FireSight: Short-Term Fire Hazard Prediction Based on Active Fire Remote Sensing Data

Julia Gottfriedsen[1,2] **Johanna Strebl**[1,2] **Max Berrendorf** **Martin Langer**[2]

Volker Tresp[1]

[1] Ludwig-Maximilians-Universität München, Munich
[2] OroraTech GmbH, Munich

Abstract

Wildfires are becoming unpredictable natural hazards in many regions due to climate change. However, existing state-of-the-art wildfire forecasting tools, such as the Fire Weather Index (FWI), rely solely on meteorological input parameters and have limited ability to model the increasingly dynamic nature of wildfires. In response to the escalating threat, our work addresses this shortcoming in short-term fire hazard prediction. First, we present a comprehensive and high fidelity remotely sensed active fire dataset fused from over 20 satellites. Second, we develop region-specific ML-based 3-7 day wildfire hazard prediction models for South America, Australia, and Southern Europe. The different models cover pixel-wise, spatial and spatio-temporal architectures, and utilize weather, fuel and location data. We evaluate the models using time-based cross-validation and can show superior performance with a PR-AUC score up to 44 times higher compared to the baseline FWI model. Using explainable AI methods, we show that these data-driven models are also capable of learning meaningful physical patterns and inferring region-specific wildfire drivers.

Climate Impact Boreal forests are the largest terrestrial carbon sink, storing more than one-third of all terrestrial carbon [45]. Wildfires are a serious threat to them; for example, between March and July 2023, hundreds of fires burned out of control in Canada, consuming more than 3% of the country's total forest area [6, 7, 26]. These wildfires can turn forests from a carbon sink to a carbon source if vegetation regrowth cannot compensate, further accelerating climate change [2, 14, 20, 40, 45]. At the same time, climate change itself is altering fire regimes worldwide. Regions with ecologically beneficial wildfires are experiencing extreme drought, often leaving insufficient flammable vegetation to support regular fire regimes; Other regions are experiencing more frequent and catastrophic wildfires [4, 14, 22]. The timely and accurate prediction of wildfire hazard is a critical element in limiting the negative impacts of wildfires by supporting the efficient deployment of firefighting resources. It can also help improve early warning systems that protect communities, ecosystems, and critical infrastructure. Machine learning models have proven to be a valuable tool for this prediction [1, 15, 31, 33].

In this paper, we present a high-quality dataset of active fires that has been compiled and harmonized from more than 20 different satellites. We demonstrate its potential for wildfire hazard modeling by building models for up to 1 week in advance wildfire prediction based on weather and fuel data. The models outperform widely used hazard indices such as the Fire Weather Index (FWI) over the entire precision-recall curve. We utilize explainable AI methods to verify that the models use physically plausible patterns. We also discuss how the models can be used to support practical decision making and highlight remaining challenges.

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.

1 Dataset

In this paper, we selected three study regions with very different fuel and weather characteristics [9, 18, 27]: *Australia*, *Brazil*, and *Greece*. We use data from October 2019 - October 2022 aggregated to daily values at a spatial resolution of 0.1 degree ($\approx$ 11 km).

Fire Hazard We approximate the hazard by predicting the occurrence of wildfires. We curated the fire event ground truth data product in our company by assimilating over 20 satellite products from both LEO and GEO satellites[1], employing a spectrum of instruments encompassing infrared and multi-spectral visual sensors to create our active fire satellite data product. A detailed description can be found in Table 1. We chose active fire data over burned area data, due to its real-time nature, ensuring immediate and accurate representation of ongoing fire dynamics. Burned area data often lags by several days, introducing inaccuracies, especially for smaller fires that might be underestimated or missed entirely [5]. The data sources are retrieved from the respective operator/data provider through different terrestrial and satellite-based links. Subsequently, they are standardized in terms of format and type to ensure compatibility for future processing steps. Hotspots representing thermal anomalies are generated in the first processing stage, either by derivation or direct ingestion as third-party data products. The hotspots, along with auxiliary data like Fire Radiative Power (FRP) and brightness temperature, are processed and aggregated along with third-party fire products into active fires clusters. We then employ advanced clustering methods to filter out persistent heat sources (e.g., factories, volcanoes), reflections (e.g., water, solar panels, steel roofs), and false positives through correlation with environmental datasets to enhance the precision of our final active fire satellite data product. Each cluster is assigned a confidence value based on the count of satellites and fire detection algorithms linked to the hotspots within a given cluster. For this study, clusters with a confidence level exceeding 0.5 were included to account for a reasonable degree of certainty.

