

PAPER • OPEN ACCESS

## Developing novel machine-learning-based fire weather indices

To cite this article: Assaf Shmuel and Eyal Heifetz 2023 *Mach. Learn.: Sci. Technol.* 4 015029

View the [article online](#) for updates and enhancements.

### You may also like

- [Anthropogenic climate change contribution to wildfire-prone weather conditions in the Cerrado and Arc of deforestation](#)  
Sihan Li, Sarah N Sparrow, Friederike E L Otto et al.
- [Climate change is increasing the likelihood of extreme autumn wildfire conditions across California](#)  
Michael Goss, Daniel L Swain, John T Abatzoglou et al.
- [Enhancing the fire weather index with atmospheric instability information](#)  
Miguel M Pinto, Carlos C DaCamara, Alexandra Hurduc et al.



## PAPER

## Developing novel machine-learning-based fire weather indices

## OPEN ACCESS

Assaf Shmuel\*  and Eyal Heifetz 

Department of Geophysics, Porter School of the Environment and Earth Sciences, Tel Aviv University, Tel Aviv 69978, Israel

\* Author to whom any correspondence should be addressed.

E-mail: [assafshmuel91@gmail.com](mailto:assafshmuel91@gmail.com)**Keywords:** machine learning, fire weather indices, forest management, wildfire risk

## RECEIVED

2 September 2022

## REVISED

14 February 2023

## ACCEPTED FOR PUBLICATION

28 February 2023

## PUBLISHED

13 March 2023

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

**Abstract**

Accurate wildfire risk estimation is an essential yet challenging task. As the frequency of extreme fire weather and wildfires is on the rise, forest managers and firefighters require accurate wildfire risk estimations to successfully implement forest management and firefighting strategies. Wildfire risk depends on non-linear interactions between multiple factors; therefore, the performance of linear models in its estimation is limited. To date, several traditional fire weather indices (FWIs) have been commonly used by weather services, such as the Canadian FWI. Traditional FWIs are primarily based on empirical and statistical analyses. In this paper, we propose a novel FWI that was developed using machine learning—the machine learning based fire weather index (MLFWI). We present the performance of the MLFWI and compare it with various traditional FWIs. We find that the MLFWI significantly outperforms traditional indices in predicting wildfire occurrence, achieving an area under the curve score of 0.99 compared to 0.62–0.80. We recommend applying the MLFWI in wildfire warning systems.

**1. Introduction**

In recent years extreme wildfire have become increasingly frequent in various regions around the globe (Liu *et al* 2022). Estimating wildfire risk is an extremely important yet challenging task. Wildfire behavior is affected by complex interactions between meteorological factors at multiple physical scales, fuel loads, topography, and anthropogenic factors (Worsnop *et al* 2020). The effect of these factors and their interactions on wildfire behavior is determined by complex physical processes which are difficult to fully model (Srock *et al* 2018). For these reasons, the commonly applied fire weather indices (FWIs) are mostly based on statistical analyses (McArthur 1967, Van Wagner and Forest 1987, Srock *et al* 2018; etc).

There are several different FWIs which are commonly used for wildfire warning systems. The most common index in Australia is the McArthur Forest Fire Danger Index (McArthur 1967). In the United States the National Fire Danger Ratings System is most common (Deeming *et al* 1977). Finally, the Canadian Forest FWI is also commonly applied (Van Wagner and Forest 1987). Over the years additional studies have aimed to increase the applicability of these FWIs and adjust them to additional vegetation types (e.g. Wotton and Beverly 2007 with the Canadian FWI).

Several factors are known to have a major impact on wildfire risk. Meteorological factors and the vegetation's fuel moisture content are probably the most significant factors (e.g. Jurdao *et al* 2012). Specifically, relative humidity (RH) has been found to be one of the most important factors by almost any study in the field (e.g. Srock *et al* 2018). Vegetation is also an important factor for wildfire risk; the most commonly used vegetation index is the normalized difference vegetation index (NDVI) which is used to estimate the density of live green vegetation (e.g. Bjånes *et al* 2021). Fuel loads are affected by the precipitation in the region in the previous year (e.g. Verhoeven *et al* 2020). Fuel loads are also affected by the time since the last regional wildfire, as previous studies have demonstrated that burn scars limit fire spread (Parks *et al* 2015). Anthropogenic factors are extremely influential both on the probability of wildfire ignition and on its propagation rate in case it ignites (e.g. Andela *et al* 2017).

