AGU ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters®

Wildfire Danger Prediction and Understanding With Deep Learning

Spyros Kondylatos[1,2] ⓘ, **Ioannis Prapas[1,2]** ⓘ, **Michele Ronco[2]** ⓘ, **Ioannis Papoutsis[1]** ⓘ, **Gustau Camps-Valls[2]** ⓘ, **María Piles[2]** ⓘ, **Miguel-Ángel Fernández-Torres[2]** ⓘ, and **Nuno Carvalhais[3,4]** ⓘ

[1]Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Athens, Greece, [2]Image Processing Laboratory (IPL), Universitat de València, Valencia, Spain, [3]Max Planck Institute for Biogeochemistry, Jena, Germany, [4]Departamento de Ciências e Engenharia do Ambiente, DCEA, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, Caparica, Portugal

Abstract Climate change exacerbates the occurence of extreme droughts and heatwaves, increasing the frequency and intensity of large wildfires across the globe. Forecasting wildfire danger and uncovering the drivers behind fire events become central for understanding relevant climate-land surface feedback and aiding wildfire management. In this work, we leverage Deep Learning (DL) to predict the next day's wildfire danger in a fire-prone part of the Eastern Mediterranean and explainable Artificial Intelligence (xAI) to diagnose model attributions. We implement DL models that capture the temporal and spatio-temporal context, generalize well for extreme wildfires, and demonstrate improved performance over the traditional Fire Weather Index. Leveraging xAI, we identify the substantial contribution of wetness-related variables and unveil the temporal focus of the models. The variability of the contribution of the input variables across wildfire events hints into different wildfire mechanisms. The presented methodology paves the way to more robust, accurate, and trustworthy data-driven anticipation of wildfires.

Plain Language Summary Climate change drives the aggravation of wildfires globally. In this context, it is crucial to better predict and understand wildfires. Our work proposes methods to accurately forecast wildfire danger and hints into the mechanisms that drive it. We use Machine Learning (ML) to predict next-day wildfire danger, using meteorological, vegetation, and human-related data, improving the results of the widely used Fire Weather Index, which is based only on meteorological forecasts. We then look to understand what the models have learned and how the different conditioning factors contribute to the final prediction. We find that ML models focus more on soil moisture and relative humidity for their predictions. They also take into account the cumulative temperature and precipitation, as well as the last day's wind speed and the relative humidity in the 2 days preceding fire ignition. These associations lead to new, data-driven ways to anticipate wildfires.

1. Introduction

Wildfires are an integral component of terrestrial ecosystems (Bond & Keeley, 2005; Scott & Glasspool, 2006) that affect the carbon cycle and shape ecological development via disturbance and regeneration (Bowman et al., 2009; Pausas & Keeley, 2009; Pausas et al., 2017). Nonetheless, wildfires negatively affect humans, especially in the wildland-urban interface, inducing catastrophic effects on infrastructure, ecosystem services, and human lives (Burke et al., 2021; Pettinari & Chuvieco, 2020). Wildfire is increasingly influenced by climate change and humans (Pausas & Keeley, 2021), who modify the natural ecosystems and alter the normal fire cycle, notably in Mediterranean-type Climate Regions (Batllori et al., 2013; Moreira et al., 2020). With the climates of these regions expected to change significantly during the next century (Klausmeyer & Shaw, 2009), it is crucial to rethink wildfire adaptation and mitigation strategies. However, wildfire is intrinsically stochastic and the result of complex interactions between all the fire drivers operating at different spatial and temporal scales (Archibald et al., 2013; Hantson et al., 2016), namely climate, vegetation, and human activity (Archibald et al., 2009), thus it is extremely difficult to model. To mitigate this difficulty, many studies have focused on identifying the links between fire weather and wildfire activity (Abatzoglou et al., 2019; Bedia et al., 2015), following the assumption that weather has the most important control on fire.

Formal analysis: Spyros Kondylatos, Ioannis Prapas, Michele Ronco
Funding acquisition: Ioannis Papoutsis
Investigation: Spyros Kondylatos, Ioannis Prapas, Michele Ronco, Ioannis Papoutsis
Methodology: Spyros Kondylatos, Ioannis Prapas, Michele Ronco, Ioannis Papoutsis, Nuno Carvalhais
Project Administration: Ioannis Papoutsis
Resources: Spyros Kondylatos, Ioannis Prapas, Ioannis Papoutsis
Software: Spyros Kondylatos, Ioannis Prapas, Michele Ronco
Supervision: Ioannis Papoutsis, Gustau Camps-Valls, Nuno Carvalhais
Validation: Spyros Kondylatos, Ioannis Prapas, Michele Ronco
Visualization: Spyros Kondylatos, Ioannis Prapas, Michele Ronco
Writing – original draft: Spyros Kondylatos, Ioannis Prapas
Writing – review & editing: Spyros Kondylatos, Ioannis Prapas, Michele Ronco, Ioannis Papoutsis, Gustau Camps-Valls, María Piles, Miguel-Ángel Fernández-Torres, Nuno Carvalhais

The challenges of modeling fire amount to limitations in anticipation and quantification of the role of fire in the Earth system and hamper our short-term diagnostic and prognostic capabilities. Thus, in general, the next day's fire danger is forecasted by indices like the Fire Weather Index (FWI; Van Wagner, 1974), which relies solely on meteorological conditions and disregards the status of other fire drivers related to vegetation and human factors. We argue that it is crucial to consider and quantify the spatio-temporal contribution of all the fire drivers when predicting the next day's fire danger. Machine Learning (ML) methods are promising, allowing to model through the complexity in a data-driven way. The use of ML is facilitated by the increasing amount of data related to fire drivers, and especially remote sensing products (Forkel et al., 2017). Several studies have been conducted leveraging ML (Jain et al., 2020) in wildfire-related research. Particularly, Reichstein et al. (2019) and Camps-Valls et al. (2021) have suggested Deep Learning (DL) as a methodology well-suited to model the complex variable interactions which are ubiquitous in Earth system problems and especially wildfires. Actually, a limited but growing amount of works exploit DL for short-term fire danger forecasting. Le et al. (2021) use a Multilayer Perceptron for predicting wildfire danger, while Zhang et al. (2019, 2021), Bergado et al. (2021), and Bjånes et al. (2021) use Convolutional Neural Networks (Krizhevsky et al., 2012) to model forest fire susceptibility. Following a different approach, Huot et al. (2020) treat burn probability estimation as a segmentation task and use UNet-like architectures (Ronneberger et al., 2015) to address it.

