EXPLAINABLE UNSUPERVISED METHODS FOR FOREST FIRE DETECTION

Chandrabali Karmakar
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
Email: chandrabali.karmakar@dlr.de

Arnab Bhowmik
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
Email: arnab.bhowmik@dlr.de

Corneliu Octavian Dumitru
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
Email: corneliu.dumitru@dlr.de

Jakob Gawlikowski
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
Email: jakob.gawlikowski@dlr.de

Shivam Goyal
German Aerospace Center (DLR)
Deggendorf Institute of Technology
Deggendorf, Germany
Email: shivam.goyal@th-deg.de

Mihai Datcu
CEOSpaceTech
University POLITEHNICA of Bucharest (UPB)
mihai.datcu@upb.ro

***Abstract*—Wildfires are among the most damaging natural hazards, and reliable satellite-based detection is essential for operational response. Many current pipelines for multispectral Sentinel-2 imagery are supervised and data-hungry, and they often behave as black boxes, limiting trust and expert auditability during fast-evolving events. We present a fully unsupervised and intrinsically explainable framework for wildfire detection that combines explainable K-Means (xKMeans), explainable Gaussian Mixture Models (xGMM), and explainable Latent Dirichlet Allocation (LDA). Each pixel is represented by a compact fire-sensitive feature vector built from Sentinel-2 bands B08, B8A, and B12 together with the Normalized Burn Ratio (NBR) and the Mid-Infrared Burn Index (BAIS1). The three models expose complementary structure: xKMeans yields human-readable, rule-based partitions; xGMM provides probabilistic memberships and entropy-based uncertainty maps that highlight ambiguous transition zones; and LDA offers global, topic-level explanations by summarizing co-occurring spectral-index patterns into interpretable fire regimes. We demonstrate the approach on multiple wildfire scenes, including case studies in Los Angeles (USA) [1], Greece [2], and an additional site in Romania based on GDACS-reported event metadata. We report qualitative and label-light quantitative consistency analyses based on model agreement and perimeter-aligned evaluation. The results indicate that the proposed workflow delineates fire-affected areas while remaining transparent enough for expert scrutiny, offering a label-efficient alternative to supervised deep-learning systems for operational wildfire monitoring.**

***Index Terms*—Wildfire detection, explainable artificial intelligence, unsupervised learning, Sentinel-2, clustering, Gaussian Mixture Models, K-Means, Latent Dirichlet Allocation, uncertainty.**

I. Introduction

Wildfires continue to intensify in frequency and severity, threatening ecosystems, infrastructure, and public safety. Sentinel-2 has become central to wildfire monitoring due to its spatial resolution, radiometric quality, and sensitivity to burn severity [3, 4]. However, many current systems depend on supervised deep-learning models [5–7], whose performance hinges on extensive labelled datasets that are rarely available during active fire emergencies.

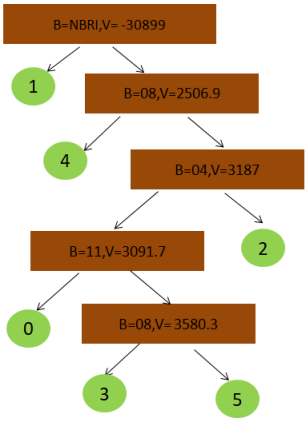

Fig. 1. xKMeans decision tree illustrating an interpretable rule set for cluster assignment [8].

A second challenge is interpretability. Deep networks typically operate as black boxes, limiting trust and preventing domain experts from understanding the model rationale. Prior explainability studies [9, 10] show that post-hoc explanations often provide only partial insight and may not faithfully reflect decision boundaries under complex fire conditions.

Unsupervised learning is a natural alternative when labels are scarce. Contrastive learning [11] and autoencoder-based anomaly detection [12] can capture latent spectral patterns without manual annotation, but many unsupervised methods still lack intrinsic interpretability: cluster semantics, uncertainty, and spectral drivers remain opaque.

We address these gaps with an explainable unsupervised framework integrating xKMeans, xGMM, and LDA. The system is designed to provide transparent internal structure at three levels: rule-based partitions for crisp interpretation, probabilistic memberships with uncertainty for operational risk awareness, and topic-feature relationships for global, human-

readable explanations of spectral-index regimes.

II. State of the Art vs. Our Method

Supervised deep-learning models dominate wildfire detection and burned-area mapping [5–7]. While accurate, they require large labelled datasets and can be difficult to interpret in high-stakes settings. Explainable AI has been explored for susceptibility and occurrence modelling [9, 10], but much of this work relies on post-hoc explanations.