Hazard Factors Two critical wildfire hazard factors are *weather*, used by virtually all approaches including physical [25], statistical [17], and AI-based [11, 21], and available *fuel*. In our approach, we combine both: We incorporate weather data from the *ERA5-Land* dataset, aggregating hourly measurements to daily at a spatial resolution of 0.1 degrees [12, 16, 36]. We also utilize fuel information via the *ESA CCI Landcover* dataset, which provides fine-grained information about what physical material is present where on the Earth's surface [13, 44]. We group the 22 classes of potential wildfire fuels into nine combined classes (cf. Table 3). We also use derived features that previous research has shown to be helpful in predicting wildfires [32, 41]: we calculate the wind speed, include the geographic location as well as the day of year. We provide further details in Table 1. All data including the fire hazard labels are projected into the ERA-5 native 0.1 degree resolution.

2 Methodology

Modelling We study four different types of models: (1) the Canadian *Fire Weather Index (FWI)* [37] is a widely adopted fire danger rating system, for example in the European Forest Fire Information System (EFFIS). It calculates daily fire danger based on previous day's weather data, but it often overestimates fire hazard and neglects spatial and temporal variations. It also assumes a uniform forest fuel type, which can lead to inaccuracies in ecosystems different from boreal forests [38]. (2) a simple *random forest* [3] baseline which makes independent predictions for each location, (3) a *U-Net* model [34] that can model spatial correlations between adjacent locations, and (4) a *3d U-Net* [10, 23] which receives a temporal sequence of past weather data for up to 7 days in addition to landcover. For further details about model architecture and the training procedure, please refer to Appendix B.

Evaluation Our fire hazard dataset is heavily imbalanced with only 0.26-3.40% of all pixels belonging to the fire class, the total percentage varying across folds, cf. Table 5 in the appendix. The highly unbalanced label set requires careful selection of evaluation metrics. For example, predicting the majority ("no burn") class would lead to high *accuracy* metric while being useless in practice. Thus, we focus on two metrics, the *precision* (How many *will burn* decisions were correct?) and *recall* (How many actually burned pixels did my model correctly predict?). These two metrics naturally compete with each other, and different decision thresholds lead to a trade-off between the

[1]LEO: Low Earth Orbit, GEO: geostationary or Geosynchronous Equatorial Orbit

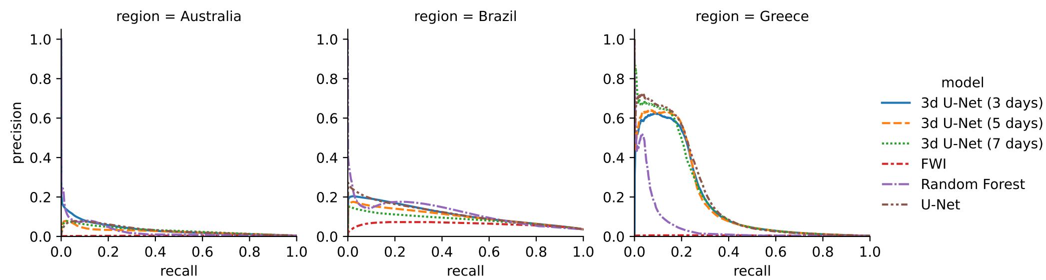


Figure 1: Precision-Recall curves for different models by region. The curves were calculated across all test folds. We generally see best results for Greece and worst for Australia. FWI is inferior to trained models across all regions and the whole precision-recall curve.

two. Therefore, we also provide precision-recall curves that visualize this trade-off. The *area under the precision recall curve (PR-AUC)* is used to summarize the curves into a single number.

Experiments In order to assess the robustness of our models, we have designed a sixfold cross-validation approach that makes optimal use of the limited data available to us, while still accounting for the intricacies of time series forecasting. To avoid data leakage, we use data from 731 consecutive days as our training set and 187 consecutive days as our validation and test sets, respectively. This results in six different folds of about two years of training data and half a year of validation and test data. Further details can be found in Appendix A.

3 Results & Discussion

We show precision-recall curves for all models and study regions in Fig. 1, as well as the aggregated PR-AUC scores in Table 7 (in the appendix). Most notably, the results for all learned models are superior to the baseline Fire Weather Index (FWI) in all regions and along the entire precision-recall curve. In terms of PR-AUC, the difference between FWI and the best learned models can be as large as 44-fold as seen for Greece. This underscores the advantages of learned models that can adapt to region-specific fire regimes, as opposed to hand-crafted indices that require manual adaptation.