In this work, we predict next day's fire danger leveraging DL for a part of the Eastern Mediterranean, centered around Greece, a representative fire-prone Mediterranean-type Climate Region. In addition, we use explainable Artificial Intelligence (xAI) to quantify variables' attribution, interpret models' predictions, and ultimately learn about fire mechanisms from data. To this end, we collect, harmonize, and share a data set containing a wide range of heterogeneous variables related to fire drivers, which include vegetation status, wetness conditions, anthropogenic factors and weather, as well as satellite-based burned areas for years from 2009 to 2021 in the region of interest. We train DL models capturing the temporal and spatio-temporal context, which are of pivotal importance in modeling wildfire dynamics. To test the generalization robustness, we evaluate the performance of the models for two different fire seasons in Greece: summer 2020, which fits the profile of a typical fire season and summer 2021, a season with several extreme wildfires (Giannaros et al., 2022). We further contrast the DL models with random forest (RF) and XGBoost as well as the FWI. In addition, we exploit xAI to uncover the main drivers that influence the occurrence and spread of wildfires and examine their temporal contribution decisions.

2. Data

We start with the curation of a 1 km × 1 km × 1 day resolution data cube (Prapas et al., 2022a) containing 90 variables related to the drivers of fire occurrence and spread (weather, vegetation, and human activity) (Hantson et al., 2016), together with historical burned areas for years 2009–2021. The study region is centered around Greece, also including parts of the Balkan peninsula and western Turkey, and covers a total area of 1,253 km × 983 km. To the best of our knowledge, this is the largest available analysis-ready database for next-day wildfire forecasting and it is made publicly available. Details on the data cube variables and harmonization process in Text S1 in Supporting Information S1. We use 25 of these variables as predictors:

1. *Daily weather data* from ERA-5 Land (Muñoz-Sabater et al., 2021) of maximum 2 m temperature, maximum wind speed, minimum relative humidity, total precipitation, maximum 2 m dewpoint temperature, and maximum surface pressure. The type of daily aggregations chosen represents the most fire-aggravating conditions.
2. *Satellite variables* from MODIS including Normalized Difference Vegetation Index (NDVI; Didan, 2015), day and night Land Surface Temperature (LST; Wan et al., 2015).
3. *Soil moisture index* from the European Drought Observatory (Cammalleri et al., 2017).
4. *Roads distance*, *waterway distance*, and *yearly population density* from WorldPop (Tatem, 2017).
5. *Elevation and Slope* from Copernicus EU-DEM (Bashfield & Keim, 2011).
6. Ten variables with the fraction of classes from Copernicus Corine Land Cover (Büttner, 2014).

The predictand is the historical burned areas from the European Forest Fire Information System (San-Miguel-Ayanz et al., 2013) containing burned areas larger than 30 hecta (ha). The burned areas are intersected with the MODIS active fires (Giglio et al., 2016) to recover the start date of the fire.

We exploit the data cube to investigate the distributions of the different variables. Figure 1a shows the probability density function of the input variables for grid cells that burned or not the next day (pairwise interactions between
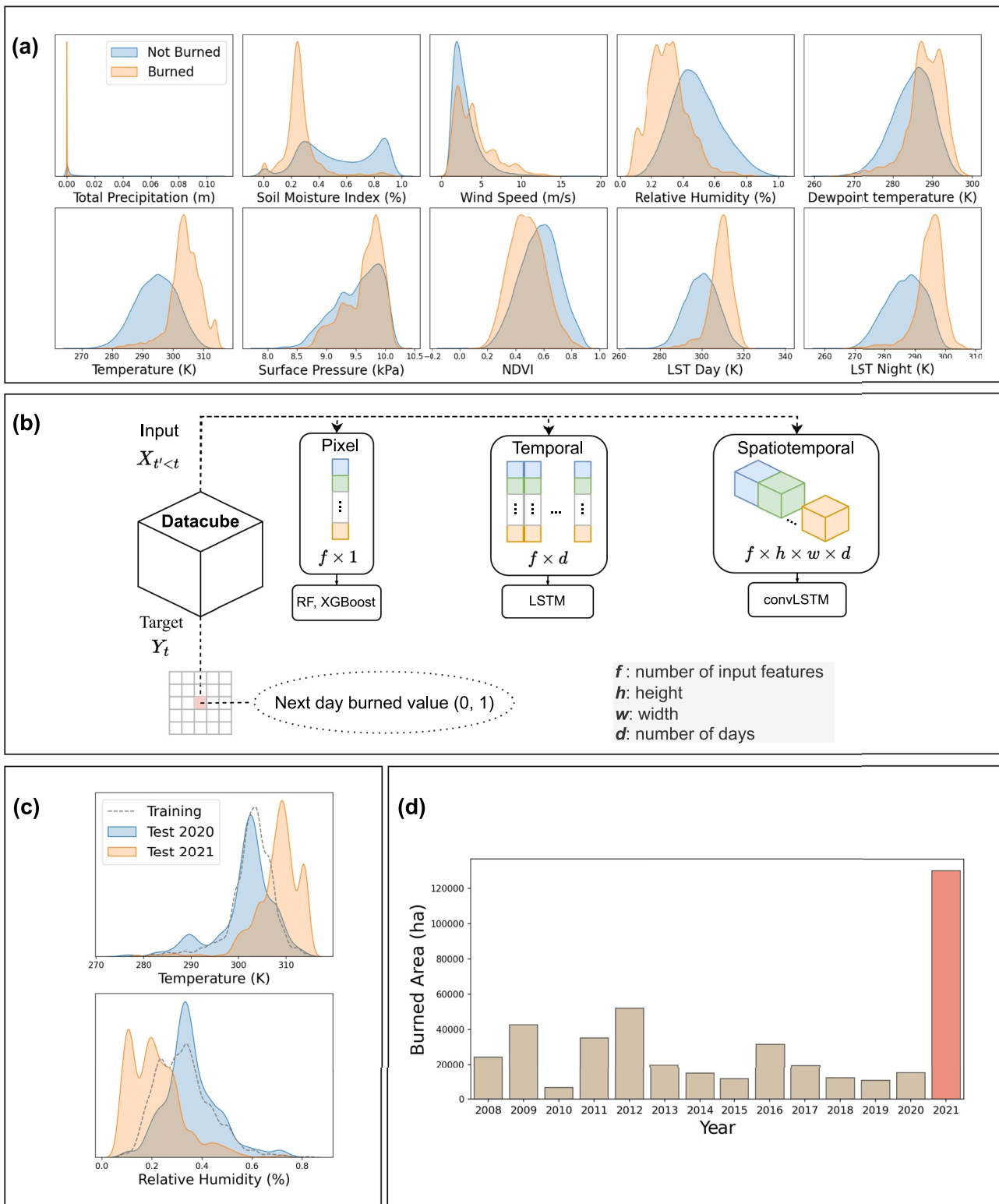
Figure 1. (a) The distribution of the values of the dynamic variables for burned and not burned cells. (b) Three data sets (pixel, temporal, and spatio-temporal) and the target are extracted from the data cube and used to train the corresponding models (random forest [RF], XGBoost, Long-Short Term Memory [LSTM], and Convolutional Long-Short Term Memory [convLSTM]). (c) The distribution of the values for temperature and relative humidity inside burned cells for the training set and the two test sets (2020 and 2021). (d) Yearly burned area in Greece.

