Deep Learning Methods for Daily Wildfire Danger Forecasting

Ioannis Prapas[1,2,*], Spyros Kondylatos[1,2,*], Ioannis Papoutsis[1], Gustau Camps-Valls[2], Michele Ronco[2], Miguel-Ángel Fernández-Torres[2], Maria Piles Guillem[2], and Nuno Carvalhais[3]

[1]IAASARS[†], National Observatory of Athens
[2]Image Processing Laboratory (IPL), Universitat de València
[3]Max Planck Institute for Biogeochemistry
[*]Authors contributed equally

Abstract

Wildfire forecasting is of paramount importance for disaster risk reduction and environmental sustainability. We approach daily fire danger prediction as a machine learning task, using historical Earth observation data from the last decade to predict next-day's fire danger. To that end, we collect, pre-process and harmonize an open-access datacube, featuring a set of covariates that jointly affect the fire occurrence and spread, such as weather conditions, satellite-derived products, topography features and variables related to human activity. We implement a variety of Deep Learning (DL) models to capture the spatial, temporal or spatio-temporal context and compare them against a Random Forest (RF) baseline. We find that either spatial or temporal context is enough to surpass the RF, while a ConvLSTM that exploits the spatio-temporal context performs best with a test Area Under the Receiver Operating Characteristic of 0.926. Our DL-based proof-of-concept provides national-scale daily fire danger maps at a much higher spatial resolution than existing operational solutions.

1 Introduction and Application Context

Globally, wildfires are a major natural hazard, which disrupts natural ecosystem services, causes loss of lives, properties and infrastructure [1], and contributes to carbon dioxide emissions. Climate change is playing an increasing role in determining wildfire regimes and is expected to aggravate wildfires in most parts of the Earth, and quite prominently in the wider Mediterranean [2]. Thus, assessing the likelihood of large fire events is of utmost importance for fire management services.

Traditionally, Fire Danger is predicted by state-wide Fire Danger Rating systems (FDRS), such as the Canadian FDRS [1] or the Australian FDRS [2]. In Europe, the European Forest Fire Information system (EFFIS) [3] uses a similar approach to the Canadian FDRS, which is based on a meteorologically derived Fire Weather Index (FWI). Our work considers vegetation characteristics captured from time-series of Earth Observation (EO) data, and information layers connected to human activity, in addition to meteorological data, and learns to predict fire danger in a data-driven way.

We create a carefully harmonized datacube that contains both historical wildfires and covariates that drive these wildfires. On top, we train Deep Learning (DL) models to forecast the next day's fire danger and evaluate these models in real conditions. This proof-of-concept is developed in

---

[†]Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing

[1]https://cwfis.cfs.nrcan.gc.ca/background/summary/fdr

[2]https://www.afac.com.au/initiative/afdrs

partnership with Fire Brigade officials, towards an operational national fire management support tool, for effectively allocating firefighting resources, designing and applying mitigation measures and timely responding to wildfire disasters.

2 Related Work and Contributions

Wildfire danger forecasting is not a typical Machine Learning (ML) task and poses three great challenges. First, wildfires are caused by the complex interactions of the fire drivers (weather conditions, land and vegetation characteristics, human activities) [4]. Second, wildfire occurrence is inherently stochastic; the same Earth system conditions may lead to major fire events or not. This makes sampling negative examples complex, as the lack of a fire event does not mean lack of fire danger. Third, wildfires are rare events, which can potentially lead to a highly imbalanced dataset.

Reichstein et al. [5] highlighted fire occurrence and spread as typical examples of applications where both spatial and temporal context are highly relevant and could profit from DL modeling. However, relevant DL solutions remain still in their infancy [6, 7]. Huot et al. [7] propose a modern DL approach for fire likelihood estimation, viewing it as a segmentation task and employing U-NET type architectures to predict the daily or weekly-aggregated fire masks of MODIS [8]. However, they make binary predictions (fire, no-fire) which do not account for the stochasticity of fire occurrence, and do not identify the fire ignitions associated with the fire masks, which can lead to data leakage, predicting fires that are already burning. Zhang et al. [6] combine weather and remote sensing data as input to a Convolutional Neural Network (CNN) to model forest fire susceptibility. However, they only exploit the spatial context, ignoring the temporal aspect of the phenomenon. Moreover, a common problem in both works is that they define random training-testing splits, which overestimates the models' performance in the real world. Instead, fire prediction should be treated as a forecasting task [9], that is to train the models in the past to predict the future. Other related work [10] either focuses on a small region or uses shallow models that cannot handle the spatio-temporal context.