A comparison between the learned models is more nuanced: There is no single best model, but different models are preferred depending on the region and the operating point within the precision-recall trade-off. Within the U-net family, the 3D U-net with a 3-day lead time is often among the best. We see a strong difference when comparing different regions, with the best results for Greece and the worst for Australia.

In the context of hazard modelling, we tend to prioritize recall over precision: Missing a fire event is more critical than false alarms. However, in practice, we are faced with limited resources, and thus precision cannot be completely ignored. The exact operational state depends on local conditions and will be different in different scenarios.

While we can show impressive improvements over widely used indices, both in terms of precision and recall, and across different operating points, the results are still far from perfect. For instance, achieving a 50% recall rate (where every other fire event goes undetected) results in a precision ranging between 5-25%. This may be because wildfires require not only favorable conditions, but also ignition events. The latter are often difficult to model: Lightning strikes are a known cause of wildfires and are often hyper-local in nature, requiring high-resolution data. Moreover, in most cases, the human factor play a decisive role [19, 24]. While inclusion of input data such as socio-economic factors may partially improve this, cases such as arson remain hard to model.

This highlights the need for an integrated approach: While hazard models can help prioritize wildfire mitigation resources and raise public awareness in times of risk, wildfires will continue to ignite. Thus, additional active fire monitoring and forecast remain crucial components in mitigating the negative impact of wildfires.

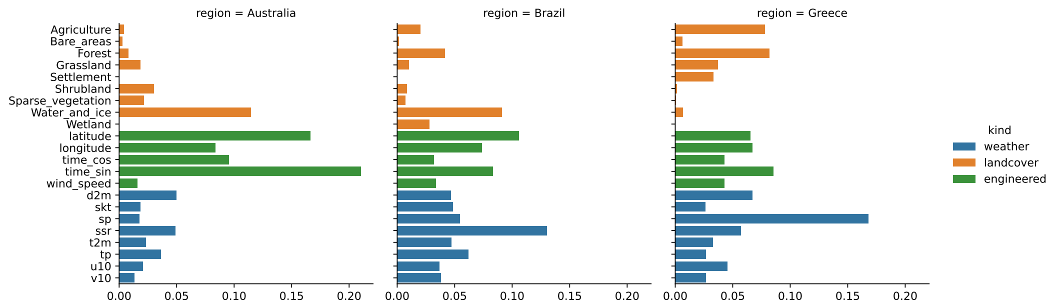


Figure 2: Relative feature importance: The bar heights indicate the importance in favour of a high fire hazard decision. We only show the absolute importance (i.e., a large bar indicates either a strong positive *or* negative impact on the decision). The bar heights have been normalized such that they sum to one for each of the three subplots. The color indicates the feature class (orange: fuel, green: engineered, blue: weather).

Explainable AI - Feature Importance Shapley values [8, 39] are a method for feature importance originating from game theory. We utilized this method to exemplarily assess the best performing model's (3D Unet) ability to learn physical relationships, cf. Fig. 2 and further details in Appendix C. The model is able to derive sensible links between fire drivers and forest fires from raw data. As shown in Fig. 2 different regions display distinct factors contributing to high fire probability predictions, supporting the need for region-specific models. In Australia, the model incorporates simple key factors: location, season, and the absence of water and ice. Location and time of the year are important due to the concentration of flammable vegetation along the coast and the shifting fire seasons between north and south, which reflects diverse fire patterns across the country [35]. In Brazil, solar radiation, latitude, and precipitation are the primary contributing factors to forest fire hazard. Latitude's significance is consistent with studies identifying distinct forest fire hazard zones at different latitudes, as well as regions with drier conditions [30]. Fig. 4 shows that dry conditions have an increasingly positive contribution towards a *will burn* prediction in the three days leading up to an ignition, whereas muggy conditions have an opposing effect. In Greece, location plays a less significant role due to its small size. Instead, factors such as vegetation, dry conditions, and wind become more crucial, consistent with findings by previous research [28].

