Beyond Event Prediction: A Loss Function for Operational Ordinal Wildfire Risk

Nicolas Caron
Université Marie et Louis Pasteur, CNRS, institut FEMTO-ST (UMR 6174)
Besançon, France
nicolas.caron@umlp.fr

Christophe Guyeux
Université Marie et Louis Pasteur, CNRS, institut FEMTO-ST (UMR 6174)
Belfort, France
christophe.guyeux@umlp.fr

Hassan Noura
Université Marie et Louis Pasteur, CNRS, institut FEMTO-ST (UMR 6174)
Belfort, France
hassan.noura@umlp.fr

Benjamin Aynes
SAD Marketing
Lille, France
b.aynes@sad-marketing.com

Abstract

Wildfire prediction is often framed as classification or regression, yet operational management relies on coherent ordinal risk levels. We propose the Cluster-Centered Latent Transition Loss (CCLLT), a novel ordinal objective designed for operational wildfire risk. Rather than predicting event occurrence directly, the proposed framework learns ordinal operational risk systems whose levels correspond to increasing real-world operational consequences. CCLLT combines latent ordinal centers, hierarchical EMA priors, monotonic ranking constraints, coverage regularization, and adaptive threshold calibration to directly optimize the ordinal consistency of risk signals. Unlike standard classification, it enforces monotonic operational transitions while preventing collapse toward dominant low-risk classes. Evaluated via a monotonic explainability framework, our approach successfully moves prediction beyond event-centered metrics to produce stable, interpretable, and operationally meaningful risk signals.

CCS Concepts

• **Theory of computation** → *Structured prediction*; • **Applied computing**; • **Networks**;

Keywords

ordinal regression, learning to rank, monotonic consistency, loss function design, threshold calibration, wildfire risk assessment, operational decision support

ACM Reference Format:
Nicolas Caron, Christophe Guyeux, Hassan Noura, and Benjamin Aynes. 2026. Beyond Event Prediction: A Loss Function for Operational Ordinal Wildfire Risk. In *Proceedings of the 35th ACM International Conference on Information and Knowledge Management (CIKM '26), November 7–11, 2026, Rome, Italy.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3799682.3839873

ACM ISBN 979-8-4007-2539-5/2026/11
https://doi.org/10.1145/3799682.3839873

1 Introduction

Wildfires represent a major environmental, economic, and human challenge whose frequency and intensity are increasing under climate change. To support emergency response strategies, we propose the Cluster-Centered Latent Transition Loss (CCLLT), a novel ordinal learning objective. Rather than optimizing event prediction directly, CCLLT learns ordinal operational risk signals that ensure higher predicted risk levels correspond to increasingly severe expected operational consequences. This fundamental distinction between "event prediction" (detecting an occurrence) and "risk signal learning" (estimating potential severity) is central to operational decision-making.

It is important to emphasize that our contribution is not a new wildfire forecasting architecture, but rather a new architecture-agnostic ordinal learning objective that can be used with GRU, LSTM, MLP, and DilatedCNN models. By exploiting large-scale meteorological, environmental, and operational data, the proposed framework goes beyond simple classification to target operational decision support, providing interpretable risk signals capable of helping firefighting services better prepare for extreme situations. The central hypothesis of this work is that operational risk is not a classification problem. It is the construction of an ordinal system whose levels must correspond to increasingly severe operational consequences. The objective is not to determine whether a wildfire will occur, but to organize future situations into a coherent operational severity system. Thus, risk learning should be treated as a system-construction problem rather than an event-prediction problem. The implementation of CCLLT is publicly available.[1]

2 Related Work

Most systems formulate wildfire forecasting as binary classification or continuous regression [4]. Methods range from early statistical models [6, 8, 10] to recent deep learning architectures (CNNs, RNNs) and foundation models [5, 11, 12]. However, these approaches primarily focus on event occurrence (mostly binary classification tasks).

[1]https://github.com/NicolasCaronPro/Learning-Operational-Ordinal-Wildfire-Risk-Signals-with-Monotonic-Ranking-Constraints

Beyond standard classification, Learning-to-Rank and ordinal learning paradigms, such as RankNet [1], optimize the relative ordering of samples. While relevant, pairwise sample ranking does not inherently structure the macroscopic transitions between predefined operational risk classes.