the variables in Figure S1 in Supporting Information S1). Interestingly, there are clear shifts in the distributions of the input variables for burned cells, occurring at higher temperatures, lower precipitation, lower soil moisture, and lower relative humidity. The substantial intersection of the input variables for burned versus nonburned areas reflects the stochasticity of the fire ignition. Similar conditions may lead or not to burned areas. These results reveal the complexity of the mechanisms that drive wildfires and motivate the employment of ML approaches, given their capability to handle such complexity in a data-driven way.

3. Methods

Predicting fire danger with ML is not straightforward and poses some challenges (Prapas et al., 2021), which are important to address:

1. Wildfires are caused by complex interactions of fire drivers operating at different temporal and spatial scales and interacting mostly in nonlinear ways. (CH1)
2. Wildfire occurrence is intrinsically stochastic; lack of a fire event does not mean a lack of fire danger. (CH2)
3. Wildfire is a physical process that affects humans and the environment in a multitude of ways. It is crucial to go beyond mere forecasting into understanding what drives the models' predictions. (CH3)

We consider the fire danger definition by Pettinari and Chuvieco (2020), that is, the assessment of the conditions that allow a fire to ignite and spread. We assume that high fire danger conditions lead to large burned areas (>30 ha), while low fire danger conditions do not. From an ML perspective, for a given cell, we consider fire danger as the probability $y_t = p(E_t | x_{t' < t})$ of a fire occurring on day t and becoming large enough to result in a burned area that includes the cell (event E_t), conditioned on the observations $x_{t' < t}$ of the different fire drivers at time $t' < t$ (observations of the meteorological variables at time $t' < t$ are forecasts for the next day). This is the information needed for the management of firefighting resources, namely the number of fire personnel and the equipment (Ager et al., 2014; Preisler et al., 2004) and is motivated by the fact that a few large fires account for most of the burned areas (Cui & Perera, 2008). We formulate the problem as a binary classification problem, training a model F that aims to predict $y_t \in \{0, 1\}$ with input $x_{t' < t}$. We train F in a supervised way, extracting input-target pairs $(x_{t' < t}, y_t)$ and minimizing the binary cross entropy loss function.

CH1 makes the selection of input $x_{t' < t}$ difficult. We extract data sets in three different modalities (instance-based, temporal, and spatio-temporal) to serve as input to appropriate data-driven models, which are trained to estimate y_t (see Figure 1b). First, we extract the *instance-based data set*, consisting of feature vectors which contain the dynamic input observations at day $t − 1$, the average of the dynamic inputs for days $t − 1$, $t − 2$, …, $t − 10$ and the static features. These become input to two different tree-based models (RF (Breiman, 2001) and XGBoost (Chen & Guestrin, 2016)), considered among the best algorithms for tabular data. Second, we extract the *temporal data set*, consisting of the time series of days $t − 1$, $t − 2$, …, $t − 10$ of the dynamic input observations, extending the 7 days memory buffer (Huot et al., 2020) by 3 days, and the static features, which are repeated in time. These serve as input to a Long-Short Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) that exploits the temporal context. Finally, we extract the *spatio-temporal data set*, consisting of 25 km × 25 km (Bjånes et al., 2021) × 10 days blocks of the dynamic input observations centered spatially around the given cell for days $t − 1$, $t − 2$, …, $t − 10$ and 25 km × 25 km patches of the static features, which are repeated in time. These serve as input to a Convolutional Long-Short Term Memory (ConvLSTM; Shi et al., 2015) that exploits the spatio-temporal context.

The targets y_t are the same for all data sets modalities; positive examples ($y_t = 1$) consist of the final burned pixels from any fire that started on day t. These can be considered as samples from the random process that generates large fires and, as such, a proxy for fire danger of this cell at day t. CH2 makes it hard to select examples from the negative class ($y_t = 0$). To decrease the risk of sampling negatives that in fact represent high fire danger cells, we select negatives from days when no fire occurred in the entire region of interest. Moreover, as positives are limited, we follow the strategy in (Huot et al., 2020) and we sample two times more negatives, to increase the number of training data without incurring a highly imbalanced data set. Finally, the negative sampling is stratified by the land cover distribution of the positive examples to prevent the models from learning trivial mappings. The code is written using the PyTorch library (Paszke et al., 2019). More details about the architectures and training are presented in Text S2 in Supporting Information S1.

Table 1
Performance of the Models in the Test Years 2020 and 2021

Model	TP(↑)	FP(↓)	TN(↑)	FN(↓)	Precision	Recall	F_1
	(a) Results 2020						
RF	740	138	2,318	488	0.843	0.603	0.703
XGBoost	888	154	2,302	340	0.852	0.723	0.782
LSTM	**927**	150	2,306	**301**	0.861	**0.755**	0.804
ConvLSTM	879	**73**	**2,383**	349	**0.923**	0.716	**0.806**
	(b) Results 2021						
RF	3,073	418	8,396	1,334	0.880	0.697	0.778
XGBoost	3,172	488	8,326	1,235	0.867	0.720	0.786
LSTM	**3,769**	402	8,412	**638**	0.904	**0.855**	**0.879**
ConvLSTM	3,543	**186**	**8,628**	864	**0.950**	0.804	0.871

Note. True Positives (TP), False Positives (FP), True Negatives (TN), False Negatives (FN), Precision, Recall, and F1-score are calculated. Details about the different metrics are provided in Text S3 in Supporting Information S1. For each metric and year, the score of the best performing model is highlighted in bold.