Outlook As our fire hazard models are trained based on historical fire occurrences, instances where conditions posed a hazard but did not result in ignition are treated as non-hazardous. To mitigate this limitation for operational utilization of the models, one potential approach is to augment the model output by incorporating Normalized Difference Moisture Index (NDMI) based fuel dryness as an overlay within the operational hazard assessment. More frequent and high resolution active fire data would further enhance the model's performance. This would help provide a more comprehensive and nuanced understanding of fire hazard, encompassing both historical ignition events and conditions conducive to fire spread. Additionally, including very relevant human factor [19, 24] in the input data besides location through e.g. socioeconomic or infrastructure data could further aid the results as a proxy for ignition.

Acknowledgments and Disclosure of Funding

We would like to express our gratitude to the ESA ϕ-lab, namely Bertrand Le Saux as well as Nicolas Longepe for their valuable support during this research project. We would also like to thank the National Observatory of Athens and Pi School for the fruitful collaboration and insight.

References

[1] Tristan Ballard, Gopal Erinjippurath, Matthew W Cooper, and Chris Lowrie. Widespread increases in future wildfire risk to global forest carbon offset projects revealed by explainable ai. In *ICLR 2023 Workshop on Tackling Climate Change with Machine Learning*, 2023.

[2] Jason Beringer, Lindsay B Hutley, David Abramson, Stefan K Arndt, Peter Briggs, Mila Bristow, Josep G Canadell, Lucas A Cernusak, Derek Eamus, Andrew C Edwards, et al. Fire in Australian savannas: from leaf to landscape. *Global Change Biology*, 21(1):62–81, 2015.

[3] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[4] California Department of Forestry and Fire Protection. Top 20 largest California wildfires. Oct 2022. Available at `https://www.fire.ca.gov/media/4jandlhh/top20_acres.pdf`.

[5] ML Campagnolo, R Libonati, JA Rodrigues, and JMC Pereira. A comprehensive characterization of modis daily burned area mapping accuracy across fire sizes in tropical savannas. *Remote Sensing of Environment*, 252:112115, 2021.

[6] Canadian Interagency Forest Fire Centre Inc. Fire statistics, 2023. Available at `https://ciffc.net/statistics`.

[7] Canadian Interagency Forest Fire Centre Inc. Situation reports, 2023. Available at `https://ciffc.net/situation/archive`.

[8] Javier Castro, Daniel Gómez, and Juan Tejada. Polynomial calculation of the shapley value based on sampling. *Computers & Operations Research*, 36(5):1726–1730, 2009.

[9] G Christakis. Attica wildfires: The intersection of governance failures and climate change, 2021. Available at `https://hphr.org/31-article-christakis`.

[10] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.

[11] Roberto Cilli, Mario Elia, Marina D'Este, Vincenzo Giannico, Nicola Amoroso, Angela Lombardi, Ester Pantaleo, Alfonso Monaco, Giovanni Sanesi, Sabina Tangaro, Roberto Bellotti, and Raffaele Lafortezza. Explainable artificial intelligence (xai) detects wildfire occurrence in the mediterranean countries of southern europe. *Scientific Reports*, 12, 09 2022.

[12] Copernicus. Copernicus climate data store, 2022. Available at `https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview`.

[13] Pierre Defourny. Product user guide and specification, Jun 2021. Available at `https://datastore.copernicus-climate.eu/documents/satellite-land-cover/D5.3.1_PUGS_ICDR_LC_v2.1.x_PRODUCTS_v1.1.pdf`.

[14] Stefan H. Doerr and Cristina Santín. Global trends in wildfire and its impacts: Perceptions versus realities in a changing world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1696):20150345, 2016.

[15] Marissa Dotter. Burnmd: A fire projection and mitigation modeling dataset. In *ICLR 2023 Workshop on Tackling Climate Change with Machine Learning*, 2023.

[16] ECMWF. ERA5: Data documentation, Nov 2022. Available at `https://confluence.ecmwf.int/display/CKB/ERA5`.

[17] Saeedeh Eskandari, Hamid Reza Pourghasemi, and John P. Tiefenbacher. Relations of land cover, topography, and climate to fire occurrence in natural regions of iran: Applying new data mining techniques for modeling and mapping fire danger. *Forest Ecology and Management*, 473:118338, 2020.

[18] Tim Fridtjof Flannery. *The future eaters: an ecological history of the Australasian lands and people*. Grove Press, 2002.

[19] Piyush Jain, Sean CP Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D Flannigan. A review of machine learning applications in wildfire science and management. *Environmental Reviews*, 28(4):478–505, 2020.