The evaluation of wildfire prediction models is often overlooked. However, recent work by Xu et al. [11] demonstrated that evaluation protocols can have a greater impact on reported performance than the model architecture itself. In operational settings, emergency services continue to rely heavily on expert-based indices such as the Canadian Fire Weather Index (FWI) because of their interpretability and proven robustness [7, 9]. Unlike conventional predictive models that focus on estimating event probability, these systems are designed to assess the potential severity of an event should it occur. To better align evaluation with this operational perspective, [2] introduced a framework for assessing ordinal risk prediction systems based on their monotonic operational consistency.the objective of this work is not event detection itself, but the learning of an interpretable ordinal operational prediction system define in 2.1.

Definition 2.1 (Operational Risk). Let $R_{z,t} \in \{0, \ldots, K\}$ denote the operational risk level and $Y_{z,t} \in \mathbb{R}_+$ the real operational variable.

Operational risk is defined as a bounded latent ordinal variable satisfying a monotonic consistency property:

$$\mathbb{E}[Y_{z,t} \mid R_{z,t} = a] < \mathbb{E}[Y_{z,t} \mid R_{z,t} = b] \quad \forall a < b.$$

3 Contribution

We introduce the Cluster-Centered Latent Transition Loss (CCLLT), a novel ordinal learning objective designed for operational wildfire risk inspired by RankNet [1]. Rather than predicting highly stochastic wildfire occurrences, CCLLT learns an ordinal risk system whose levels correspond to increasing operational consequences. The proposed loss combines latent ordinal centers, hierarchical EMA priors, monotonic transition constraints, and coverage regularization to directly optimize ordinal consistency while preserving predictive coverage. Unlike RankNet, which focuses on pairwise ordering between individual samples, CCLLT structures the latent space at the class level and can be integrated into a wide range of architectures, including MLPs, recurrent networks, and temporal convolutional models. To explicitly highlight the novelty of CCLLT compared to the standard RankNet formulation, Table 1 provides a direct comparison. While RankNet operates purely at the individual sample level, CCLLT structures the latent space at the ordinal class level through latent centers, hierarchical EMA priors, and coverage regularization.

Table 1: Methodological differences between RankNet and CCLLT.

Feature	RankNet	CCLLT (Proposed)
Operational Level	Individual Samples	Ordinal Classes
Latent Structure	Pairwise Score Diff.	Latent Ordinal Centers
Regularization	None	Coverage & Midpoint
Temporal Stabilization	None	EMA Centers

4 Experimental Setup

Experiments are conducted on wildfire intervention data collected from three French departments (Ain, Doubs and Alpes-Maritimes). Data from 2017–2020 and 2022 are used for training, 2021 for validation and 2023 for testing. The final benchmark contains 157 explanatory variables from Caron et al. [3] database. It's include meteorological, environmental, historical, topographical, and socioeconomic variables. The data spans 2,089 unique dates, including 7,006 positive wildfire interventions.

The objective is to learn an operational risk signal composed of five ordinal levels ranging from 0 (Very Low) to 4 (Extreme). For each department, day, and operational target, the model predicts a single risk level. This five-level scale was selected because it is commonly used in French wildfire risk assessment systems and operational decision-making frameworks.

For each department and each day, three operational targets are considered independently: the number of wildfire incidents, the intervention duration, and the resources deployed by emergency services.

The goal is therefore not to predict the exact value of these operational variables, but rather to learn an ordinal risk level whose ordering remains consistent with the observed operational consequences.

The proposed loss is evaluated using several architectures commonly employed for wildfire prediction, including GRU and LSTM temporal models, a Multi-Layer Perceptron (MLP), and a Dilated Convolutional Neural Network (DilatedCNN). These architectures allow the evaluation of the proposed ordinal framework across recurrent, fully connected, and temporal convolutional paradigms. Hyperparameters are optimized using Optuna while model selection relies on the monotonic explainability framework introduced in [2].

Let $s_i \in \mathbb{R}$ denote the latent risk score predicted for sample i, and let $p_{i,k}$ denote the soft probability that sample i belongs to ordinal class k.

Figure 1 provides an overview of the proposed Cluster-Centered Latent Transition Loss (CCLLT). The framework first estimates soft ordinal centers from model predictions, stabilizes them through hierarchical EMA priors, and finally enforces monotonic transitions between ordinal risk levels through a ranking objective.