Our work builds upon the presented methods and moves a step forward on better estimating fire likelihood, making the following contributions: **First, we create, harmonize, curate and openly publish a large datacube** [11] containing fire driver variables and burned areas for the period 2009-2020 in Greece, a fire-prone country of the Mediterranean. **Second, we rigorously formulate fire danger forecasting as an ML problem.** For that purpose, we identify the fire events associated with large burned areas, along with their starting date, which allows us to model the joint probability of fires occurring and becoming large. Then, we treat fire prediction as a forecasting task by doing a time split to create the training, validation and test sets. Moreover, we do a careful negative sampling in order to account for the dataset imbalance. **Third, we propose and compare a variety of DL models** that are able to capture spatial, temporal or spatio-temporal context. We present an operational prototype, which predicts the next day's fire danger at a national scale in Greece.

3 Data harmonization

Our dataset covers most of Greece and surrounding area, and contains climate, vegetation and human activity variables, which influence the ignition and spread of a fire [1]. The target, as explained in Section 4.1, is defined by the burned area resulting from a fire that started on a given day. We gather all variables and harmonize them into a 1 km $\times$ 1 km $\times$ 1 day resolution datacube, which covers an area of 700 km $\times$ 562 km for the years 2009-2020.

More specifically, we gather the following data: 1. **Daily weather data** from ERA-5 Land [12], for 5 hours (04:00, 08:00, 12:00, 16:00, 20:00) of 2 m temperature, 10 m wind u-component, 10 m wind v-component and total precipitation, available in 9 km spatial resolution. 2. **Satellite variables** from MODIS downloaded from Nasa's portal [3], including Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation (Fpar), available in a 8-daily temporal and 500 m spatial resolution; Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), available in a 16-daily temporal and 500 m spatial resolution; and Day/Night Land Surface Temperature (LST), available in a daily temporal and 1 km spatial resolution. 3. **Roads density.** Road networks are extracted from OpenStreetMap[4], in order to calculate street density at 1 km

---

[3]https://modis.gsfc.nasa.gov/data/
[4]https://www.openstreetmap.org/

spatial resolution. 4. **Population density** extracted from WorldPop, where yearly measurements are available. 5. **Land cover** extracted from Copernicus Corine Land Cover (CLC) [13] for years 2006, 2012 and 2018, at 100 m spatial resolution. 6. **Topography variables** from Copernicus EU-DEM [14] including elevation, aspect and slope at 30 m spatial resolution. 7. **Historical burned areas** from the European Forest Fire Information System (EFFIS) [3] containing burned areas larger than 30 hectares (ha). The EFFIS burned areas are intersected with the MODIS active fires product [8] to recover the start date of the fire.

The data sources are aggregated into a datacube of 19 features. The dynamic attributes (13 features) are the daily min/max 2 m temperature, min/max u-component and v-component of wind, max total precipitation, Fpar, LAI, Day/Night LST, NDVI and EVI. To simulate a real-time setting, the satellite attributes are shifted forward one day, as MODIS provides near real-time data at a daily temporal resolution and a day's observations are available after the satellite passes the location. The weather attributes are not shifted as we assume them to be available from forecasts. The static attributes (6 features) are the CLC, mean elevation, aspect and slope, along with roads and population density. We **publish this datacube** in order to be reused by the community [11].

4 Experiments

4.1 Experimental setup

Wildfire emergency response decision makers are primarily interested in the next day spatial probability of large wildfires [15]. If FI is the event that a wildfire ignites and FL is the event that a wildfire becomes large (> 30 ha), we wish to predict the joint probability of a fire occurring and becoming a large fire $FI \cap FL$ [15]. This translates to $P(FI \cap FL) = P(FI)P(FL|FI)$.

To model $P(FI \cap FL)$, we extract our training, validation and testing sets from the datacube introduced above, as follows:

For a given pixel (i.e. a cell representing a 1 km × 1 km square region) and a given day, we extract input-target pairs for 4 different modeling modalities. First, the *pixel dataset*, where we extract the input attributes and their last 10-days average (only for the dynamic input attributes). Second, the *temporal dataset*, where we extract the last 10-days time-series of the input attributes. Third, the *spatial dataset*, where we extract 25 km × 25 km patches centered spatially around the given pixel. Fourth, the *spatio-temporal dataset*, where we extract 25 km × 25 km × 10 days blocks centered spatially around the given pixel. Static variables are repeated in the time dimension. All input variables are used, except for the CLC, because it is used to determine the negative samples, as discussed next. *The target is for all types of datasets the same*; the next-day burned value of the given pixel. This dataset extraction process is visualized in Figure 1, where each type of dataset is then fed to a different type of ML model.
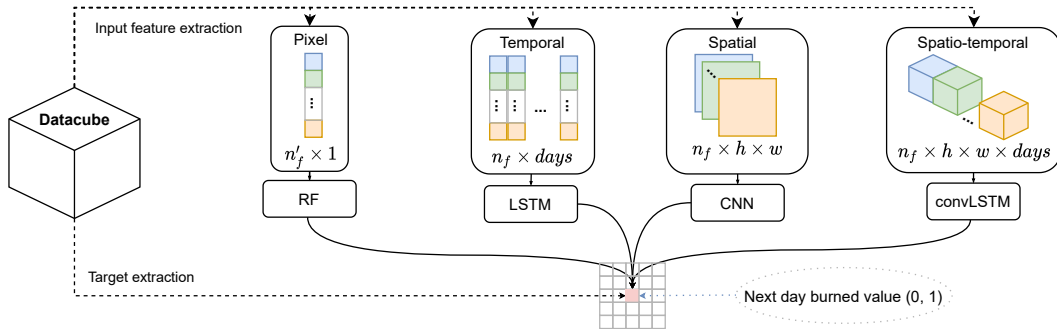

Figure 1: Four different types of dataset samples (pixel, temporal, spatial, and spatio-temporal) are extracted from the datacube and then fed to the corresponding models (RF, LSTM, CNN, and convLSTM). For a dataset sample, n_f is the number of input features, $days$ is the number of days in the time series, h is the height and w is the width, wherever applicable. The pixel dataset has a different number of input features (n_f'), because it also includes aggregations of the dynamic features.

3

Positive examples (pixels contained in burned areas) are all included in the datasets. We sample two times more negative examples (pixels not contained in burned areas), extracted from days when no fire occurred. Furthermore, we stratify the negative sampling to follow the land cover distribution of the positive examples (burned pixels). This stratification strategy makes the distinction between positives and negatives harder for the models, and prevents them from learning trivial mappings.

We do a time split using the years 2009-2018 for training, 2019 for validation and 2020 for testing. All the datasets consist of 15731 training (10380 non-fire, 5351 fire), 1337 validation (1075 non-fire, 262 fire), and 1286 testing (975 non-fire, 311 fire) examples.

4.2 Models and training

We train a different model for each type of dataset introduced above, as depicted in Figure 1.

For the pixel dataset, we consider a Random Forest (RF) baseline model, which has been extensively used in related work [10]. Its hyperparameters are chosen based on the validation Area Under the Receiver Operating Characteristic (AUROC). The final RF consists of 100 trees, with no max depth set, the minimum number of samples required to split is 2 and the minimum number of samples required to be at a leaf node is 1.

For the temporal, spatial and spatio-temporal datasets, typical large DL models quickly overfit, because our datasets are relatively small. Thus, we develop some custom models, whose architectures and hyperparameters have been chosen based on the validation loss. All linear layers, but the last, are followed by a dropout with probability $p = 0.5$ and the ReLU activation function. The data is standardized before serving as input. The DL models are trained for 50 epochs with the binary cross-entropy loss with ℓ_2-norm regularization, the Adam optimizer and a batch size of 128.

For the temporal dataset, we use a Long-Short Term Memory (LSTM) architecture that is able to capture temporal dynamics. An LSTM layer with 64 neurons is followed by two linear hidden layers with 64 and 32 neurons, respectively, and an output 2-class softmax layer. The model is trained with a 0.01 weight decay and 0.001 learning rate.

For the spatial dataset, we use a CNN architecture, known to effectively capture spatial features. A CNN layer with 16 filters with 3×3 kernels, padding of 1, and stride of 1 is followed by a 2×2 max-pooling layer. This is followed by a dropout and two linear layers with 16 and 8 neurons, respectively, before the final 2-class softmax layer. The model is trained with a 0.03 weight decay and 0.0004 learning rate.

For the spatio-temporal dataset, we use a ConvLSTM [16], which is known to capture spatio-temporal features of the input, combining the strengths of the LSTM and CNN components. A ConvLSTM layer with 16 filters with 3×3 kernels is used, followed by an architecture similar to the aforementioned CNN. The model is trained with a 0.03 weight decay and 0.0001 learning rate.

4.3 Results

The evaluation is performed over the test set, which contains all the fire events during the year 2020. In Table 1, we report the precision, recall and F_1 score for the positive class (burned), together with the AUROC.