[20] Michael Jerrett, Amir S. Jina, and Miriam E. Marlier. Up in smoke: California's greenhouse gas reductions could be wiped out by 2020 wildfires. *Environmental Pollution*, 310:119888, 2022.

[21] Spyros Kondylatos, Ioannis Prapas, Michele Ronco, Ioannis Papoutsis, Gustau Camps-Valls, María Piles, Miguel-Ángel Fernández-Torres, and Nuno Carvalhais. Wildfire danger prediction and understanding with deep learning. *Geophysical Research Letters*, 49(17):e2022GL099368, 2022.

[22] Reason L Machete and Kebonyethata Dintwe. Cyclic trends of wildfires over sub-saharan africa. *Fire*, 6(2):71, 2023.

[23] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.

[24] Tichaona Mukunga, Matthias Forkel, Matthew Forrest, Ruxandra-Maria Zotta, Nirlipta Pande, Stefan Schlaffer, and Wouter Dorigo. Effect of socioeconomic variables in predicting global fire ignition occurrence. *Fire*, 6(5):197, 2023.

[25] Natural Resources Canada, 2022.

[26] Natural Resources Canada. How much forest does canada have?, Feb 2023. Available at `https://natural-resources.canada.ca/our-natural-resources/forests/state-canadas-forests-report/how-much-forest-does-canada-have/17601`.

[27] Ubirajara Oliveira, Britaldo Soares-Filho, Mercedes Bustamante, Leticia Gomes, Jean P Ometto, and Raoni Rajão. Determinants of fire impact in the brazilian biomes. *Frontiers in Forests and Global Change*, 5:735017, 2022.

[28] A Papadopoulos, AK Paschalidou, PA Kassomenos, and G McGregor. Investigating the relationship of meteorological/climatological conditions and wildfires in greece. *Theoretical and applied climatology*, 112:113–126, 2013.

[29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[30] Vânia R Pivello, Ima Vieira, Alexander V Christianini, Danilo Bandini Ribeiro, Luciana da Silva Menezes, Christian Niel Berlinck, Felipe PL Melo, José Antonio Marengo, Carlos Gustavo Tornquist, Walfrido Moraes Tomas, et al. Understanding brazil's catastrophic fires: Causes, consequences and policy needed to prevent future tragedies. *Perspectives in Ecology and Conservation*, 19(3):233–255, 2021.

[31] Ioannis Prapas, Akanksha Ahuja, Spyros Kondylatos, Ilektra Karasante, Eleanna Panagiotou, Lazaro Alonso, Charalampos Davalas, Dimitrios Michail, Nuno Carvalhais, and Ioannis Papoutsis. Deep learning for global wildfire forecasting. *arXiv preprint arXiv:2211.00534*, 2022.

[32] Haiganoush K Preisler and Anthony L Westerling. Statistical model for forecasting monthly large wildfire events in western united states. *Journal of Applied Meteorology and Climatology*, 46(7):1020–1030, 2007.

[33] Eduardo R Rodrigues, Campbell D Watson, Bianca Zadrozny, and Gabrielle Nyirjesy. Firo: A deep-neural network for wildfire forecast with interpretable hidden states. In *NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning*, 2022.

[34] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[35] Jeremy Russell-Smith, Cameron P Yates, Peter J Whitehead, Richard Smith, Ron Craig, Grant E Allan, Richard Thackway, Ian Frakes, Shane Cridland, Mick CP Meyer, et al. Bushfires 'down under': patterns and implications of contemporary australian landscape burning. *International Journal of Wildland Fire*, 16(4):361–377, 2007.

[36] Helen Setchell. Ecmwf reanalysis v5, Nov 2022. Available at `https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5`.

[37] Brian J Stocks, BD Lawson, ME Alexander, CE Van Wagner, RS McAlpine, TJ Lynham, and DE Dube. The Canadian forest fire danger rating system: an overview. *The Forestry Chronicle*, 65(6):450–457, 1989.

[38] S. W. Taylor, Douglas G. Woolford, C. B. Dean, and David L. Martell. Wildfire prediction to inform fire management: Statistical science challenges. *Statistical Science*, 28(4):586 – 615, 2013.

[39] William Thomson. The shapley value, a crown jewel of cooperative game theory. *Handbook of the Shapley Value*, pages 1–15, 2019.

[40] Alexandra Tyukavina, Peter Potapov, Matthew C Hansen, Amy H Pickens, Stephen V Stehman, Svetlana Turubanova, Diana Parker, Viviana Zalles, André Lima, Indrani Kommareddy, et al. Global trends of forest loss due to fire from 2001 to 2019. *Frontiers in Remote Sensing*, 3:825190, 2022.