While the traditional FWIs are undoubtedly effective and have been widely used in fire alert systems throughout the world, they do have several disadvantages. First, the traditional FWIs have been planned

according to specific regions of the world, and they may be inadequate in other regions (Giuseppe *et al* 2020). Second, the traditional FWIs have been developed using traditional statistical methods, which have been known to underperform machine learning (ML)-based models (e.g. Son *et al* 2022). Finally, traditional FWIs are mostly based on a limited number of independent variables; specifically, some of the variables described in the previous paragraph are not reflected in the traditional FWIs (Kondylatos *et al* 2022). One promising direction of research that could address exactly these issues is the application of ML models. ML-based models are known to outperform linear regressions and other traditional statistical methods. ML models can provide accurate predictions even when based on a large number of variables with complex interactions. In fact, their advantage over traditional statistical methods is especially distinct in such circumstances.

The goal of this paper is to develop novel machine learning based fire weather indices (MLFWIs). ML models have been applied in almost every scientific field in recent years and have provided outstanding results. ML models have also been widely used for the purpose of estimating wildfire risk. A comprehensive review of the work in the field has been summarized in a review by Jain *et al* (2020). Scholars have applied various ML models for this purpose, including random forests (RFs), Support Vector Machine, Artificial Neural Networks, Adaboost, and more. The vast majority of these studies are performed in a regional scale. To name a few examples, both Castelli *et al* (2015) and Wood (2021) developed models predicting burned areas in the Montesinho Natural Park in Portugal; Cao *et al* (2017) developed a wildfire susceptibility mapping model for the Yunnan Province in Southern China with a 10 day temporal resolution; Van Beusekom *et al* (2018) developed a RF classifier for wildfire occurrence in Puerto Rico with a daily resolution; Wang and Wang (2019) developed RF models in South Central US with a monthly resolution; Bjånes *et al* (2021) developed a deep learning ensemble model in two regions in Chile with a monthly resolution; etc.

While regional-scale studies have successfully demonstrated the potential of ML models, they are limited in several aspects. First, large datasets are necessary to properly train ML models; regional studies are mostly more limited in this sense. Second, the applicability of these studies is limited to the region in which the models were trained. Finally, in terms of studying the influence of various factors on wildfire risk, regional models do not provide information on the impact of region-dependent factors; for example, a model which is trained on a flat plain will not reflect the influence of topographic slope on wildfire risk.

In recent years scholars have made substantial advancements in generating global wildfire datasets based on satellite observations. Global wildfire datasets such as (Chuvieco *et al* 2018, Andela *et al* 2019, Lizundia-Loiola *et al* 2020) are freely available and have opened the way for studies on global wildfire risk estimation, including studies which develop ML models for this purpose. Chuvieco *et al* (2021) have analyzed the annual variability of biomass burning using ML models in 0.25° resolution grid; Forkel *et al* (2019) predicted regional burned areas in a monthly resolution in a 1.89° × 2.5° resolution grid; Zhang *et al* (2021) have developed convolutional neural network and multilayer perceptron (MLP) classification models to create wildfire risk mapping in a monthly 0.25° resolution grid. Shmuel and Heifetz (2022) estimated wildfire risk in a monthly 0.25° resolution global dataset.

These state-of-the-art models on global datasets have greatly contributed to wildfire risk estimation on a global scale; however, they are all performed in a monthly temporal resolution, which is not sufficient for the purpose of creating wildfire alerts. The goal of the current study is to develop ML models in a daily resolution and to evaluate the accuracy of these models in comparison to the traditional fire weather indices. Since we aim to develop MLFWIs that can be used globally, we only include simple and widely available factors in our models. We compare the MLFWIs to 14 traditional fire weather indices and subindices and find that they substantially outperform them in wildfire risk prediction.

We do not only use the traditional FWIs as a benchmark accuracy, but also examine the contribution of combining them in the ML-based models as input data. By doing so we integrate the accumulated wisdom in the field and build upon it instead of training the models from scratch.