While RF achieves the best precision, it has the lowest recall and F_1 score, together with the CNN, i.e. they both

Table 1: Performance of the models on the test set (year 2020)

Model	Precision	Recall	F_1	AUROC
RF	**0.832**	0.508	0.631	0.898
LSTM	0.741	**0.762**	**0.751**	0.920
CNN	0.732	0.553	0.63	0.910
ConvLSTM	0.798	0.646	0.714	**0.926**

miss fire events. The LSTM and ConvLSTM perform better regarding these metrics. However, since we want to output the fire likelihood and not just a hard binary prediction, we consider the AUROC to be the most important metric for the evaluation. In terms of AUROC, the RF achieves the weakest performance, while ConvLSTM achieves the best one, suggesting that the best performance can be achieved when both spatial and temporal contexts are combined. Moreover, whilst all DL models outperform RF, LSTM and ConvLSTM have better AUROC than CNN, showing that temporal context is more important than spatial context for wildfire danger forecasting.
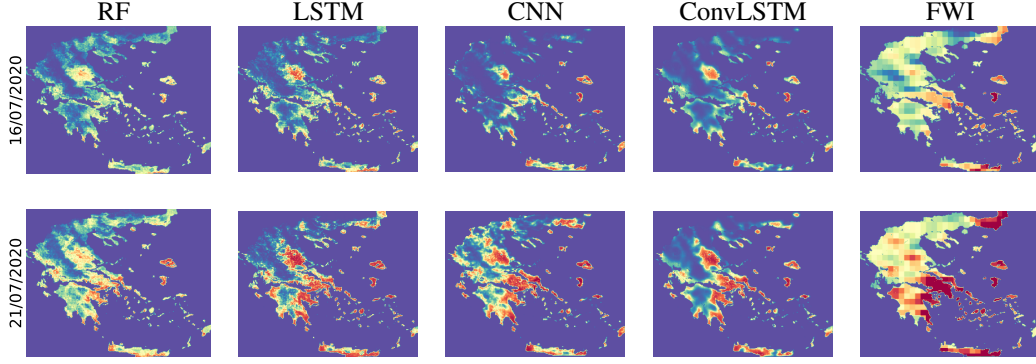
4

Figure 2: Maps produced by different models and the day's FWI for two dates with medium (top) and high (bottom) fire danger. Extreme danger is shown in dark red and no danger in blue.

However, a single metric like AUROC does not tell the whole story. In Figure 2 we visualize all models' predictions for two separate days in 2020. Next to them, we plot the FWI, which is commonly used operationally as an index related to fire danger and is provided at a $8 \text{ km} \times 8 \text{ km}$ spatial resolution. At a first glance, all the models seem to follow the same wildfire danger patterns, with some subtle differences, yet all of them provide finer details and more realistic susceptibility patterns compared to FWI. The RF map outputs generally lower probabilities than the other models. Lastly, CNN and ConvLSTM maps are visually smoother, as they exploit the spatial context which is overlapping in neighboring cells.

5   Discussion

The results on the reported metrics for daily fire labels (Table 1), demonstrate the DL models' potential to estimate wildfire danger. While, in terms of AUROC we see the advantage of DL models, when looking at the other metrics there is not a clear winner. However, the precision and recall values can be tuned by choosing a class weight for the positive class or by sampling more or less negative examples. Visually, the generated national-scale daily prediction maps (Figure 2) have structure, contain both large and local scale patterns, and show spatio-temporal variability.

Further research will scale to the greater fire-prone Mediterranean area. This will allow to test how well our models generalize to similar biomes. It will also add data, which can allow us to improve with deeper and more sophisticated DL architectures to account for richer temporal information to detect legacy effects, as well as long-range spatial connections.

While DL models are able to make predictions with high accuracy, their deterministic nature prevents them from providing uncertainty estimates for their predictions. Wildfire occurrence, however, presents a high degree of stochasticity, therefore uncertainty assessment is critical in decision making. Hence, Bayesian ML, DL [17] approaches offer a promising route in that direction.

Given that wildfires are a phenomenon with huge impact on human lives and the ecosystem, explaining can be more important than fitting. In order to increase trustworthiness of DL-based solutions and use them as operational tools against fire disasters, it is crucial to interpret and inspect our model's predictions through explainable AI [18]. This can shed light to the limitations of these models and help us refine our models to become more amenable, simpler and with accountable decisions.

6   Conclusion

In this work, we formulated daily fire danger forecasting as a machine learning problem and produced a harmonized country-wide datacube [11], from which we extract datasets in four modalities. We implemented some simple, yet effective prototype DL models exploiting the spatial and temporal context of the inputs, which outperformed a baseline RF classifier in terms of AUROC, suggesting that DL can be a valuable tool for wildfire forecasting.