Unsupervised methods provide an attractive alternative when labels are scarce. Contrastive representation learning (FireCLR) [11] enables cross-sensor generalisation, and autoencoders have been used to identify spectral anomalies linked to fire activity [12]. However, many unsupervised approaches still lack intrinsic interpretability: cluster semantics, uncertainty, and decision boundaries remain opaque, limiting operational trust.

Our approach differs by using intrinsically explainable unsupervised models. Instead of explaining a black-box classifier after training, we structure the pipeline so that explanations are native outputs: xKMeans provides explicit decision rules, xGMM yields uncertainty-aware probability maps, and LDA gives a compact topic view that links fire f to combinations of bands and indices.

III. Study Area and Data

A. Analysed Fire Events

We evaluate the framework on multiple wildfires with contrasting conditions: (i) a fire near Los Angeles (USA) [1] dominated by shrubland and chaparral fuels, (ii) a forest fire in Greece [2] affecting dense pine and mixed broadleaf stands, and (iii) an event in Romania representing a different forest structure and background heterogeneity.

For Romania, we use the GDACS-reported wildfire WF 1005046 (episode 1), with start date 20 Mar 2022 and last detection 28 Mar 2022, an estimated burned area of 5039 ha, and an exposed population of 154 in the burned area [13]. GDACS forest-fire events and burned-area estimates are generated through the Global Wildfire Information System, based on near-real-time products driven by thermal anomalies. We use the GDACS event metadata to define the temporal window and region of interest for selecting Sentinel-2 Level-2A acquisitions and for perimeter-aligned, label-light evaluation.

For each event, we use cloud-minimised Sentinel-2 Level-2A scenes acquired close to the burning period and clipped to a study tile around the affected region. For every pixel in the study tiles we extract the compact fire-sensitive feature set defined in Section IV-A. This representation captures vegetation loss (NIR and narrow-NIR), moisture and burn response (SWIR), and established burn-sensitive indices.

B. Additional Sites for Extended Experiments

Within the wider project we also prepare Sentinel-2 data sets for further regions from recent fire seasons, including (i) the Brazil–Bolivia–Paraguay tri-border in the Amazon, (ii) fires near Sydney (Australia), and (iii) a large event in Andalusia (Spain). Due to the page limits of this IGARSS paper, these additional sites are consolidated in an extended version together with full quantitative analyses.

Band number		Central wavelength
Pixel spacing 10 m	2 - Blue	492.4 nm
	3 - Green	559.8 nm
	4 - Red	664.6 nm
	8 - Near Infrared (NIR)	832.8 nm
Pixel spacing 20 m	5 - Vegetation red edge	704.1 nm
	6 - Vegetation red edge	740.5 nm
	7 - Vegetation red edge	782.8 nm
	8A - Narrow NIR	864.7 nm
	11 - Shortwave IR (SWIR)	1613.7 nm
	12 - SWIR	2202.4 nm
Pixel spacing 60 m	1 - Coastal aerosol	442.7 nm
	9 - Water vapour	945.1 nm
	10 - SWIR - Cirrus	1373.5 nm

Fig. 2. Spectral bands of Sentinel-2 data provided by ESA.

Sentinel-2 provides 13 spectral bands in the visible, near-infrared, and shortwave-infrared ranges at 10 m, 20 m, and 60 m resolution. Figure 2 summarises the band layout.

IV. Methods

A. Feature Engineering

Each pixel is represented by

$$\mathbf{f}(x) = [B08(x),\, B8A(x),\, B12(x),\, \mathrm{NBR}(x),\, \mathrm{BAIS1}(x)].$$

Sentinel-2 L2A band values are stored as scaled integers; before computing NBR and BAIS1 we rescale B08, B8A, and B12 by dividing by 10000 to obtain reflectance in the range [0,1].

$$\mathrm{NBR}(x) = \frac{B08(x) - B12(x)}{B08(x) + B12(x)} \tag{1}$$

$$\mathrm{BAIS1}(x) = \sqrt{(1 - B08(x))\,(1 - B12(x))\,(B12(x) - B08(x))}. \tag{2}$$

All features are z-normalised per scene to support stable clustering across different backgrounds.

B. xKMeans: Interpretable Partitions

xKMeans provides a clear partition of the feature space into k clusters. The goal is a cluster map that can be explained through the feature signature: for example, a fire-dominated cluster is expected to show reduced NIR response and a burn-ratio shift relative to surrounding vegetation.