The objectives of the paper are as follows:

- (a) To enable accurate wildfire danger predictions by applying advanced ML models using various data.
- (b) To analyze the most significant factors affecting wildfire risk, both individually and in interaction with additional factors.
- (c) To demonstrate the potential of applying ML-based FWIs in actual fire warning systems.
- (d) To examine the contribution of adding traditional FWIs (in addition to the raw data) as input data when training ML-based models.
- (e) To examine the potential of predicting extreme wildfires using ML-based models.

The paper is organized as follows. We begin by describing our data and the methods used to develop the MLFWIs. We then present the main results of the study. We present the prediction accuracies of the MLFWIs and compare it with the performance of 14 traditional indices and subindices. We also analyze the influence

of the various factors and their interactions on the models, providing scientific insights and understanding of the mechanism by which the models work. Finally, we compare the performance of the MLFWIs to that of traditional indices in predicting the 100 largest wildfires in the dataset. We conclude with a discussion on the results and their significance, recommendations for practical applications of the models, and suggestions for future research.

## 2. Research design

### 2.1. Data

The target (dependent) variable, wildfire occurrence, is based on the global wildfire dataset published by Artés *et al* (2019). The dataset includes Shapefiles describing the daily-burned polygons at the individual fire level with global coverage. The polygons are available in a 250-meter resolution. For the purpose of estimating wildfire occurrence, we use a 0.25° global grid with daily binary values. For each region and day, the value 1 is assigned if any new fire ignited in that specific region and time and the value 0 is assigned otherwise. We cover all global wildfires of 2016, resulting in a total of 44 460 216 observations in 121 476 regions, including 2 409 079 burning days of 1 024 926 different wildfires. As the number of burned observations is significantly smaller than the number of unburned observations, we balance the data by randomly sampling unburned observations so that their number would be equal to the number of burned observations (e.g. Hasanin and Khoshgoftaar 2018). The prediction accuracies of an unbalanced data would probably be higher, but would not reflect the true performance of the model. For example, if 90% of the observations were unburned, simply guessing that all observations are unburned would lead to an accuracy of 90%.

We use various features (independent variables) in our models, including meteorological factors, topography, fuel loads, and anthropogenic factors. We include 2-meter temperature, precipitation, RH and 10-meter wind velocity, based on the ERA5 hourly reanalysis data (Hersbach *et al* 2020). Specifically for precipitation and RH, for each observation we include three features: present value, mean value in the previous month, and mean value in the previous year. We also include a variable for incoming short-wave solar radiation, obtained as a daily mean in 0.25° resolution regions (Troccoli 2020).

Topographic slope can affect the propagation of wildfires (e.g. Pimont *et al* 2012). We include the mean slope in each region, based on the dataset published by Amatulli *et al* (2018).

Population density is known to have a significant effect on wildfire danger (e.g. Andela *et al* 2017). We include a variable which describes the mean population density in each 0.25° region based on the dataset of the Center for International Earth Science Information Network—Columbia University (2018).

We include the NDVI, a dimensionless parameter which is commonly used to estimate the density of live green vegetation. The NDVI variable is calculated as the difference between near-infrared and red reflectance, divided by their sum (Didan *et al* 2015):  $NDVI = \frac{NIR - Red}{NIR + Red}$ . We obtained monthly 2016 NDVI values for each 0.25° region from the NASA Earth Observations website (Didan *et al* 2015).

We use 14 different fire weather indices and subindices. These indices include three groups: (1) Canadian Forest Service Fire Weather Index Rating System; (2) Australian McArthur Mark 5 Rating System; (3) U.S. Forest Service National Fire-Danger Rating System. The variables in each group include (1a) fire weather index; (1b) build up index; (1c) danger index; (1d) drought code; (1e) duff moisture code; (1f) initial fire spread index; (1g) fine fuel moisture code; (1h) fire daily severity rating; (2a) Keetch-Byram drought index; (2b) fire danger index; (3a) spread component; (3b) energy release component; (3c) burning index; (3d) ignition component. These variables were available in a 0.25° resolution from the Copernicus Climate Change Service (Fire Danger Indices Historical Data from the Copernicus Emergency Management Service 2021).

In some of the models we examine the contribution of regional wildfire history to the models' performance. In these cases we include two variables, one describing the area burned in the 0.25° region in the year before the observation, and one describing the mean regional burned area since 2003. Both variables are based on the ECMWF data (Lizundia-Loiola *et al* 2020). Both variables do not include information on the day of the observation to prevent data leakage.

