Manual Patrols to Automated Detection: Leveraging Aerial Imagery, Computer Vision, and Large Language Models for Wildfire Risk Mitigation

Haranath Varanasi[1] and Zining Yang[1,2]

[1]Southern California Edison, Rosemead, CA 91770, USA
[2]Cal Poly Pomona, Pomona, CA 91768, USA

ABSTRACT

Wildfire risk around electrical transmission and distribution infrastructure has grown significantly because of climate change, vegetation encroachment, prolonged drought cycles, and extreme weather events. Electric utilities are required to inspect their electric assets on regular basis. The team traditionally rely on ground patrols, helicopter inspections, and manual review of aerial photographs to evaluate asset condition and detect burn indicators. While effective, these methods are time-consuming, labor-consuming, costly, and limited by human capacity, making frequent monitoring impractical at scale. This paper presents an integrated framework that uses aerial imagery, convolutional neural networks (CNNs), computer vision (CV) segmentation, and multimodal large language models (LLMs) to automate the detection of charring, scorch marks, vegetation encroachment, and other wildfire risk factors. The approach reduces manual inspection burdens, increases monitoring frequency, saves cost, and enables proactive wildfire mitigation.

Keywords: Wildfire mitigation, Aerial imagery, Computer vision, Large language models

INTRODUCTION

Wildfires pose a growing threat to electric grid infrastructure and surrounding communities. Since the damaging fires in 2017 and 2018 in California, there has been increasing regulatory scrutiny, and electric utilities are required to conduct asset inspection on regular basis. Utilities traditionally conduct manual and helicopter based inspections, but these approaches are time-consuming, labor-consuming, and difficult to scale.

To address these challenges, this work demonstrates an automated aerial imagery based detection pipeline designed to identify charring, scorch marks, ash patterns, and other early wildfire precursors. Improving early detection improves public safety, reduces ignition risk, and speeds up mitigation workflows, while also improves operational efficiency. By combining computer vision models with multimodal large language models, this workflow will make inspections quicker and easier to manage, moving from fixing problems after they happen to preventing them before they start. Automated systems can improve: frequency and consistency of inspections,

Received February 7, 2026; Revised April 4, 2026; Accepted April 24, 2026; Available online July 20, 2026

speed of identifying high-risk conditions, and workforce and operational efficiency.

Figure 1: Wildfire example.

LITERATURE REVIEW

Recent advances in drone technology and computer vision have strengthened wildfire monitoring and burn severity assessment. Satellite and aerial imagery are commonly used to map burned areas, assess severity, and analyze post-fire recovery. We observed a shift from index-based methods towards machine learning and deep learning approaches that better capture spatial variability and complex burn patterns (Szpakowski and Jensen, 2019; Nelson et al., 2024). UAV imagery enables high-resolution detection of indicators such as charring, ash deposition, and vegetation stress, supporting fine-scale severity classification (Simes et al., 2024). Multispectral and 3D photogrammetric methods further improve severity estimation by capturing structural and spectral burn characteristics (Batchelor et al., 2025). Despite these advances, challenges remain in labeling effort, pattern variability, and cross-environment generalization.

Deep learning models such as convolutional neural networks (CNNs) have shown success in burn severity mapping and vegetation identification (Ban et al., 2020; Belenguer-Plomer et al., 2021; Zhu et al., 2017). More recently, multimodal large language models (LLMs) such as GPT-4V and Google Gemini extend image understanding to open-ended assessment and zero-shot classification (Radford et al., 2021; Bommasani et al., 2021; Zhang et al., 2024). These models can support synthetic labeling, dataset augmentation, and image triage when labeled wildfire imagery is scarce (Wu et al., 2023).

METHODOLOGY

The methodology contains three major components: CNN-based classification, segmentation-based detection, and dataset augmentation and triage using multimodal LLMs.

CNN-Based Image Classification

A baseline CNN model classified aerial images into "charring" vs. "no charring." Although accuracy reached approximately 80%, challenges included:

- Pattern specificity:
  The model detected black charring well but struggled with subtle indicators such as ash deposits, browning of wood, and partial scorch patterns.
- High dependency on training image diversity:
  Each new burn pattern required its own training set, reducing generalizability.
- Limited dataset availability:
  Curating thousands of labeled examples through manual search was highly labor-intensive and slow.

Segmentation-Based Approach

Manual pixel-level or region-level segmentation improved explainability but required labour-intensive labelling. The model struggled with generalizing across image conditions.

Table 1: Challenges in segmentation-based approach.