[41] Sally S-C Wang, Yun Qian, L Ruby Leung, and Yang Zhang. Identifying key drivers of wildfires in the contiguous us using machine learning and game theory interpretation. *Earth's Future*, 9(6):e2020EF001910, 2021.

[42] Adrian Wolny. Wolny/pytorch-3dunet: 3d u-net model for volumetric semantic segmentation written in pytorch, Jun 2023. Available at `https://github.com/wolny/pytorch-3dunet`.

[43] Adrian Wolny, Lorenzo Cerrone, Athul Vijayan, Rachele Tofanelli, Amaya Vilches Barro, Marion Louveaux, Christian Wenzl, Sören Strauss, David Wilson-Sánchez, Rena Lymbouridou, Susanne S Steigleder, Constantin Pape, Alberto Bailoni, Salva Duran-Nebreda, George W Bassel, Jan U Lohmann, Miltos Tsiantis, Fred A Hamprecht, Kay Schneitz, Alexis Maizel, and Anna Kreshuk. Accurate and versatile 3d segmentation of plant tissues at cellular resolution. *eLife*, 9:e57613, jul 2020.

[44] Daniele Zanaga, Ruben Van De Kerchove, Dirk Daems, Wanda De Keersmaecker, Carsten Brockmann, Grit Kirches, Jan Wevers, Oliver Cartus, Maurizio Santoro, Steffen Fritz, Myroslava Lesiv, Martin Herold, Nandin-Erdene Tsendbazar, Panpan Xu, Fabrizio Ramoino, and Olivier Arino. Esa worldcover 10 m 2021 v200, October 2022.

[45] Bailu Zhao, Qianlai Zhuang, Narasinha Shurpali, Kajar Köster, Frank Berninger, and Jukka Pumpanen. North american boreal forests are a large carbon source due to wildfires from 1986 to 2016. *Scientific reports*, 11(1):7723, 2021.

Figure 3: We chose different study regions spread across the globe with diverse ecosystems and fire regimes.

A Data Details

- We provide a detailed description of all features in Table 1.
- Moreover, Table 2 lists all satellites used for deriving the active fire data.
- Table 3 gives details about the categories we derived from ESA CCI Landcover classes.
- Fig. 3 shows the study regions.
- For an introduction to circular time encoding, cf., e.g., `https://developer.nvidia.com/blog/three-approaches-to-encoding-time-information-as-features-for-ml-models/#approach-2-cyclical-encoding-with-sine-cosine-transformation`
- We use a 6-fold time-based cross-validation.
  - We provide a detailed description of how days were assigned to folds in Table 4.
  - Table 5 shows the active fire frequency by fold and region.

B Training Details

- We normalize all weather input features and the engineered ones to have zero mean and unit variance; the landcover is already in a normalized, one-hot format.
- To account for the class imbalance, we use a loss weight of 50 for the positive (fire) class.
- Our UNet implementation is based on the PyTorch 3D UNet repository[2] from [43, 42]. Random forests were implemented in `scikit-learn` [29].
- More details about the models' hyperparameters can also be found in Table 6.

C Feature Importance Details

- We calculated the Shapely values [8][3] using the `captum`[4] Python package
- Since we are dealing with a segmentation problem, we get a feature importance tensor of the same shape as the input ($number\_of\_features \times lead\_time \times longitude \times latitude$), containing negative and positive attributions towards the decision for each output pixel.
- In order to get an aggregated view for each region (cf. Fig. 2, we first took the absolute values of the attributions scores, and then aggregated all inputs for the same features using mean aggregation.
- For each sample in each region we chose the pixel with the largest predicted fire hazard that actually burned.

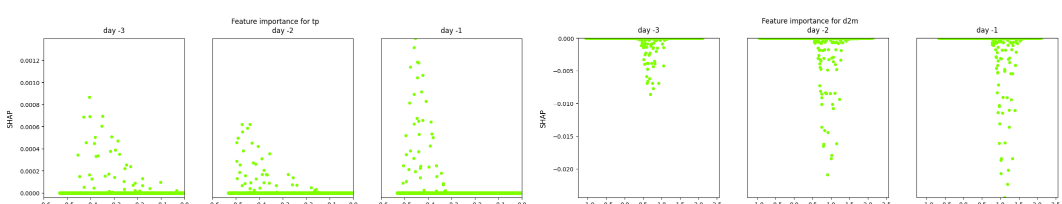
[2]https://github.com/wolny/pytorch-3dunet