The features and their sources are summarized in table 1.

### 2.2. Methodology

We develop various classification models which predict the risk of ignition in each region and day. The performance of classification models can be measured by several different metrics. One of the most common metrics is the area under the curve (AUC) metric, which describes the area under the relative operating characteristic (ROC) curve (Hanley and McNeil 1982). The ROC-AUC score is given in the range of 0–1—a completely random classification would obtain an ROC-AUC score of 0.5; scores in the range of 0.5–0.7 are

**Table 1.** Summary of features and data sources.

Variable	Abbreviation	Source
Daily ignition	Ignition	Artés <i>et al</i> (2019)
2 m temperature	Temp	Hersbach <i>et al</i> (2020)
Relative humidity	RH	
10-m wind speed	Wind_speed	
Precipitation	prec	
Mean relative humidity in previous month	RH_1_month	
Mean precipitation in previous month	Prec_1_month	
Mean relative humidity in previous year	RH_12_months	
Mean precipitation in previous year	Prec_12_months	
Mean slope	Slope	Amatulli <i>et al</i> (2018)
Population density	Population	Center for International Earth Science Information Network—Columbia University (2018)
Normalized difference vegetation index	NDVI	Didan <i>et al</i> (2015)
Incoming short-wave solar radiation	Radiation	Troccoli (2020)
Daily fire weather index	FWI	Fire Danger Indices
Daily build up index	BUI	Historical Data from the
Daily danger index	Danger	Copernicus Emergency
Daily drought code	Drought	Management Service (2021)
Daily duff moisture code	DMC	
Daily initial fire spread index	ISI	
Daily fine fuel moisture code	FFMC	
Daily fire daily severity rating	Severity	
Daily Keetch-Byram drought index	KBDI	
Daily fire danger index	FFDI	
Daily spread component	Spread	
Daily energy release component	Energy	
Daily burning index	BI	
Daily ignition component	IC	
Regional wildfire history	Fire_history	Lizundia-Loiola <i>et al</i> (2020)

considered poor; score in the range of 0.7–0.9 are considered moderate; and score of 0.9 and above are considered good (Bradley 1997, Cao *et al* 2017). In binary classification models the ROC-AUC metric is considered preferable to the accuracy metric (Ling *et al* 2003). However, as different metrics have different advantages and disadvantages, it is common to present several different accuracy metrics. In addition to the ROC-AUC metric, we present the PR-AUC which is the AUC of the Precision-Recall Curve (Boyd *et al* 2013), the accuracy metric, the true positive rate (TPR, also known as sensitivity) and the true negative rate (TNR, also known as specificity). These metrics are best defined using the components of the confusion matrix—true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN):

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

$$TNR = \frac{TN}{TN + FP} \quad (3)$$

We develop four different classification models: (i) RF (Biau and Scornet 2016). (ii) Extreme Gradient Boosting (XGBoost) (Chen and Guestrin 2016). (iii) MLP, a form of Neural Network (Ramchoun *et al* 2016), and (iv) logistic regression (Lever *et al* 2016). We perform a simple train-test split where 25% of the observations are used for testing. We perform these analyses using Python's Scikit-learn package (Pedregosa *et al* 2011), apart from the XGBoost model which is based on the XGBoost package (Chen and Guestrin 2016).

We tune the hyperparameters of the RF, XGBoost, and MLP models to improve their performance. For the XGBoost and RF models, we evaluate the following hyperparameters: number of estimators between 100 and 400 and maximal tree depth between 8 and 12. We only present the results of the highest performance models in the Results section. As for the MLP model, we evaluate its performance in the range of 1–3 hidden layers, with each layer including 50–150 neurons.

After comparing the performances of the ML models and the traditional indices, we also examine whether including the traditional indices in the training of the ML models improves their performance. In addition, we examine the contribution of a regional wildfire history variable to the models' performances.

One of the main objectives of this study is to discover the most influential factors in terms of wildfire risk and to understand how they affect the probability of wildfire occurrence. One possible method of doing so is by applying the SHAP (SHapley Additive exPlanations) values analysis (e.g. Mangalathu *et al* 2020). The basic idea behind this method is to quantify the effect of each feature on the model by comparing the model's output including the feature's value to its output without it. The process is repeated for each observation in the dataset and can provide information both on the mean effect and on the feature's influence in a specific observation. Furthermore, this method provides information on the direction of the effect (i.e. whether high and low values of each feature had a positive or negative effect on the model's output).