4.1 Hierarchical Soft Ordinal Centers

Latent ordinal centers are estimated dynamically using soft memberships. For a grouping structure g, the latent center associated with ordinal class k is defined as:

$$\hat{\mu}_k^{(g)} = \frac{\sum_{i \in \mathcal{G}^{(g)}} p_{i,k} y_i}{\sum_{i \in \mathcal{G}^{(g)}} p_{i,k} + \varepsilon},$$

where $\mathcal{G}^{(g)}$ denotes the set of samples belonging to grouping scale g, $p_{i,k}$ is the soft probability that sample i belongs to ordinal class k, and y_i denotes the real operational target.

To stabilize ordinal structures across optimization steps, we introduce a hierarchical shrinkage mechanism over several grouping scales:

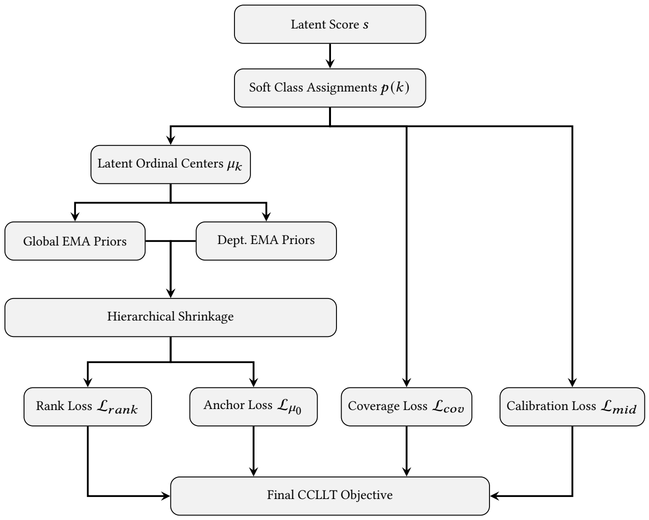

Figure 1: Overview of the proposed CCLLT objective. Latent ordinal centers are stabilized through hierarchical EMA priors and optimized using ranking, anchor, coverage, and calibration losses.

$$\mu_k^{(g)} = \frac{m_k^{(g)}\,\hat{\mu}_k^{(g)} + \sum_{\ell=1}^{g-1} \lambda_\ell^{(g)}\,\mu_k^{(\ell)}}{m_k^{(g)} + \sum_{\ell=1}^{g-1} \lambda_\ell^{(g)}},$$

where $m_k^{(g)}$ denotes the effective mass of class k at scale g, and $\lambda_\ell^{(g)}$ controls the shrinkage toward coarser ordinal systems.

Consequently, ordinal consistency is enforced independently inside each grouping structure rather than globally across all samples. In the wildfire context, this distinction is essential: the objective is not to compare a historically high-risk region against a low-risk one, but instead to verify that the ordinal risk structure remains coherent inside each region individually.

Finally, the hierarchical shrinkage toward larger-scale ordinal systems allows sparse or underrepresented groups to remain statistically stable. Groups containing very few samples can therefore still contribute to training while borrowing robustness from higher aggregation levels. This mechanism prevents unstable ordinal centers while preserving local operational specificity.

In this work, two hierarchical resolutions are considered. The first corresponds to the *global* level, where all samples are used to define a shared ordinal system across the entire dataset. The second corresponds to the *departmental* level, where an independent ordinal system is learned for each French department. This hierarchical formulation allows the model to capture both large-scale ordinal trends shared across all regions and local operational characteristics specific to each department.

4.2 EMA Ordinal Priors

The latent centers defined in the previous section are estimated from mini-batches and may therefore suffer from high variance, especially when some ordinal classes are rare or completely absent from the current batch. Relying exclusively on batch statistics would make the learned ordinal structure unstable and highly dependent on local sampling fluctuations.

To address this limitation, we maintain Exponential Moving Average (EMA) estimates of the latent centers at both global and departmental scales. For each ordinal class k, the corresponding prior is updated according to

$$\mu_k^{EMA} \leftarrow \rho\,\mu_k^{EMA} + (1-\rho)\,\hat{\mu}_k,$$

where ρ denotes the EMA momentum and $\hat{\mu}_k$ is the current batch estimate.