TABLE I
RECENT WILDFIRE DETECTION AND BURNED-AREA MAPPING METHODS

Reference	Data	Method	Contribution
Hu et al. (2021) [3]	Sentinel-2	Multi-criteria detection	Spectral thresholding for active fire identification across biomes.
Zhang et al. (2021) [5]	Sentinel-2	CNN	Deep-learning framework for automated active fire detection.
Roteta et al. (2021) [4]	Landsat/S2	GEE workflow	Operational burned-area mapping in cloud platforms.
Sui et al. (2024) [6]	Sentinel-2	Attention U-Net	Bi-temporal burned-area classification with attention.
Tiengo et al. (2025) [7]	Sentinel-2	Rao's Q metric	Burned-area discrimination using spectral diversity.
Abdollahi & Pradhan (2023) [9]	Multi-source	XAI	Explainability for susceptibility prediction and transparency.
Cilli et al. (2022) [10]	Europe-wide	XAI	Interpretable wildfire occurrence modeling in Mediterranean regions.
Hamilton et al. (2022) [11]	Sentinel-2, Planet	Contrastive learning	Unsupervised feature learning for burned-area mapping.
Üstek et al. (2024) [12]	MODIS/S2	Autoencoder	Deep unsupervised anomaly detection for fire-related spectral changes.

C. xGMM: Soft Memberships and Uncertainty

xGMM models the feature distribution as a Gaussian mixture:

$$p(x) = \sum_{j=1}^{k} \pi_j \mathcal{N}(x|\mu_j, \Sigma_j). \tag{3}$$

Each pixel receives responsibilities $\gamma_j(x)$, which are used as soft memberships. We summarize uncertainty with entropy:

$$H(x) = -\sum_{j=1}^{k} \gamma_j(x) \log \gamma_j(x). \tag{4}$$

This gives a practical uncertainty layer: high entropy highlights mixed pixels and transition zones where the model distributes probability across multiple clusters.

D. LDA: Topic-Level Explanations

LDA summarizes co-occurring band-index patterns into topics that explain what the clusters represent at scene level. Continuous features are discretised into bins; each pixel becomes a short set of discretised terms. LDA learns topic-term distributions and per-pixel topic mixtures. In practice, this allows us to connect a fire-dominated cluster with a consistent combination of high-SWIR / low-NIR / shifted burn-ratio bins, and to separate it from intact vegetation topics.

E. Complementarity of xGMM and LDA

xGMM and LDA provide two different but compatible explanation layers:

- xGMM answers: *How strongly does this pixel belong to the fire-like cluster, and where is the model uncertain?* The output is a probabilistic membership and an uncertainty map.
- LDA answers: *Which spectral-index patterns appear together in this scene, and which pattern dominates a region?* The output is a topic-level description that helps interpret clusters with human-readable feature combinations.

Together, xGMM provides local confidence and boundary behaviour, while LDA provides global meaning for the cluster signatures.

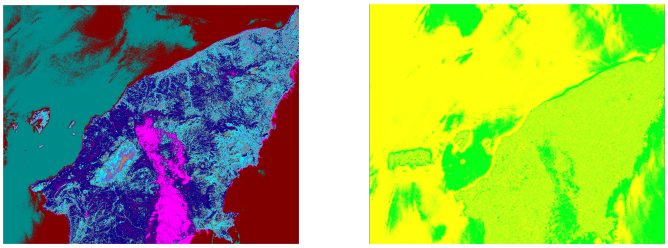

Fig. 3. Representative spatial outputs over a Greece test area. Left: xKMeans-derived cluster map highlighting the fire-dominated cluster. Right: xGMM probability map showing confidence gradients around burned regions.

V. RESULTS

The framework was evaluated on Sentinel-2 scenes from Los Angeles (USA), Greece, and Romania. All models were applied to the same fire-sensitive feature representation defined in Section IV-A, and the outputs were compared in terms of spatial coherence, spectral separability, and explainability.

A. Spatial Outputs and Uncertainty

Figure 3 shows representative outputs for a Greece test area. The xKMeans-derived cluster map isolates a spatially coherent fire-affected region. The xGMM probability map highlights high confidence in the core burned area and decreasing confidence toward the perimeter, reflecting mixed pixels and gradual transitions. The same interpretability products are generated for all sites, including Romania, enabling perimeter inspection via uncertainty gradients and rule-based cluster interpretation.

B. Cluster Statistics from xGMM

Figure 4 summarises the spectral structure learned by xGMM. In burned regions, the fire-dominated cluster is characterised by reduced NIR (B08), increased SWIR2 (B12), and a pronounced shift toward negative NBR, consistent with established burn interpretation. Cluster-wise statistics also support cross-site comparability by providing a compact, auditable signature of the fire-like regime.