The SHAP value analysis is presented in a graph with one row for each feature. The features are ordered by their effect on the model in descending order. In each row, each dot represents the value of a specific feature in a specific observation; the color of the dot reflects the value of the feature, and its location on the horizontal axis reflects its effect on the model's output. To keep the graph interpretable, we randomly sample 5000 observations from the entire dataset.

While the SHAP analysis is an effective method of evaluating the influence of a single feature, it does not reflect the combined effects of multiple features. For this purpose, we also present two-way (relative to two variables) partial dependence plots (PDPs) of wildfire occurrence based on the highest performing model. The two-way PDPs provide clearer insights on the individual and combined effects of different variables on the model outcome.

Finally, we evaluate which of the various models could best predict the 100 largest wildfires in the dataset. To perform this comparison two issues must be addressed. First, to prevent data leakage, the 100 largest wildfires must not be part of the training data as well as the testing data. For this reason, we retrain the different models after removing the 100 largest wildfires from the training data. The second issue that must be resolved is that the output of the different indices and models is not set on the same range. To make a fair comparison, for each index or model we present the danger value in percentile, compared to all its values in dataset. The value in percentiles reflects the level of danger in comparison to different times and locations. A perfect prediction would mean that the spatiotemporal observations of the 100 largest wildfires receive the highest possible danger values compared to other times and locations in the dataset.

## 3. Results

### 3.1. Wildfire ignition

We present the ROC-AUC, PR-AUC, accuracy, TPR and TNR scores of the XGBoost, RF and MLP models alongside the performances of the logistic regression based on the meteorological data and logistic regressions based on each of the 14 traditional fire weather indices (table 2 and figure 1). The XGBoost model achieved the highest score, with a ROC-AUC of 0.98 on a scale of 0–1 and an accuracy of 0.95. The RF model obtained the second-best ROC-AUC and accuracy scores, 0.92 and 0.85 respectively. This result is in line with previous studies which have demonstrated that tree ensembles tend to outperform neural networks in tabular data (e.g. Shwartz-Ziv and Arnon 2022). The full logistic regression obtained a lower ROC-AUC score of 0.68, while the performances of the traditional indices were limited to ROC-AUC scores of 0.62–0.80. Similar results were obtained for the PR-AUC metric, with the XGBoost obtaining the highest PR-AUC score (0.94) followed by the RF model (0.89). Among the traditional indices and subindices, the highest performance was obtained by the Duff Moisture Code and the Keetch-Byram Drought Index, both achieving a 0.80 ROC-AUC score. The full ROC curves for the ML models and logistic regression are presented in figure 1.