These priors serve three complementary objectives. First, they allow the ordinal system to emerge progressively during training instead of being determined solely by the current mini-batch. Second, they provide meaningful estimates for classes that are not present in the current batch. Finally, they produce a stable representation of the relative spacing between ordinal levels at the end of training, which can subsequently be used to evaluate monotonic transitions.

To avoid propagating highly noisy estimates during the initial optimization phase, EMA updates are activated only after the first two training epochs. The resulting global and departmental priors are then incorporated into the hierarchical shrinkage mechanism.

4.3 Monotonic Transition Ranking

For each transition pair (a, b) with $a < b$, we define:

$$\Delta_{a\to b} = \mu_b - \mu_a.$$

The transition ranking loss is defined as:

$$\mathcal{L}_{rank} = \sum_{k=1}^{C-1} w_k\,\mathrm{BCEWithLogits}(\Delta_{a\to b} - m_{a,b}, 1).$$

This formulation directly penalizes inverted ordinal transitions while preserving smooth optimization dynamics.

4.4 Coverage Regularization

To prevent collapse toward dominant low-risk classes, a cumulative distribution regularization term is introduced:

$$\mathcal{L}_{coverage} = \left\| F(\hat{p}_g) - F(q_g) \right\|_2^2.$$

This formulation directly constrains the ordinal structure of predicted risk distributions and preserves higher-risk operational states during training.

4.5 Low-Risk Anchor Regularization

Finally, an additional regularization term is introduced to anchor the lowest ordinal level and stabilize the lower bound of the operational risk scale. In practice, the latent center associated with class 0 is encouraged to remain as small as possible through the following penalty:

$$\mathcal{L}_{\mu_0} = \mathrm{Softplus}(\mu_0),$$

where μ_0 denotes the averaged latent center of the first ordinal class over the active clusters.

This regularization prevents upward drift of the lowest-risk representation during training and ensures that the *Very Low* risk level remains associated with the weakest operational intensity. Consequently, the ordinal structure becomes better anchored globally, while the transition ranking loss continues to enforce monotonic increases between higher-risk classes.

4.6 Final Objective

The final loss combines ranking, coverage regularization, low-risk calibration, and threshold alignment:

$$\mathcal{L} = \lambda_{trans}\mathcal{L}_{rank} + \lambda_{coverage}\mathcal{L}_{coverage} + \lambda_{\mu_0}\mathcal{L}_{\mu_0} + \lambda_{mid}\mathcal{L}_{mid}.$$

5 Evaluation as Monotonic Explainability

Because the objective is to learn an operational risk system rather than maximize event-detection accuracy, evaluating CCLLT requires metrics capable of assessing the consistency of the entire ordinal structure. The proposed framework is evaluated using the monotonic explainability strategy introduced in [2]. Traditional event-detection metrics (like precision, recall, or F1-score) are insufficient for operational risk systems. Similar recall values may correspond to very different underlying risk structures; a model might correctly detect events but randomly assign severity levels, which is operationally unusable. Monotonic consistency is therefore a necessary property. The objective is not purely predictive performance, but also ensuring operational interpretability so that increasing predicted risk levels systematically translate to increasingly severe field conditions.

Given a predicted ordinal risk score S and an operational variable Y, the relationship is estimated using a spline regression with spatial and temporal fixed effects:

$$Y = f(S) + Z + D + \varepsilon,$$

where Z and D denote zone and date fixed effects. The monotonic transitions between ordinal levels are then defined as:

$$\Delta_{a\to b} = \mu(b) - \mu(a),$$

where $\mu(s)$ denotes the expected operational outcome associated with risk level s.

For each transition order k, four complementary monotonic metrics are computed:

- MED_k: median monotonic increase across transitions,
- MIN_k: worst observed transition,
- $VIOL_k$: frequency of monotonicity violations,
- NEG_k: average magnitude of negative transitions.

These quantities are combined into a transition consistency score:

$$\mathrm{SCORE}_k = \frac{\mathrm{MED}_k + \mathrm{MIN}_k}{2} - \mathrm{NEG}_k\left(1 + \mathrm{VIOL}_k\right).$$

Positive values indicate coherent monotonic operational behavior, while negative values reveal ordinal inconsistencies and inverted operational transitions. In addition to monotonic explainability metrics, recall is also reported in order to evaluate the ability of the framework to preserve event detection capacity while learning operational ordinal structures.