Challenge	Impact
High labeling cost	Maintenance/installation safety
Pattern variability	Maintenance/installation time to complete
Rare burn textures	Requires niche examples
Background noise	Increases false positives

LLM-Augmented Active Learning

To overcome limited labeled data and variable patterns,We developed a Well-architected human-in-the-loop (HITL) active learning pipeline. It essentially uses Large Language Models (LLMs) as high-level "reasoners" to discover new patterns, while Convolutional Neural Networks (CNNs) handle the heavy lifting for known patterns.

The primary challenge in detecting structural burn indicators is the morphological diversity of patterns such as surface scorching, deep charring, and ash deposition. To address this, we define a set of n specialized classifiers, M = {M1, M2, . . . , Mn}, where each Mi is optimized for a specific visual signature.

Heuristic-Based Ensemble-OR Triage

In the initial discovery phase, the priority is to minimize the Miss Rate (maximize recall). We utilize a ”union” logic where a positive signal is triggered if at least one specialized model exceeds its detection threshold τi. The final triage prediction PTriage is defined as:

$$P\_Triage(x)=\begin{cases}1 & \text{if } \mathrm{V}\_i=1^{n}(M\_i(x)\geq\tau\_i)\\ 0 & \text{otherwise}\end{cases}$$

This ensures that the ”long-tail” patterns—those often ignored by more generalized models are captured and forwarded to the LLM-Human validation loop.

Reliability-weighted Soft Fusion For production-level inference, we transition to a precision-focused aggregation. Instead of binary logic, we compute a weighted sum of the probability scores P (y|Mi, x). The final fused score Sf used is calculated as:

$$S_{fused}(x)=\sum_{i=1}^{n} w_i\cdot P(y\mid M_i,x)$$

Where the weights wi are assigned based on the historical precision of model Mi against the ground truth established by the LLM and human reviewers. These weights are constrained such that:

$$\sum_{i=1}^{n} w_i=1$$

This mechanism allows the system to suppress false positives caused by models that may be “over-sensitive” in certain environmental conditions (e.g., high glare or shadow regions).

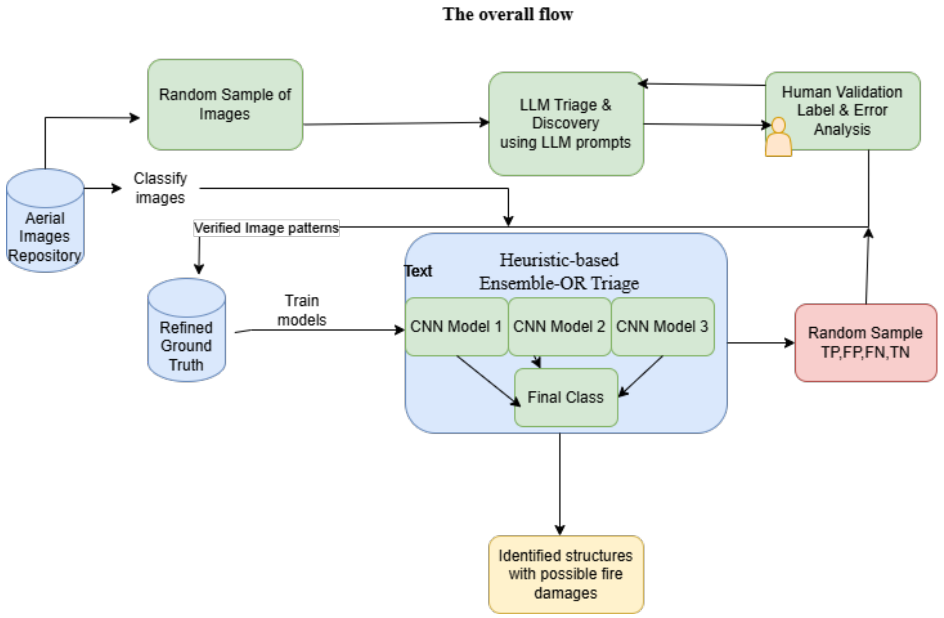

Figure 2: Image depicting overall flow.

The figure above shows the overall flow, including the following components.

1. Initial Discovery: Perform random sampling across diverse structures to identify novel visual patterns.
2. Zero-shot Triage: Utilize Multimodal LLMs to flag potential burn indicators, establishing a high-level candidate pool.
3. Ground Truth Establishment: Human validation of LLM outputs to create a "gold standard" training set.
4. High-Recall Inference: Deployment of Heuristic-based Ensemble-OR Triage to capture multi-morphological burn signatures.
5. False Negative Recovery: Cross-referencing CNN-ignored samples against high-confidence LLM detections to identify "unlearned" patterns.
6. Precision Refinement: Applying Reliability-weighted Soft Fusion to aggregate model scores and filter environmental noise.
7. Active Learning Loop: Continuous re-routing of recovered False Negatives back to the training pipeline for iterative model refinement.

