environmental and earth sciences proceedings

MDPI

Proceeding Paper

Image-Based Wildfire Behavior Classification Using Convolutional Neural Networks †

Jakov Bejo *, Ljiljana Šerić and Damir Krstinić

Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Ul. Ruđera Boškovića 32, 21000 Split, Croatia; ljiljana@fesb.hr (L.Š.); dkrst@fesb.hr (D.K.)

* Correspondence: Jakov.Bejo.00@fesb.hr

† Presented at the 2nd International Workshop on Extreme Wildfire Events (X-Fire 2026), Prague, Czech Republic, 23–25 June 2026.

Abstract

After ignition, fire behavior is governed by a complex interaction of fuel, weather and topography. In practice, fire behavior is labeled as wind, topography or fuel-driven depending on the dominant driver. These labels are typically assigned through expert judgment and post-fire analysis rather than real-time classification. In this paper, we propose a model based on the ConvNeXtV2 convolutional neural network (CNN), capable of detecting the dominant wildfire driver directly from high-resolution surveillance images. The model is trained using 16 labeled sequences with targeted augmentation, class-balanced sampling and loss weighting to counter strong class imbalance and heavy repetition within sequences. The network shows good results in both fuel- and wind-driven situations but struggles with topography-driven wildfires, even when trained on full-resolution images. Despite the small dataset, our approach illustrates how modern CNNs can complement expert predictions.

Keywords: wildfire behavior; image classification; convolutional neural network; ConvNeXtV2; deep learning; fire surveillance

1. Introduction

Wildfire behavior can usually be described by its driving force, with the most common categories being wind-driven, topography-driven and fuel-driven. These types are usually based on expert judgment after the event rather than tied to measurable variables [1,2]. Related work has shown that a few meteorological, topographic and fuel descriptors can be used for wildfire classifiers [3], but these models demand the accumulation of data from multiple databases. We propose a new approach of utilizing modern computer vision techniques to mimic expert judgment. With this approach, we remove the need for complex data retrieval and harness the emerging technology and data from wildfire surveillance systems [4,5]. Our goal is to test whether a compact, fine-tuned ConvNeXtV2 [6] CNN image classification model can properly distinguish between different wildfire behavior drivers.

2. Data and Methods

Video material of wildfires from the coastal region of Croatia was collected from the OIV (Odašiljači i veze d.o.o.) FireDetectAI video surveillance system [7], operational in Croatia. A lightweight web tool built on a PHP–MySQL stack was used to browse sequences, assign labels and log assumptions about driver types. Each video or sequence

Academic Editor: Theodore M. Giannaros

Published: 9 July 2026

could be segmented when behavior changed over time, and each segment was labeled with its specific driver type. Sequences were labeled by domain experts based on smoke plume geometry, spread direction relative to terrain and prevailing wind conditions at the time of the event. Examples of each class are shown in Figure 1. It is worth noting that all images were screened manually and noisy images were removed. Sequences where fire behavior changed with time were split and used separately. The final dataset contains 5421 images grouped in 16 wildfire sequences, split into labeled frame groups.

Figure 1. From left to right: fuel-, wind-, and topography-driven wildfire examples.

Sequences used for validation were separated from training data before any other augmentation. The results were split such that no single wildfire could be in both validation and training parts. Each frame is labeled according to its parent sequence or its parent sequence segment in case of multiple behaviors during the same wildfire. During both training and inference, the model relies on single-frame input. Although a single snapshot does not carry any temporal or movement cues, the wildfire drivers often has very distinct spatial patterns which the model can pick up on, such as high vertical plumes for fuel-driven fires, horizontal plumes for wind-driven, and terrain following smoke spread for topography-driven wildfires.

Frames are extracted at a 1920×1080 resolution and passed through a two-step preprocessing pipeline. First, the frame is cropped to 1920×1024 in a way that removes top overlays while preserving the sky and horizon. Second, in the case of half-resolution runs, the cropped image is downscaled to 960×512 using Lanczos resampling, which maintains the native 16:9 aspect ratio and preserves edge detail relevant for smoke and terrain structure. ConvNeXtV2 accepts these dimensions without further padding or warping. To reduce overfitting on the small repetitive dataset, we apply random horizontal flips, occasional vertical flips, mild rotations and color jitter during training. Before augmentation, data were split into a 90/10 ratio for train/valid, with a detailed breakdown shown in Table 1. The ConvNeXtV2-Base backbone was fine-tuned in two stages with the AdamW algorithm: a 10-epoch head-only warmup, followed by full fine-tuning with layer-wise LR decay (5×10^{-6} to 2×10^{-5}) and a 5-epoch warmup.

Table 1. Dataset distribution pre-augmentation and class-balancing configuration.

Driver Class	Image Count (Train + Val)	Class Loss Weight
Fuel	1811	0.923
Topography	1353	1.235
Wind	1848	0.874

Class representation reflects local wildfire patterns, with wind-driven fires more common than fuel- or topography-driven cases. We counter this imbalance by combining inverse-frequency class weights in the loss function. Label smoothing with $\varepsilon = 0.1$ further discourages overconfident predictions on majority classes. At inference, we reuse the same deterministic preprocessing. Due to the lack of data, testing performance on separate images was not possible, so we evaluated the results based on per-class precision and recall.