Dealing with forecasting (Oliveira et al., 2021), we do a temporal split for the evaluation. We use years 2009–2018 for training and 2019 for validation. For testing, we use two sets, one for year 2020, representing a typical fire season and one for year 2021, representing an extreme one (Giannaros et al., 2022). This is supported both by the unprecedented values of temperature and relative humidity (Figure 1c) and by the overwhelming difference in total burned area for 2021 (Figure 1d). Eventually, all three data sets consist of 40,554 training (27,036 nonfire, 13,518 fire), 3,900 validation (2,600 nonfire, 1,300 fire), 3,684 testing (2,456 nonfire, 1,228 fire) samples for 2020, and 13,221 testing (8,814 nonfire, 4,407 fire) samples for 2021. Precision, Recall, and F1-score are used as evaluation metrics, common for measuring models' performance in imbalanced learning scenarios. Moreover, the Receiver Operating Characteristics (ROC) curve and Area Under the ROC curve (AUROC) are used to compare the predictive skill of the models against the FWI. Details about the metrics in Text S3 in Supporting Information S1.

The quantitative evaluation (Sections 4.1, 4.2) is enhanced with xAI that interprets the model's predictions (Section 4.3) and addresses CH3. Specifically, we exploit Shapley values (SHAP) (Lundberg & Lee, 2017), Partial Dependency Plots (PDPs; Molnar et al., 2020), and Integrated Gradients (IGs; Sundararajan et al., 2017) to better understand the drivers behind models' predictions and their dynamics for modeling wildfire danger. The methods were implemented with Python captum (Kokhlikyan et al., 2020). Details about the xAI methods in Text S4 in Supporting Information S1.

4. Results and Discussion

4.1. Evaluation of Machine Learning Methods

The evaluation of the ML methods used is performed over two distinct test sets, years 2020 and 2021 (Table 1). The DL models (LSTM and ConvLSTM) perform well with an F1-score greater than 0.8 and better than RF and XGBoost according to all the metrics. ConvLSTM provides less false positives and higher precision, which could be attributed to the holistic view that the spatial context grants. Contrary, the LSTM achieves the lowest number of false negatives and highest recall. Concerning F1-score, the difference between the LSTM and ConvLSTM is not evident, revealing that the temporal context is at a high degree sufficient at the scale and spatial resolution we are investigating.

The results are better for 2021 than for 2020 for all the models. Precision is higher, probably because negatives are more distinguishable. We sample negatives on days when no fire occurred, but few days did not have fires in the summer of 2021. Thus, more negatives are sampled outside the summer, which might make their classification easier. Nevertheless, the fact that the recall is higher points to a good generalization of the models to extreme conditions. Although intuitive from a physics perspective to have high fire danger when there is a high temperature and low relative humidity, it is not obvious for ML models, especially since such conditions were not encountered during training (Figure 1c). We reveal that the models have learned physically plausible associations in Section 4.3.

The evaluation is accompanied by a qualitative inspection of the danger maps produced by the LSTM model. Figure 2a illustrates the spatio-temporal variability of the predictions for 10 consecutive days of 2020. The results indicate the capacity of the model to identify fire danger and delineate it. Figure 2b shows the 3-day temporal evolution of the predictions until 3 August 2021, when *three extreme wildfires ignited. Just these three wildfires burned 70,908 ha, which is 1.2× times the annual average for the years 2008–2020* (Giannaros et al., 2022). Remarkably, the model's prediction on the day of the fire was significantly higher than in previous days, highlighting the timeliness of the LSTM in identifying the conditions that lead to major fire events. More prediction maps in Figure S1 in Supporting Information S1.
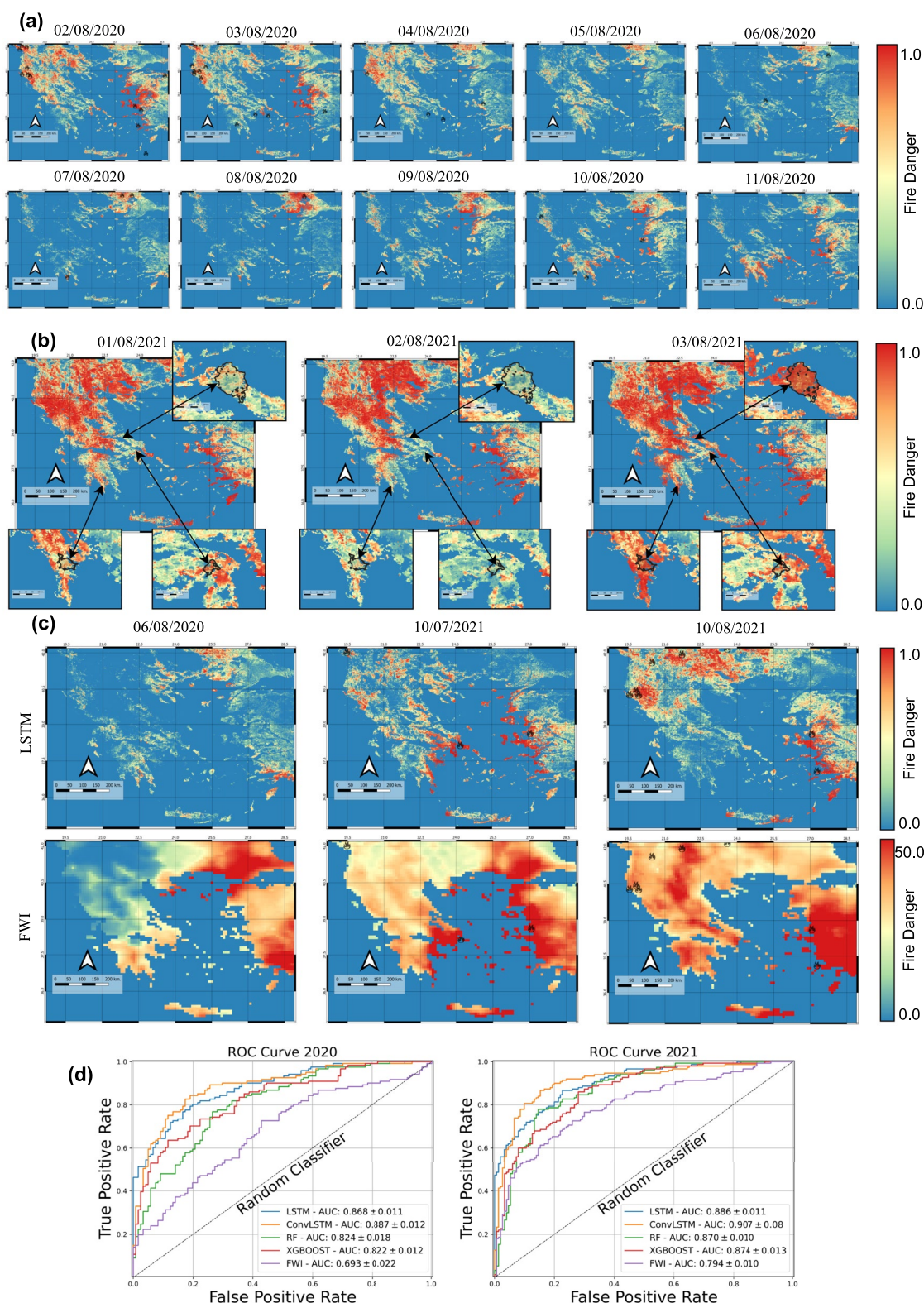
Figure 2. (a) Fire danger maps for 10 consecutive days of 2020 produced by the Long-Short Term Memory (LSTM). Black fire symbols represent the fire ignitions at that day. (b) Temporal evolution of fire danger for three consecutive days in 2021, leading to the ignition of three major fire events in Greece, in 3 August 2021. The black line in the zoomed maps designates the final burnt area of each event. (c) Fire danger maps produced by the LSTM and Fire Weather Index (FWI) for 3 days. Black fire symbols represent the fire ignitions at that day. (d) Receiver Operating Characteristics (ROC) curves of FWI, random forest (RF), XGBoost, LSTM, Convolutional Long-Short Term Memory (ConvLSTM), presented along with the AUROC mean and standard deviation from the 10 different experiments.