[3]`https://christophm.github.io/interpretable-ml-book/shapley.html` may be a more accessible introductory resources

[4]`https://captum.ai/`

Table 1: A list of all features with a description and their units.

source	variable	description	unit
ERA-5 land	`ssr`	Surface net solar radiation, clear sky	J/m^2
	`u10`	10m u-component of wind, i.e., eastward component of the wind, 10m above the Earth's surface	m/s
	`v10`	10m v-component of wind, i.e., northward component of the wind, 10m above the Earth's surface	m/s
	`d2m`	2m dewpoint temperature, i.e., the temperature to which the air, at 2m above the Earth's surface, would have to be cooled for saturation to occur.	K
	`t2m`	2m temperature, i.e., the temperature of air at 2m above the Earth's surface.	K
	`skt`	Skin temperature, i.e., the temperature of the uppermost surface layer.	K
	`sp`	Surface pressure, i.e., the pressure of the atmosphere on the Earth's surface.	Pa
	`tp`	Total precipitation, i.e., the accumulated liquid and frozen water that falls to the Earth's surface.	m
ESA CCI Landcover	agriculture	cf. Table 3	one-hot encoded
	forest		
	grassland		
	wetland		
	settlement		
	shrubland		
	sparse vegetation		
	bare areas		
	water and ice		
engineered	wind speed	Wind speed calculated from `u10` and `v10`	m/s
	sine day of year	Sine of circular encoding of day of year, i.e., $\sin\left(2\pi \cdot \frac{\textit{day-of-year}}{365}\right)$	unit-less
	cosine day of year	Cosine of circular encoding of day of year, i.e., $\cos\left(2\pi \cdot \frac{\textit{day-of-year}}{365}\right)$	unit-less
	longitude	Degrees East/West	degrees
	latitude	Degrees North/South	degrees
	active fires	active fire dataset integrated across 20 satellites described in Table 2	binary



(a) Total precipitation (b) 2m dewpoint temperature

Figure 4: Feature importance for one correct *will burn* prediction in Brazil, predicted by 3D U-Net with a lead time of three days. On *x-axis*, the feature values are plotted, on *y-axis* the Shapley values. Low total precipitation shows a growing positive influence on a *will burn* prediction in the three days leading up to the ignition of a fire. The 2m dewpoint temperature, on the other hand, is indicative of humid conditions and has an increasingly inhibitory contribution the closer to the fire.

Table 2: A description of the satellites used to produce the active fire data product. LEO refers to low earth orbit satellites and GEO refers to geostationary satellites. Different satellites have different instruments that can operate in the visual or infrared spectrum. We group satellites with identical instruments into row groups, e.g. the row group of Sentinel-3A and Sentinel-3B means that they both have an SLSTR and an OLCI instrument. There are additional third-party derived data products based on the instrument measurements. We integrate them all into a single active fire data product.

Type	Satellite(s)	Product	Description	Native Resolution
LEO	Aqua Terra	3rd party	MODIS Instrument-based active fire data product	1,000 m
LEO	Suomi-NPP NOAA-20	infrared 3rd party	VIIRS instrument VIIRS instrument-based active fire data product	750 m 375 m
LEO	MetOp-B MetOp-C	infrared	AVHRR/3 instrument	1,100 m
LEO	Sentinel-2A Sentinel-2B	visual	MSI instrument	10 m
LEO	Sentinel-3A Sentinel-3B	infrared visual 3rd party	SLSTR instrument OLCI instrument SLSTR instrument-based active fire data product	1,000 m 300 m 1,000 m
LEO	Landsat-8	visual	OLI instrument	30 m
LEO	Landsat-9	visual	OLI-2 instrument	30 m
LEO	Fengyun-3D	infrared	MERSI-2 instrument	1,000 m
GEO	GOES-16 GOES-17	3rd party	ABI instrument-based active fire data product	2,000 m
GEO	GK-2A	3rd party	AMI instrument-based active fire data product	2,000 m
GEO	Himawari-8	3rd party	AHI instrument-based fire data product for Australia	2,000 m
GEO	Meteosat-8 Meteosat-9 Meteosat-10 Meteosat-11	3rd party	SEVIRI instrument-based active fire data product	3,000 m

- For the scatter plot (cf. Fig. 4), the raw Shapley values were plotted against the feature values of select input features.