Table 2: Average operational monotonicity performance across all targets, clusters and architectures. Higher is better for k_i and Recall. Bold values represents the best value of each criteria.

Loss	Recall	k_1	k_2	k_3	k_4
FWI (Expert)	0.730	-0.096	-0.059	-0.019	-0.001
Cross Entropy	0.748	0.159	0.096	0.082	-0.003
RankNet	0.688	0.088	0.172	0.169	**0.133**
CCLLT	0.893	0.053	**0.184**	**0.220**	0.078
CCLLT Ablation Study					
w/o EMA Priors	0.659	-0.026	-0.008	0.011	0.026
w/o Coverage Reg.	0.867	0.043	0.105	0.113	0.073
w/o Midpoint Calib.	**0.948**	**0.100**	0.155	0.065	0.002

6 Performance

For readability, results are averaged across architectures, departments, and operational targets. Table 2 reports average monotonicity scores across all operational targets, architectures, and clustering configurations.

First, CCLLT achieves the highest average recall (0.893) among all baselines, outperforming RankNet (0.688) and the operational FWI index (0.730). This demonstrates that explicitly structuring the latent space through latent centers not only preserves predictive coverage but actively improves event detection capacity compared to pairwise sample ranking.

Second, while predicting a positive risk more frequently than RankNet (as shown by its significantly higher recall), CCLLT learns a risk system of similar operational coherence, and even improves upon RankNet for critical higher-order transitions ($k_3 = 0.220$ vs 0.169). Furthermore, both ranking-based approaches (RankNet and CCLLT) demonstrate a clear improvement in monotonic consistency compared to standard classification with Cross Entropy. Finally, all data-driven models tend to produce much more coherent operational systems than the expert FWI index, which exhibits negative average transitions across the board (e.g., $k_1 = -0.096, k_2 = -0.059$), highlighting its frequent ordinal inversions.

Overall, directly constraining monotonic transitions is a strong alternative to pairwise sample ranking, preserving predictive performance while ensuring a much more robust ordinal consistency across severe risk levels.

6.1 Ablation Study

The ablation study provides insight into the role of each component of CCLLT. First, EMA priors appear essential for learning a stable ordinal structure. Removing them causes a substantial drop in recall (0.893 to 0.659) and deteriorates monotonic consistency across all transition orders, indicating that the latent ordinal centers become significantly less reliable during optimization. Second, coverage regularization contributes to both predictive coverage and ordinal quality. Although the overall degradation remains moderate, removing this term consistently reduces recall and weakens higher-order transitions, suggesting that preserving a balanced risk distribution helps maintain meaningful ordinal structures. Finally, midpoint

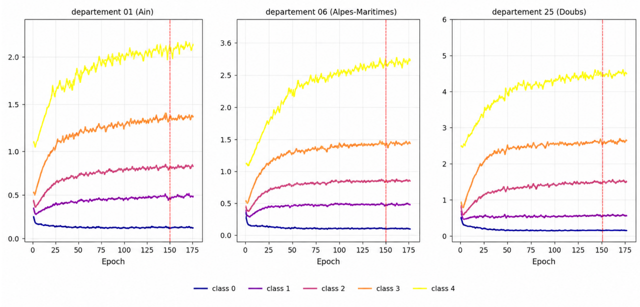

Figure 2: Evolution of EMA ordinal centers for the *the number of wildfire incidents* target.

calibration primarily affects the organization of severe risk levels. While its removal increases recall (0.948), the performance on distant transitions drops markedly (k_3: 0.220⟶0.065, k_4: 0.078⟶0.002), indicating that the model tends to over-predict high-risk states without properly separating them. Overall, the results suggest that EMA stabilization is the most critical component, while coverage regularization and midpoint calibration further improve the quality and interpretability of the learned ordinal system.

6.2 Interpretability of the Learned Ordinal System

Figure 2 illustrates the evolution of EMA ordinal centers. The centers remain strictly ordered without crossings throughout optimization, preserving the ordinal structure of the risk scale, and stabilize after approximately 100 epochs, demonstrating the robustness of the EMA mechanism.

Importantly, the learned systems differ substantially across departments. While ordinal ordering is preserved globally, the distances between successive risk levels vary significantly (e.g., the distance between low and extreme risk is larger in Alpes-Maritimes than in Ain). This indicates risk levels should be interpreted relative to local operational conditions rather than a unique national scale, directly justifying the hierarchical formulation and departmental ordinal priors of CCLLT.