4.2. Comparison With FWI

Although the performance of the DL models is good, a comparison with FWI (Text S5 in Supporting Information S1), a model that is used in practice for wildfire management, is critical for enhancing their reliability. We compare the two methods for the two test years, 2020 and 2021.

First, we use open-source code (Steinfeld, 2022) to calculate the FWI in a spatial resolution of 9 km × 9 km using the meteorological variables from ERA-5 Land. Then, we aggregate the predictions of the ML models to the same spatial resolution. This makes the results consistent in terms of spatial resolution to better compare the predictive skill of the methods. For every fire event, we compute the models' predictions and FWI value at the 9 km × 9 km cell closest to the ignition point of the fire. We sample as many negatives as fire events from days when there was no fire and evaluate the performance of the different methods using ROC curves and AUROC (Figure 2d). We repeat this procedure 10 times and report the average and standard deviation of the AUROC.

The results (Figure 2d) demonstrate that ML models outperform FWI for both years. Moreover, all models have better scores in 2021 (similar to Section 4.1). The inclusion of EO data, vegetation indices, and human-related data contribute to a better forecasting of fire danger, as even RF and XGBoost, that cannot handle the temporal context, outperform FWI in terms of AUROC. DL models that additionally handle the temporal and spatio-temporal context perform best. This shows that it is important to model the nonlinear interactions and temporal evolution of the fire drivers.

Figure 2c shows the maps produced by the LSTM and FWI for some selected days in 2020 and 2021, along with the positions of fires that ignited these days. LSTM outputs range in [0 and 1]. FWI outputs are clipped to 50, as higher values are considered as extreme fire danger (San-Miguel-Ayanz et al., 2013). Visually, LSTM provides much finer scale details than FWI. Moreover, FWI overestimates fire danger in many cases, as contrary to the LSTM, FWI does not consider variables related to fire-proneness (e.g., land cover and vegetation).

This comparison against FWI favors the ML models but does not call for its replacement. FWI is an indicator of weather-based fire behavior potential and does not consider burned area, which is intrinsic to the definition of our models and evaluation. FWI is also tested in time, meaningful, and interpretable at its full value range. Thus, in order to increase the trustworthiness of the data-driven models, we perform a comprehensive xAI analysis for the LSTM in the next section.

4.3. What Drives Model's Predictions?

First, we identify, among all the inputs, which variables drive the predictions of fire danger for the burned pixels in both test sets. To this end, we use the SHAP to compute the marginal contribution of each feature. Figure 3a shows the SHAP of the 15 most important covariates in descending order. Positive (negative) SHAP means a contribution to higher (lower) fire danger. Remarkably, among the most important drivers we get soil and wetness indicators (soil moisture index and relative humidity), temperature variables (dewpoint temperature, LST Day, and temperature), and proxies for fuel moisture (NDVI) and fire spread (wind speed).

Figure 3b shows the PDPs for some of the key fire drivers (other examples can be found in Figures S2 in Supporting Information S1). Both the average and the standard deviation reveal information on the learned wildfire mechanisms. Particularly, the bell-shape pattern in NDVI reflects low fire danger under low NDVI (sparse or scarce vegetation conditions) but also under high NDVI, which could translate into wet/healthy vegetation operating as a fire spread deterrent but could also translate into a local decoupling between large weather patterns and micro-meteorological conditions. The PDPs of soil moisture and relative humidity enhance the argument that vegetation productivity and aridity are important factors for wildfire occurrence. Low values of these variables cause fuel dryness and, thus, increase the probability of having large fires (Figure 3b). Very warm weather conditions (i.e., LST Day and Night) and high winds also lead to higher danger. Wind speed, in turn, has a nonlinear trend and high standard deviation, which can be explained by the fact that wind speed is associated with fire spread but is not a sufficient nor necessary condition for a large fire.