D Additional Results

- Table 7 shows the PR-AUC score per model and region.

Table 3: Combination of ESA CCI Landcover classes in our experiments according to [13]. *Category* denotes our combined categories, while *Landcover Classes* corresponds to the numeric classes with their descriptions in *Landcover Names*.

Category	Landcover Classes	Landcover Names
Agriculture	10, 11, 12	Cropland, rainfed
	20	Cropland, irrigated or post-flooding
	30	Mosaic cropland (>50%) / natural vegetation; (tree, shrub, herbaceous cover) (<50%)
	40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)
Forest	50	Tree cover, broadleaved, evergreen, closed to open (>15%)
	60, 61, 62	Tree cover, broadleaved, deciduous, closed to open (>15%)
	70, 71, 72	Tree cover, needleleaved, evergreen, closed to open (>15%)
	80, 81, 82	Tree cover, needleleaved, deciduous, closed to open (>15%)
	90	Tree cover, mixed leaf type (broadleaved and needle-leaved)
	100	Mosaic tree and shrub (>50%) / herbaceous cover (<50%)
	160	Tree cover, flooded, fresh, or brackish water
	170	Tree cover, flooded, saline water
Grassland	110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)
	130	Grassland
Wetland	180	Shrub or herbaceous cover, flooded, fresh/saline/brackish water
Settlement	190	Urban areas
Sparse vegetation	140	Lichens and mosses
	150, 151, 152, 153	Sparse vegetation (tree, shrub, herbaceous cover) (<15%)
Shrubland	120, 121, 122	Shrubland
Bare areas	200, 201, 202	Bare areas
Water and ice	210	Water bodies
	220	Permanent snow and ice

Table 4: Time series cross-validation. We use data from 731 consecutive days as our training set, and 187 consecutive days as validation and test set, to make the best use of our data while avoiding data leakage. This leads to six different splits of approximately two years of training data and half a year of validation, and test data for each run.

<table>
<tr><th rowspan="2">fold</th><th colspan="5">days</th></tr>
<tr><th>0-186</th><th>187-373</th><th>374-730</th><th>731-918</th><th>919-1105</th></tr>
<tr><td>1</td><td colspan="3">train</td><td>validation</td><td>test</td></tr>
<tr><td>2</td><td colspan="3">train</td><td>test</td><td>validation</td></tr>
<tr><td>3</td><td>validation</td><td colspan="3">train</td><td>test</td></tr>
<tr><td>4</td><td>test</td><td colspan="3">train</td><td>validation</td></tr>
<tr><td>5</td><td>validation</td><td>test</td><td colspan="3">train</td></tr>
<tr><td>6</td><td>test</td><td>validation</td><td colspan="3">train</td></tr>
</table>

Table 5: Frequency of active fire in each of the folds and regions.

fold	1	2	3	4	5	6	total
Australia	0.50%	0.50%	0.22%	0.41%	0.39%	0.22%	0.37%
Brazil	4.47%	4.47%	2.21%	3.84%	3.18%	2.21%	3.40%
Greece	0.43%	0.43%	0.08%	0.27%	0.24%	0.08%	0.26%

Table 6: Training parameters for our experiments.

training parameter	value(s)	comments
positive weight	50	
criterion	BCE or Dice	BCE: Greece, Brazil Dice: Australia
crop size	60	
crop true	true	
center crop	false	
learning rate	0.001	
max epochs	300	
batch size	8	
num workers	4	
timesteps	1, 3, 5, or 7	depending on lead time
feature maps	[16, 32, 64, 128, 256]	
# groups	2	
pooling kernel size	[1, 2, 2]	only required if lead time >1

Table 7: PR-AUC scores by region and model in percent. Larger values indicate better prediction. Bold font shows the largest entries per region. We notice that all learned models perform superior to the FWI. For Brazil, the best result is obtained by Random Forest, while on the other two, 3D U-Net performs best. Comparing regions, the best scores can be obtained for Greece, while Australia shows considerably worse performance.

	PR-AUC		
	Australia	Brazil	Greece
FWI	0.39%	6.24%	0.51%
Random Forest	2.64%	**11.75%**	4.93%
U-Net	3.23%	11.46%	**19.93%**
3d U-Net (3 days)	**3.60%**	11.31%	17.95%
3d U-Net (5 days)	2.53%	10.22%	18.32%
3d U-Net (7 days)	3.10%	8.73%	18.84%