Acknowledgments and Disclosure of Funding

References

[1] M. Lucrecia Pettinari and Emilio Chuvieco. Fire Danger Observed from Space. *Surveys in Geophysics*, 41(6):1437–1459, November 2020. ISSN 1573-0956. doi: 10.1007/s10712-020-09610-8. URL `https://doi.org/10.1007/s10712-020-09610-8`.

[2] Marco Turco, Juan José Rosa-Cánovas, Joaquín Bedia, Sonia Jerez, Juan Pedro Montávez, Maria Carmen Llasat, and Antonello Provenzale. Exacerbated fires in Mediterranean Europe due to anthropogenic warming projected with non-stationary climate-fire models. *Nature communications*, 9(1):1–9, 2018. Publisher: Nature Publishing Group.

[3] Jesús San-Miguel-Ayanz, Ernst Schulte, Guido Schmuck, and Andrea Camia. The european forest fire information system in the context of environmental policies of the european union. *Forest Policy and Economics*, 29:19–25, 2013. ISSN 1389-9341. doi: https://doi.org/10.1016/j.forpol.2011.08.012.

[4] Stijn Hantson, Almut Arneth, Sandy P. Harrison, Doug I. Kelley, I. Colin Prentice, Sam S. Rabin, Sally Archibald, Florent Mouillot, Steve R. Arnold, and Paulo Artaxo. The status and challenge of global fire modelling. *Biogeosciences*, 13(11):3359–3375, 2016. Publisher: Copernicus Publications.

[5] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, and Nuno Carvalhais. Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743):195–204, 2019. Publisher: Nature Publishing Group.

[6] Guoli Zhang, Ming Wang, and Kai Liu. Forest fire susceptibility modeling using a convolutional neural network for yunnan province of china. *International Journal of Disaster Risk Science*, 10(3):386–403, 2019. Publisher: Springer.

[7] Fantine Huot, R. Lily Hu, Matthias Ihme, Qing Wang, John Burge, Tianjian Lu, Jason Hickey, Yi-Fan Chen, and John Anderson. Deep Learning Models for Predicting Wildfires from Historical Remote-Sensing Data. *arXiv:2010.07445 [cs]*, October 2020. URL `http://arxiv.org/abs/2010.07445`. arXiv: 2010.07445.

[8] Louis Giglio, Wilfrid Schroeder, and Christopher O. Justice. The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sensing of Environment*, 178:31–41, 2016.

[9] Mariana Oliveira, Luís Torgo, and Vítor Santos Costa. Evaluation procedures for forecasting with spatiotemporal data. *Mathematics*, 9(6):691, 2021.

[10] Piyush Jain, Sean C.P. 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. doi: 10.1139/er-2020-0019. URL `https://doi.org/10.1139/er-2020-0019`.

[11] Ioannis Prapas, Spyros Kondylatos, and Ioannis Papoutsis. A datacube for the analysis of wildfires in greece, June 2021. URL `https://doi.org/10.5281/zenodo.4943354`.

[12] Joaquín Muñoz-Sabater, Emanuel Dutra, Anna Agustí-Panareda, Clément Albergel, Gabriele Arduini, Gianpaolo Balsamo, Souhail Boussetta, Margarita Choulga, Shaun Harrigan, Hans Hersbach, et al. Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data Discussions*, pages 1–50, 2021.

[13] György Büttner. Corine land cover and land cover change products. In *Land use and land cover mapping in Europe*, pages 55–74. Springer, 2014.

[14] A. Bashfield and A. Keim. Continent-wide DEM creation for the european union. In *34th International Symposium on Remote Sensing of Environment. The GEOSS Era: Towards Operational Environmental Monitoring. Sydney, Australia*, pages 10–15. Citeseer, 2011.

[15] A. A. Ager, H. K. Preisler, B. Arca, D. Spano, and M. Salis. Wildfire risk estimation in the mediterranean area. *Environmetrics*, 25(6):384–396, 2014. doi: https://doi.org/10.1002/env.2269. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2269`.

[16] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28:802–810, 2015.

[17] Laurent Valentin Jospin, Wray Buntine, Farid Boussaid, Hamid Laga, and Mohammed Bennamoun. Hands-on Bayesian Neural Networks – a Tutorial for Deep Learning Users. *arXiv:2007.06823 [cs, stat]*, July 2020. URL `http://arxiv.org/abs/2007.06823`. arXiv: 2007.06823.

[18] Wojciech Samek, Grégoire Montavon, Sebastian Lapuschkin, Christopher J Anders, and Klaus-Robert Müller. Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3):247–278, 2021.