It is interesting to study how LSTM is exploiting the time behavior of the covariates. Figure 3c shows the average IGs per day for some of the most relevant variables (see Figures S3 in Supporting Information S1 for additional variables) that highlight two dynamic distinct behaviors: accumulation effects and short-time interactions. Reasonably, relative humidity and max wind speed become important predictors only in the last days before the
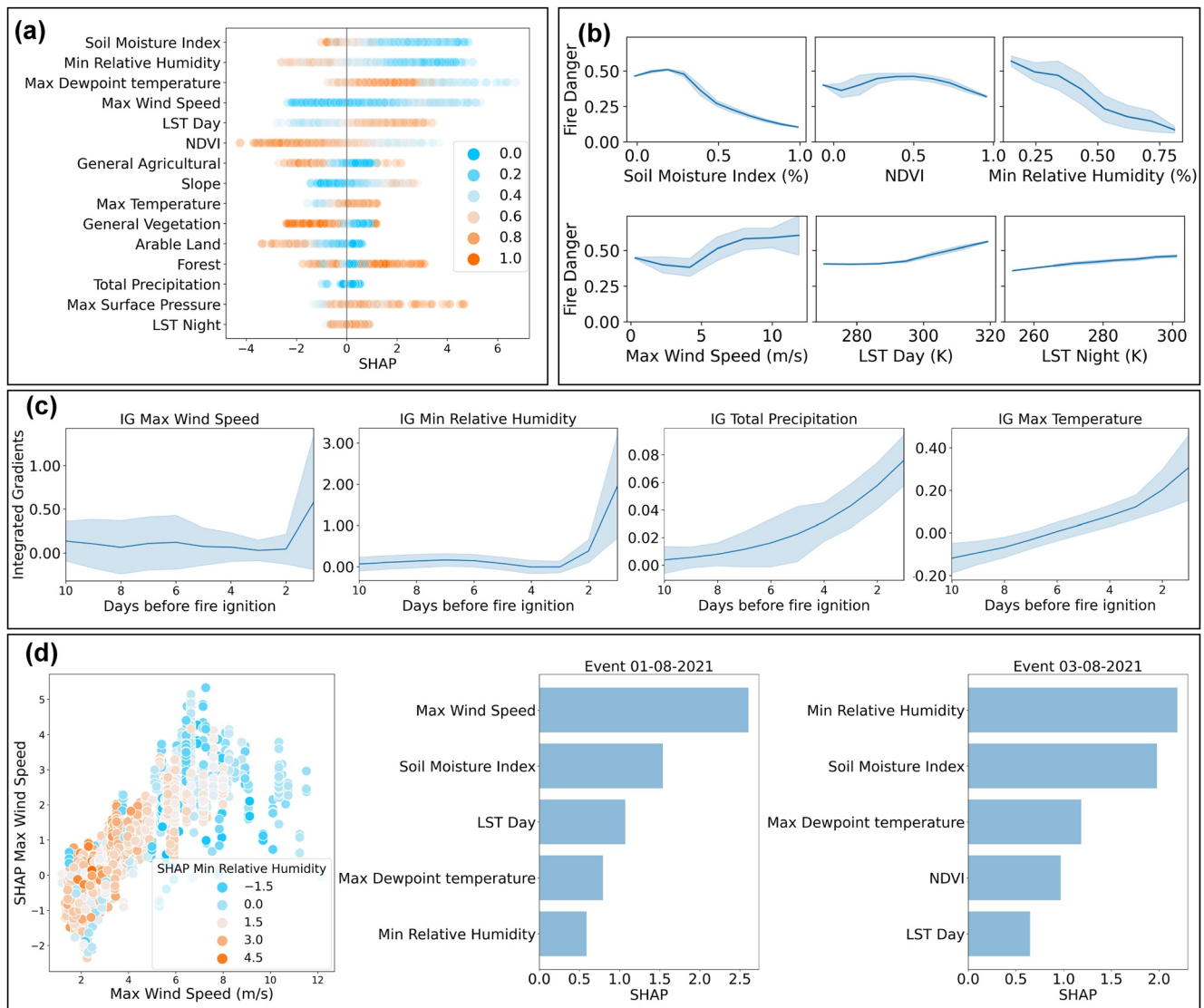
Figure 3. (a) Shapley values (SHAP) for the 15 most important variables (importance computed based on the absolute SHAP value of all the fire drivers). Dots represent positive examples in the test set and are colored based on the scaled value of the relevant variable. *Sorted by mean absolute SHAP value*. (b) Partial Dependency Plot line plots (mean and standard deviation) for six variables considering all samples in the test sets. Values on the *x* axis are averaged over time and reported in the original units. (c) The 10-day evolution of Integrated Gradients (IGs; mean and standard deviation) over all burned pixels for four of the variables. (d) Left figure: *x*-axis contains the values of wind speed, the *y*-axis contains the SHAP of the maximum wind speed for all the burned pixels in the test sets. The dots are colored based on the SHAP of the minimum relative humidity. Middle and right plots: five largest SHAP values for two wildfire events in 2021.

fire event starts. However, the effect of air temperature and total precipitation on fire danger builds up over the whole 10-day time series. This demonstrates the importance of the increasing fire-prone conditions (dryness and hotness) when predicting fire danger. The impact of using longer time series, beyond the 10-day synoptic scale adopted in this work, requires further exploration.

Major fire events are commonly driven by either low relative humidity or high wind speed (Ruffault et al., 2020). We demonstrate the ability of LSTM to identify these two clusters of wildfires, in a scatter plot showing the interactions between these two variables in Figure 3d. This is also confirmed by two extreme fires in 2021, that each one of them was mainly driven by one of these two different mechanisms (Figure 3d).

To sum up, the use of xAI has unveiled the capability of the LSTM to learn physical interactions in accordance with our physical knowledge, preparing the ground toward a more trustworthy data-driven wildfire forecasting.

5. Conclusion

Predicting wildfire danger is an especially challenging problem, mostly due to the stochastic nature of wildfires from ignition to spatial propagation, and the nonlinear interactions of the fire drivers. In this work, we address it using DL models, which capture the different temporal and spatio-temporal associations of environmental, meteorological, and human-related drivers in a data-driven way. The models demonstrate great predictive skill, superior to FWI. Notably, they also generalize well, even for the 2021 extreme fire season, hinting on the capacity of DL to learn plausible physical associations directly from data.

This notion is further enhanced with xAI methods, which reveal what the models have learned. The relationships emerging from the DL models highlight the importance of fuel-related variables (NDVI, soil moisture, and relative humidity), which should therefore be used in combination with meteorological drivers to assess the likelihood of large fire events. In addition, the xAI analysis unveils the ability of the models to distinguish between dryness-driven and wind-driven wildfires. It also shows that variables' temporal scales are critical for predicting wildfire danger, ranging from the same day wind speed, the relative humidity conditions for the two preceding days, and the cumulative 10-day temperature and precipitation.