7 Conclusion

We argue that operational wildfire risk should be learned as an ordinal system rather than predicted as isolated events. We introduced CCLLT, a novel ordinal learning objective that constrains monotonic transitions between ordinal risk levels via latent centers, hierarchical shrinkage, and EMA priors.

Experiments on three operational targets demonstrate that the proposed loss produces more coherent monotonic risk structures while preserving recall. Furthermore, the learned centers reveal that operational risk systems differ across departments, supporting the modeling of risk through hierarchical local structures.

The objective is not to determine whether a wildfire will occur, but to organize future situations into a coherent operational severity system. Results suggest that learning such systems requires explicitly modeling ordinal consistency rather than optimizing event prediction alone. Ultimately, operational risk is not a classification problem. It is the construction of an ordinal system whose levels must correspond to increasingly severe operational consequences.

Acknowledgments

All computations have been performed on the Mésocentre de calculs de Franche-Comté supercomputer facilities.

Declaration

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Christopher J. C. Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. 2005. Learning to Rank Using Gradient Descent. In *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*. Association for Computing Machinery, New York, NY, USA, 89–96. doi:10.1145/1102351.1102363

[2] Nicolas Caron, Christophe Guyeux, Hassan Noura, Maxime Coulmeau, and Benjamin Aynes. 2026. Risk Is Not the Target: A Monotonic Framework for Evaluating Wildfire Operational Risk Signals. In *Proceedings of the IEEE International Conference on Computers, Software, and Applications (COMPSAC)*. Accepted for publication; proceedings not yet available.

[3] Nicolas Caron, Hassan Noura, Christophe Guyeux, and Benjamin Aynes. 2025. Localized Forest Fire Risk Prediction: A Department-Aware Approach for Operational Decision Support. In *2025 Computing, Communications and IoT Applications (ComComAp)*. 251–258. doi:10.1109/ComComAp68359.2025.11353174

[4] J. Doe and A. Smith. 2020. Approaching Wildfire Prediction as a Classification Problem. *Journal of AI and Environmental Science* 45 (2020), 123–134.

[5] Spyros Kondylatos, Ioannis Prapas, Gustau Camps-Valls, and Ioannis Papoutsis. 2023. Mesogeos: A multi-purpose dataset for data-driven wildfire modeling in the Mediterranean. arXiv:2306.05144 [cs.CV] https://arxiv.org/abs/2306.05144

[6] Slobodan Milanović, Nenad Marković, Dragan Pamučar, Ljubomir Gigović, Pavle Kostić, and Sladjan D. Milanović. 2020. Forest Fire Probability Mapping in Eastern Serbia: Logistic Regression versus Random Forest Method. (2020). https://doi.org/10.3390/f12010005

[7] Vassiliki Varela et al. 2018. Fire Weather Index (FWI) classification for fire danger assessment applied in Greece. *Tethys Journal of Weather and Climate of the Western Mediterranean* (12 2018). doi:10.3369/tethys.2018.15.03

[8] Lara Vilar, Douglas G. Woolford, David L. Martell, and M. Pilar Martín. 2010. A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain. *International Journal of Wildland Fire* 19, 3 (2010), 325–337. doi:10.1071/WF09030

[9] Van Wagner. 1987. Development and structure of the canadian forest fire weather index system. Canadian Forest Service. (1987). https://cfs.nrcan.gc.ca/publications?id=19927

[10] Xin Wu, Gui Zhang, Zhigao Yang, Sanqing Tan, Yongke Yang, and Ziheng Pang. 2023. Machine Learning for Predicting Forest Fire Occurrence in Changsha: An Innovative Investigation into the Introduction of a Forest Fuel Factor. *Remote Sensing* 15, 17 (2023). doi:10.3390/rs15174208

[11] Yangshuang Xu, Yuyang Dai, Liling Chang, Qi Wang, and Yushun Dong. 2026. Does Your Wildfire Prediction Model Actually Work, or Just Score Well? arXiv:2605.18911 [cs.LG] https://arxiv.org/abs/2605.18911

[12] Guoli Zhang, Ming Wang, and Kai Liu. 2019. Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China. *International Journal of Disaster Risk Science* 10, 3 (2019), 386–403. doi:10.1007/s13753-019-00233-1