We report the spectral signature of the fire-dominated cluster produced by our pipeline. Figure 5 shows density comparisons between the fire cluster (Cluster 5) and the remaining clusters (aggregated) for B08 (NIR), B12 (SWIR2), and NBRI (burn

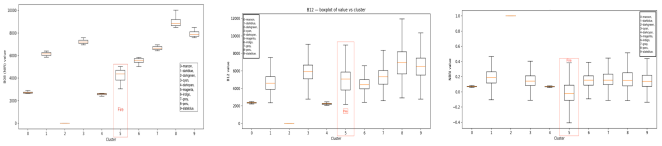

Fig. 4. xGMM cluster-wise boxplots for B08, B12, and NBR. The fire-dominated cluster exhibits low NIR, high SWIR2, and strongly negative NBR, consistent with burned surfaces.

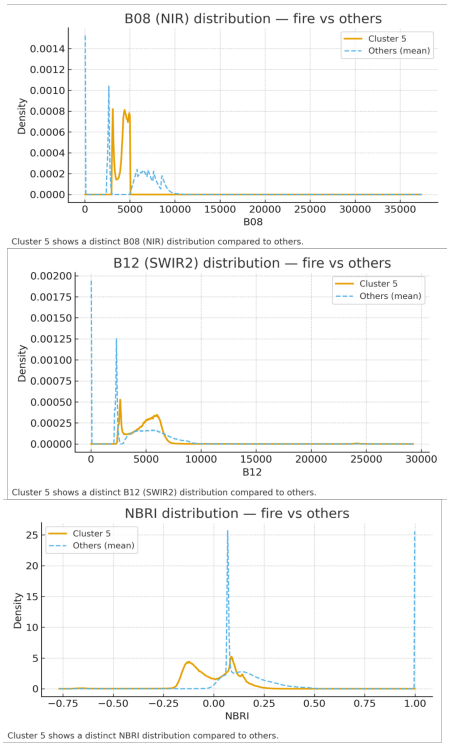

Fig. 5. Density comparisons from our results for the fire-dominated cluster (Cluster 5) versus the remaining clusters. Left: B08 (NIR). Middle: B12 (SWIR2). Right: NBRI. The fire cluster shows a distinct signature across the three variables, supporting an interpretable separation between fire-affected and non-fire regions.

ratio). These distributions come directly from our scene-level results.

The B08 plot shows that the fire cluster occupies a different NIR range compared to the other clusters. This matches the expected loss of healthy vegetation response in burned areas. The B12 plot shows a shifted SWIR2 distribution for the fire cluster relative to the rest of the scene, consistent with burn-related surface change. The NBRI plot shows a clear distribution shift between the fire cluster and other clusters, reflecting burn-ratio sensitivity to fire-affected pixels. Taken together, these three plots provide a compact, interpretable explanation of what the model captured: the detected fire cluster is not defined by a single feature alone, but by a joint signature across NIR, SWIR2, and burn ratio.

C. LDA Topics as Global Explanations

Figure 6 summarizes the type of global explanation provided by LDA. Across scenes, LDA yields a small number of

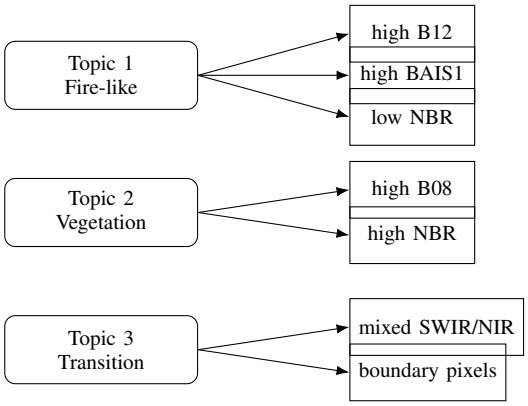

Fig. 6. LDA topic-level explanation. Fire-like topics are associated with high B12 and BAIS1 and low NBR, vegetation topics with high B08 and high NBR, and transition topics with mixed spectral behaviour near boundaries.

stable topics that align with interpretable fire regimes. Fire-like topics are characterised by discretised terms corresponding to high B12 and BAIS1 combined with low NBR, while vegetation topics emphasise high B08 and high NBR. Transition topics capture mixed spectral behaviour around cluster boundaries and partially affected pixels. By comparing topic prevalence inside the fire-dominated cluster versus outside, we obtain a scene-level explanation consistent with xKMeans rules and xGMM probability structure, providing a practical interpretability check without requiring dense labels.