Concluding, in parallel to the large performance improvement that DL models establish in wildfire danger forecasting, the physical rationale behind the drivers of the predictions emphasizes the role of data-driven learning toward a reliable and trustworthy approach for predicting and understanding wildfire danger and its dynamics.

Data Availability Statement

The data cube is in Zenodo (Prapas et al., 2022a, https://doi.org/10.5281/zenodo.6475592). The data sets are also in Zenodo (Prapas et al., 2022b, https://doi.org/10.5281/zenodo.6528394). Both are under Creative Commons Attribution 4.0 International Public License. The code and instructions for training the models are in Zenodo (Prapas & Kondylatos, 2022, https://doi.org/10.5281/zenodo.6524771), under MIT license.

References

Abatzoglou, J. T., Williams, A. P., & Barbero, R. (2019). Global emergence of anthropogenic climate change in fire weather indices. *Geophysical Research Letters*, *46*(1), 326–336. https://doi.org/10.1029/2018GL080959

Ager, A. A., Preisler, H. K., Arca, B., Spano, D., & Salis, M. (2014). Wildfire risk estimation in the Mediterranean area. *Environmetrics*, *25*(6), 384–396. https://doi.org/10.1002/env.2269

Archibald, S., Lehmann, C. E. R., Gómez-Dans, J. L., & Bradstock, R. A. (2013). Defining pyromes and global syndromes of fire regimes. *Proceedings of the National Academy of Sciences*, *110*(16), 6442–6447. https://doi.org/10.1073/pnas.1211466110

Archibald, S., Roy, D. P., van Wilgen, B. W., & Scholes, R. J. (2009). What limits fire? An examination of drivers of burnt area in Southern Africa. *Global Change Biology*, *15*(3), 613–630. https://doi.org/10.1111/j.1365-2486.2008.01754.x

Bashfield, A., & Keim, A. (2011). Continent-wide DEM creation for the European Union. In *34th International Symposium on Remote Sensing of Environment – The GEOSS Era: Towards Operational Environmental Monitoring* (pp. 10–15). Citeseer.

Batllori, E., Parisien, M.-A., Krawchuk, M. A., & Moritz, M. A. (2013). Climate change-induced shifts in fire for Mediterranean ecosystems. *Global Ecology and Biogeography*, *22*(10), 1118–1129. https://doi.org/10.1111/geb.12065

Bedia, J., Herrera, S., Gutiérrez, J. M., Benali, A., Brands, S., Mota, B., & Moreno, J. M. (2015). Global patterns in the sensitivity of burned area to fire-weather: Implications for climate change. *Agricultural and Forest Meteorology*, *214–215*, 369–379. https://doi.org/10.1016/j.agrformet.2015.09.002

Bergado, J. R., Persello, C., Reinke, K., & Stein, A. (2021). Predicting wildfire burns from big geodata using deep learning. *Safety Science*, *140*, 105276. https://doi.org/10.1016/j.ssci.2021.105276

Bjånes, A., De La Fuente, R., & Mena, P. (2021). A deep learning ensemble model for wildfire susceptibility mapping. *Ecological Informatics*, *65*, 101397. https://doi.org/10.1016/j.ecoinf.2021.101397

Bond, W. J., & Keeley, J. E. (2005). Fire as a global 'herbivore': The ecology and evolution of flammable ecosystems. *Trends in Ecology & Evolution*, *20*(7), 387–394. https://doi.org/10.1016/j.tree.2005.04.025

Bowman, D. M. J. S., Balch, J. K., Artaxo, P., Bond, W. J., Carlson, J. M., Cochrane, M. A., et al. (2009). Fire in the Earth system. *Science*, *324*(5926), 481–484. https://doi.org/10.1126/science.1163886

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Burke, M., Driscoll, A., Heft-Neal, S., Xue, J., Burney, J., & Wara, M. (2021). The changing risk and burden of wildfire in the United States. *Proceedings of the National Academy of Sciences*, *118*(2), e2011048118. https://doi.org/10.1073/pnas.2011048118

Büttner, G. (2014). CORINE land cover and land cover change products. In *Land use and land cover mapping in Europe* (pp. 55–74). Springer.

Cammalleri, C., Vogt, J. V., Bisselink, B., & de Roo, A. (2017). Comparing soil moisture anomalies from multiple independent sources over different regions across the globe. *Hydrology and Earth System Sciences*, *21*(12), 6329–6343. https://doi.org/10.5194/hess-21-6329-2017

Camps-Valls, G., Tuia, D., Zhu, X. X., & Reichstein, M. (2021). *Deep learning for the Earth sciences: A comprehensive approach to remote sensing, climate science and geosciences*. John Wiley & Sons. Retrieved from https://github.com/DL4ES

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Cui, W., & Perera, A. H. (2008). What do we know about forest fire size distribution, and why is this knowledge useful for forest management? *International Journal of Wildland Fire*, *17*(2), 234–244. https://doi.org/10.1071/wf06145

Didan, K. (2015). *MOD13A2 MODIS/terra vegetation indices 16-day L3 global 1 km SIN grid V006* Distributed by NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD13A2.006 (Sparse or scarse vegetation conditions).

Forkel, M., Dorigo, W., Lasslop, G., Teubner, I., Chuvieco, E., & Thonicke, K. (2017). A data-driven approach to identify controls on global fire activity from satellite and climate observations (SOFIA V1). *Geoscientific Model Development*, *10*(12), 4443–4476. https://doi.org/10.5194/gmd-10-4443-2017

Giannaros, T. M., Papavasileiou, G., Lagouvardos, K., Kotroni, V., Dafis, S., Karagiannidis, A., & Dragozi, E. (2022). Meteorological analysis of the 2021 extreme wildfires in Greece: Lessons learned and implications for early warning of the potential for pyroconvection. *Atmosphere*, *13*(3), 475. https://doi.org/10.3390/atmos13030475

Giglio, L., Schroeder, W., & Justice, C. O. (2016). The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sensing of Environment*, *178*, 31–41. https://doi.org/10.1016/j.rse.2016.02.054

Hantson, S., Arneth, A., Harrison, S. P., Kelley, D. I., Prentice, I. C., Rabin, S. S., et al. (2016). The status and challenge of global fire modelling. *Biogeosciences*, *13*(11), 3359–3375. https://doi.org/10.5194/bg-13-3359-2016

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huot, F., Hu, R. L., Ihme, M., Wang, Q., Burge, J., Lu, T., et al. (2020). Deep learning models for predicting wildfires from historical remote-sensing data. *arXiv:2010.07445 [cs]*. Retrieved from http://arxiv.org/abs/2010.07445