3. Results and Discussion

The half-resolution model (960 $\times$ 512) was trained for 30 epochs in 14 h on a Tesla T4 GPU and gives a validation accuracy of 0.56. The model strongly favors fuel and wind, recovering fuel-dominated fires with a recall of over 75%, but a lower precision at only 0.39, treating many ambiguous scenes as wind-driven. Topography-driven fires are rarely correctly predicted, so precision for this class is very low despite cases having much higher geographical elevation compared to others. Wind-driven events exhibited the best results, with the highest overall accuracy of 0.77 and more balanced recall behavior. Exact numbers are presented in Table 2.

Table 2. Per-class precision for two model versions.

Model/Class	Fuel-Driven	Topography-Driven	Wind-Driven
Half resolution	0.39	0.25	0.77
Full resolution	0.41	0.24	0.79

This pattern suggests that the network latches onto strong smoke plume signatures and horizontal smoke structures but struggles when terrain cues are subtle or partially occluded. Topography-driven cases likely fail because surveillance cameras view terrain at oblique angles that flatten elevation cues, making the classifier rely more on the shape of the smoke. Class-balancing techniques and label strategies help stabilize training but cannot overcome the small number of distinct fire sequences and an imbalanced dataset, so performance remains uneven across classes. A full-resolution test conducted only on original-resolution images showed slightly better results for wind- and fuel-driven behavior, but no improvement for topography-driven.

4. Conclusions

In this paper, we investigated the capability of the ConvNeXtV2 model to classify wildfire surveillance images into fuel-, wind-, and topography-driven behavior classes using a small, unbalanced dataset. The classifier achieves moderate overall performance, with reasonable discrimination of fuel- and wind-dominated events, but fails at recognizing topography-driven fires. Its strengths lie in exploiting the spatial detail of high-resolution imagery, but results remain uneven across classes. Even with the under-representation of certain classes, both the full- and half-resolution models managed to correctly classify half of the overall validation split, providing a good starting point for further research.

The main limitation of this work is the very small number of independent wildfire sequences and the high redundancy of frames within each sequence, which restricts both the model's learning capacity and the robustness of validation scores. Future work will focus on expanding the dataset with more events and more varied viewpoints, refining augmentation and fine-tuning strategies to better target topography-driven cases, and exploring alternative architectures. In particular, replacing the current model with an object detector such as YOLO [8] within the early detection pipeline could improve both accuracy and localization of behavior types while retaining the practical advantages of surveillance-based monitoring. Limitations of the topography-driven class could be improved by integration with GIS data.

Author Contributions: Conceptualization, J.B., L.Š. and D.K.; methodology, J.B., L.Š. and D.K.; software, J.B.; validation, J.B.; formal analysis, J.B.; investigation, J.B.; resources, J.B.; data curation, J.B. and D.K.; writing—original draft preparation, J.B.; writing—review and editing, L.Š. and D.K.; visualization, J.B.; supervision, L.Š. and D.K.; project administration, L.Š. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by a Short-Term Scientific Mission Grant from the COST Action CA22164—European Network on Extreme fiRe behaviOr (NERO).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset used for experiments described in this paper is still in development stage and due to privacy and copyright reasons will not be published at the current time.

Acknowledgments: This work is supported by a Short-Term Scientific Mission Grant from the COST Action CA22164—European Network on Extreme fiRe behaviOr (NERO).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
GPU	Graphics Processing Unit
GIS	Geographic Information System
OIV	Odašiljači i Veze (eng. Transmitters and Communications Ltd.)
YOLO	You Only Look Once

References

1. Lecina-Díaz, J.; Alvarez, A.; Retana, J. Extreme fire severity patterns in topographic, convective and wind-driven historical wildfires of Mediterranean pine forests. *PLoS ONE* **2014**, *9*, e85127. [CrossRef] [PubMed]
2. Finney, M.A.; McAllister, S.S.; Grumstrup, T.P.; Forthofer, J.M. *Wildland Fire Behaviour: Dynamics, Principles and Processes*; CSIRO Publishing: Clayton, Australia, 2021.
3. National Wildfire Coordinating Group. *S290: Intermediate Wildland Fire Behavior*; Student Workbook; NWCG: Boise, Idaho, 2019.
4. Bugarić, M.; Krstinić, D.; Šerić, L.; Stipaničev, D. Current trends in wildfire detection, monitoring and surveillance. *Fire* **2025**, *8*, 356. [CrossRef]
5. Stipaničev, D.; Vuko, T.; Krstinić, D.; Štula, M.; Bodrožić, L. Forest fire protection by advanced video detection system—Croatian experiences. In Proceedings of the Third TIEMS Annual Conference, Seoul, Republic of Korea, 23–26 May 2006; pp. 1–8.
6. Woo, S.; Debnath, S.; Hu, R.; Chen, X.; Liu, Z.; Kweon, I.S.; Xie, S. ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders. *arXiv* **2023**, arXiv:2301.00808.
7. OIV. *OIV FireDetect AI—Smart Communication System for Early Detection of Fire*; Odašiljači i veze d.o.o.: Zagreb, Croatia, 2025.
8. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.