Table 2: Comparative performance of detection strategies across burn patterns.

Burn Pattern	Ensemble-OR Triage		Soft Fusion Aggregation	
	Recall	FPR	Recall	Precision
Surface Scorching	0.96	0.15	0.88	0.92
Structural Charring	0.98	0.12	0.94	0.95
Ash Deposition	0.92	0.18	0.85	0.89
Overall Loop	0.95	0.15	0.89	0.92

DATA AND FEATURE ENGINEERING

The data includes high-resolution aerial imagery of poles, crossarms, insulators, and surrounding vegetation. Feature Classes include the following

- Color-based features: scorch coloration, blackening
- Texture features: ash grain, char patterns
- Environmental features: vegetation dryness
- Metadata: GPS, altitude, viewing angle

RESULTS

Evaluation of the CNN and LLM-assisted pipeline showed increased detection performance, better generalization, and reduced manual review load.

Table 3: Model result comparison.

Model	Accuracy	Recall	F1-Score
Baseline CNN	0.7	0.75	0.8
Segmentation CNN	0.7105	0.81	0.81
CNN + LLM-Triage	0.825	0.93	0.89

CONCLUSION

The hybrid aerial imagery and LLM-assisted pipeline provides a scalable method for identifying wildfire hazards near electrical infrastructure. This work demonstrates improved operational efficiency, enhanced detection capabilities, and a pathway to real-time automated wildfire risk mitigation.

By automating the identification of critical indicators specifically charring, scorch marks, and the accumulation of dry vegetation utilities can move beyond periodic snapshots toward a continuous risk-assessment model. The integration of CV segmentation and multimodal LLMs allows for the high-fidelity categorization of asset health, enabling teams to prioritize repairs based on actual thermal or biological threats rather than rigid, calendar-based schedules.

Furthermore, the implementation of this framework yields significant socio-economic benefits.

- **Operational Efficiency:** The system eliminates thousands of unnecessary field service hours and hazardous manual trips for workers, significantly reducing the carbon footprint and safety risks associated with traditional ground and helicopter patrols.
- **Financial Impact:** By streamlining the inspection lifecycle, the company realizes a substantial increase in the bottom line through reduced O&M (Operations and Maintenance) costs and the avoidance of catastrophic litigation and restoration expenses associated with utility-caused wildfires.
- **Future Readiness:** This automated pipeline provides a foundation for predictive maintenance planning, allowing utilities to forecast vegetation growth patterns and structural degradation before they escalate into emergencies.

Ultimately, this research bridges the gap between raw aerial data and actionable intelligence, offering a proactive shield for both the power grid and the communities it serves.

REFERENCES

Batchelor, J.L., Hudak, A.T., Bright, B.C., Meddens, A.J.H. and Klauberg, C. (2025). Drone-based multispectral photogrammetric point clouds to improve fire severity assessment. *Remote Sensing of Environment*.

Ban, Y., Zhang, P., Nascetti, A., Bevington, A. and Wulder, M. (2020). Near real-time wildfire detection using satellite SAR time series and deep learning. Scientific Reports, 10, 1320.

Belenguer-Plomer, M.A., Tanase, M.A., Chuvieco, E. and Bovolo, F. (2021). CNN-based burned area mapping using radar and optical data. Remote Sensing of Environment, 260, 112468.

Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E. et al. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

Nelson, D.M., He, Y. and Moore, G. (2024). Trends and applications in wildfire burned area mapping: Remote sensing data, cloud geoprocessing platforms, and emerging algorithms. *Remote Sensing of Environment*.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J. et al. (2021). Learning transferable visual models from natural language supervision. In: Proceedings of the 38th International Conference on Machine Learning (ICML), pp. 8748–8763.

Simes, T., Pádua, L. and Moutinho, A. (2024). Wildfire burnt area severity classification from UAV-based RGB and multispectral imagery. *Remote Sensing*, 16(1), 30.

Szpakowski, D.M. and Jensen, J.L.R. (2019). A review of the applications of remote sensing in fire ecology. *Fire Ecology*, 15(1).

Wu, C., Fevry, T., Wan, C., Jojic, N. and Zettlemoyer, L. (2023). Multimodal learning with large-scale vision-language models. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Zhang, Y., Chen, X., Wang, J. and Li, K. (2024). Evaluation of multimodal large language models for medical and general-domain image understanding. Artificial Intelligence in Medicine, 145, 102676.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F. and Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review. IEEE Geoscience and Remote Sensing Magazine, 5(4), pp. 8–36.