Jain, P., Coogan, S. C., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. *Environmental Reviews*, *28*(4), 478–505. https://doi.org/10.1139/er-2020-0019

Klausmeyer, K. R., & Shaw, M. R. (2009). Climate change, habitat loss, protected areas and the climate adaptation potential of Species in Mediterranean ecosystems worldwide. *PLoS One*, *4*(7), e6392. https://doi.org/10.1371/journal.pone.0006392

Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Alsallakh, B., Reynolds, J., et al. (2020). Captum: A unified and generic model interpretability library for PyTorch. *arXiv*. Retrieved from http://arxiv.org/abs/2009.07896

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (Vol. 25). Curran Associates, Inc. Retrieved from https://papers.nips.cc/paper/2012/hash/c399862d3b9d6b-76c8436e924a68c45b-Abstract.html

Le, H. V., Hoang, D. A., Tran, C. T., Nguyen, P. Q., Tran, V. H. T., Hoang, N. D., et al. (2021). A new approach of deep neural computing for spatial prediction of wildfire danger at tropical climate areas. *Ecological Informatics*, *63*, 101300. https://doi.org/10.1016/j.ecoinf.2021.101300

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 4768–4777).

Molnar, C., Casalicchio, G., & Bischl, B. (2020). Interpretable machine learning – A brief history, state-of-the-art and challenges. *arXiv preprint arXiv:2010.09337*, *1323*, 417–431. https://doi.org/10.1007/978-3-030-65965-3_28

Moreira, F., Ascoli, D., Safford, H., Adams, M. A., Moreno, J. M., Pereira, J. M. C., et al. (2020). Wildfire management in Mediterranean-type regions: Paradigm change needed. *Environmental Research Letters*, *15*(1), 011001. https://doi.org/10.1088/1748-9326/ab541e

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data Discussions*, *13*(9), 1–50. https://doi.org/10.5194/essd-13-4349-2021

Oliveira, M., Torgo, L., & Santos Costa, V. (2021). Evaluation procedures for forecasting with spatiotemporal data. *Mathematics*, *9*(6), 691. https://doi.org/10.3390/math9060691

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 32, pp. 8024–8035). Curran Associates, Inc. Retrieved from http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Pausas, J. G., & Keeley, J. E. (2009). A burning story: The role of fire in the history of life. *BioScience*, *59*(7), 593–601. https://doi.org/10.1525/bio.2009.59.7.10

Pausas, J. G., & Keeley, J. E. (2021). Wildfires and global change. *Frontiers in Ecology and the Environment*, *19*(7), 387–395. https://doi.org/10.1002/fee.2359

Pausas, J. G., Keeley, J. E., & Schwilk, D. W. (2017). Flammability as an ecological and evolutionary driver. *Journal of Ecology*, *105*(2), 289–297. https://doi.org/10.1111/1365-2745.12691

Pettinari, M. L., & Chuvieco, E. (2020). Fire danger observed from space. *Surveys in Geophysics*, *41*(6), 1437–1459. https://doi.org/10.1007/s10712-020-09610-8

Prapas, I., & Kondylatos, S. (2022). *Orion-AI-Lab/wildfire_forecasting: Initial code release*. Zenodo. https://doi.org/10.5281/zenodo.6524771

Prapas, I., Kondylatos, S., & Papoutsis, I. (2022a). *FireCube: A daily datacube for the modeling and analysis of wildfires in Greece*. Zenodo. https://doi.org/10.5281/zenodo.6475592

Prapas, I., Kondylatos, S., & Papoutsis, I. (2022b). *Training data for submitted paper "wildfire danger prediction and understanding with Deep learning"*. Zenodo. https://doi.org/10.5281/zenodo.6528394

Prapas, I., Kondylatos, S., Papoutsis, I., Camps-Valls, G., Ronco, M., Fernandez-Torres, M.-A., et al. (2021). Deep learning methods for daily wildfire danger forecasting. *arXiv:2111.02736 [cs]*. Retrieved from http://arxiv.org/abs/2111.02736

Preisler, H. K., Brillinger, D. R., Burgan, R. E., & Benoit, J. (2004). Probability based models for estimation of wildfire risk. *International Journal of Wildland Fire*, *13*(2), 133–142. https://doi.org/10.1071/wf02061

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical image computing and computer-assisted intervention – MICCAI 2015* (pp. 234–241). Springer International Publishing.

Ruffault, J., Curt, T., Moron, V., Trigo, R. M., Mouillot, F., Koutsias, N., et al. (2020). Increased likelihood of heat-induced large wildfires in the Mediterranean Basin. *Scientific Reports*, *10*(1), 13790. https://doi.org/10.1038/s41598-020-70069-z

San-Miguel-Ayanz, J., Schulte, E., Schmuck, G., & Camia, A. (2013). The European forest fire information system in the context of environmental policies of the European Union. *Forest Policy and Economics*, *29*, 19–25. https://doi.org/10.1016/j.forpol.2011.08.012

Scott, A. C., & Glasspool, I. J. (2006). The diversification of Paleozoic fire systems and fluctuations in atmospheric oxygen concentration. *Proceedings of the National Academy of Sciences*, *103*(29), 10861–10865. https://doi.org/10.1073/pnas.0604090103

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W.-c. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, *28*, 802–810.

Steinfeld, D. (2022). Calculation of indices for forest fire risk assessment in weather and climate data. Retrieved from https://github.com/steidani/FireDanger

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 3319–3328).

Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data*, *4*(1), 1–4. https://doi.org/10.1038/sdata.2017.4

Van Wagner, C. E. (1974). *Structure of the Canadian forest fire weather index* (Vol. 1333). Environment Canada, Canadian Forestry Service.

Wan, Z., Hook, S., & Hulley, G. (2015). *MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006*. Distributed by NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD11A1.006

Zhang, G., Wang, M., & Liu, K. (2019). Forest fire susceptibility modeling using a convolutional neural network for Yunnan province of China. *International Journal of Disaster Risk Science*, *10*(3), 386–403. https://doi.org/10.1007/s13753-019-00233-1

Zhang, G., Wang, M., & Liu, K. (2021). Deep neural networks for global wildfire susceptibility modelling. *Ecological Indicators*, *127*, 107735. https://doi.org/10.1016/j.ecolind.2021.107735