VI. Conclusion and Future Work

We presented an intrinsically explainable, fully unsupervised framework for wildfire detection in Sentinel-2 imagery, combining xKMeans, xGMM, and LDA on a compact set of fire-sensitive spectral features. The models provide complementary explainability: xKMeans yields human-readable partitions, xGMM supplies uncertainty-aware probability maps for boundary inspection, and LDA delivers topic-level semantic summaries that support global auditing of fire regimes. We include multiple sites, including an additional Romania event anchored by GDACS metadata, to broaden the evaluation and demonstrate the portability of the explainable outputs.

Future work will complete the full set of experiments across all prepared locations, incorporate multi-temporal Sentinel-2 sequences and additional sensors, and report comprehensive quantitative evaluation against operational reference products and robustness analyses in an extended paper.

Acknowledgment

Part of this work at POLITEHNICA Bucharest was funded by the European Union (NextGenerationEU instrument) through the National Recovery and Resilience Plan, “PNRR-III-C9-2022 – I5 Establishment and operationalization of Competence Centers” competition, “Competence Center for Climate Change Digital Twin for Earth forecasts and societal redressment: DTEClimate” project, contract no. 760008/30.12.2022, code 7/16.11.2022.

REFERENCES

[1] European Space Agency (ESA), "Los Angeles ablaze," ESA Multimedia (Image), Jan. 2025, published: 08 Jan. 2025. Image ID: 505318. Accessed: 09 Jan. 2026. [Online]. Available: https://www.esa.int/ESA_Multimedia/Images/2025/01/Los_Angeles_ablaze

[2] ——, "Rhodes wildfire forces thousands to flee," ESA Multimedia (Image), Jul. 2023, published: 24 Jul. 2023. Image ID: 481171. Accessed: 09 Jan. 2026. [Online]. Available: https://www.esa.int/ESA_Multimedia/Images/2023/07/Rhodes_wildfire_forces_thousands_to_flee

[3] X. Hu, Y. Ban, and A. Nascetti, "Sentinel-2 msi data for active fire detection in major fire-prone biomes: A multi-criteria approach," *International Journal of Applied Earth Observation and Geoinformation*, vol. 101, p. 102347, 2021.

[4] E. e. a. Roteta, "Landsat and sentinel-2 based burned area mapping tools in google earth engine," *Remote Sensing*, vol. 13, no. 4, p. 816, 2021.

[5] Q. e. a. Zhang, "Towards a deep-learning-based framework of sentinel-2 imagery for automated active fire detection," *Remote Sensing*, vol. 13, no. 23, p. 4790, 2021.

[6] T. e. a. Sui, "Biau-net: Wildfire burnt area mapping using bi-temporal sentinel-2 imagery and attention u-net," *International Journal of Applied Earth Observation and Geoinformation*, vol. 132, p. 104034, 2024.

[7] J. e. a. Tiengo, "Burned areas mapping using sentinel-2 and rao's q index: Application to mediterranean islands," *Remote Sensing*, vol. 17, no. 4, p. 737, 2025.

[8] M. Moshkovitz, S. Dasgupta, C. Rashtchian, and N. Frost, "Explainable k-means and k-medians clustering," in *International conference on machine learning*. PMLR, 2020, pp. 7055–7065.

[9] A. Abdollahi and B. Pradhan, "Explainable artificial intelligence (xai) for interpreting the contributing factors feed into the wildfire susceptibility prediction model," *Science of The Total Environment*, vol. 879, p. 163004, 2023.

[10] P. e. a. Cilli, "Explainable artificial intelligence (xai) detects wildfire occurrence in the mediterranean countries of southern europe," *Scientific Reports*, vol. 12, no. 16781, 2022.

[11] D. e. a. Hamilton, "Fireclr: Unsupervised contrastive representation learning for burned area mapping," in *NeurIPS Workshop on Tackling Climate Change with Machine Learning*, 2022. [Online]. Available: https://arxiv.org/abs/2211.14654

[12] I. e. a. Üstek, "Deep autoencoders for unsupervised anomaly detection in wildfire prediction," *arXiv preprint*, 2024. [Online]. Available: https://arxiv.org/abs/2411.09844

[13] GDACS, "Forest fire (5039 ha) in romania, 20 mar 2022 (gdacs id wf 1005046, episode 1)," Global Disaster Alert and Coordination System, 2022, accessed: 2025-12-22. [Online]. Available: https://www.gdacs.org/Wildfires/report.aspx?eventtype=WF&eventid=1005046&episodeid=1