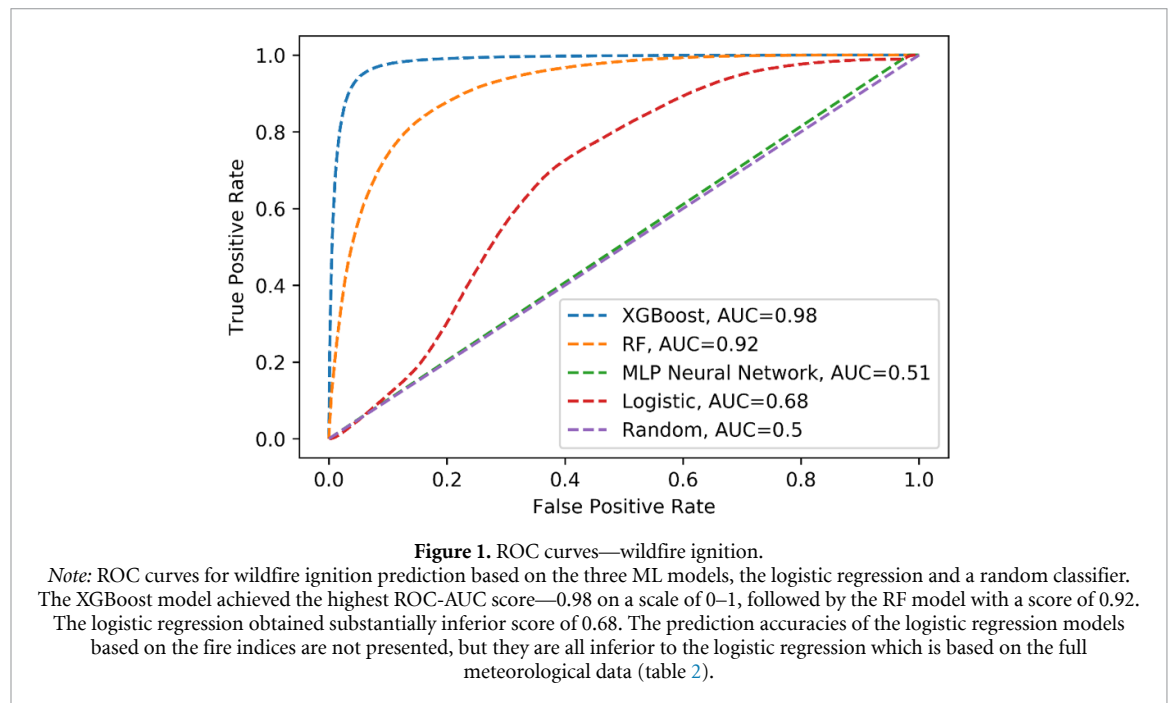
In table 3 we compare the performance of the previous LR and ML-based models (labeled LR-1 and MLFWI-1) to the performance of models which additionally include the traditional fire weather indices (LR-2 and MLFWI-2), and to models which additionally include the regional wildfire history (LR-3 and MLFWI-3). All three MLFWIs are based on the highest performance model, XGBoost. Both the traditional FWIs and the regional wildfire history increased the performance of the XGBoost model. The ROC-AUC score increased from 0.984 in MLFWI-1–0.990 in MLFWI-3 and the accuracy increased from 0.946 to 0.964 respectively. Similarly, the PR-AUC score increased from 0.938 in MLFWI-1–0.947 in MLFWI-3. The addition of the traditional FWIs had a negligible effect on the LR model, but inserting the regional wildfire history variable substantially increased its performance from an accuracy of 0.50–0.73. However, even the full LR model (LR-3) was substantially less accurate than the basic MLFWI (MLFWI-1).

Figures 2 and 3 provide interpretation of the most significant features and their effect on the model prediction. Figure 2 presents a SHAP values analysis for the XGBoost which provided the highest prediction accuracy. Figure 3 presents two-way PDP plots for each variable pair in the XGBoost model. Both the SHAP

**Table 2.** Summary of model performances—wildfire ignition risk.

Model	ROC-AUC	PR-AUC	Accuracy	TPR	TNR	Hyperparameters tested	Best parameters
RF	0.92	0.89	0.84	0.83	0.85	Max depth: 8–12 n_estimators: 100–400	12 400
XGBoost	0.98	0.94	0.95	0.93	0.96	Max depth: 8–12 n_estimators: 100–400	9 400
MLP	0.51	0.75	0.51	0.51	0.99	Hidden layers: 1–3 # neurons in layer: 50–150	1 50
LR	0.68	0.60	0.51	0.50	0.72	—	—
Random	0.50	0.50	0.50	0.50	0.50	—	—
Fire Weather Index	0.77	0.71	0.71	0.74	0.69	—	—
Build up Index	0.79	0.74	0.70	0.77	0.76	—	—
Danger Index	0.76	0.75	0.71	0.74	0.69	—	—
Drought Code	0.68	0.59	0.63	0.66	0.62	—	—
Duff Moisture Code	0.80	0.75	0.70	0.78	0.65	—	—
Initial Fire Spread Index	0.75	0.68	0.69	0.72	0.66	—	—
Fine Fuel Moisture Code	0.77	0.72	0.70	0.65	0.80	—	—
Fire Daily Severity Rating	0.77	0.71	0.68	0.75	0.65	—	—
Keetch-Byram Drought Index	0.80	0.75	0.74	0.72	0.76	—	—
Fire Danger Index	0.76	0.71	0.70	0.74	0.68	—	—
Spread Component	0.62	0.60	0.62	0.63	0.61	—	—
Energy Release Component	0.73	0.69	0.67	0.70	0.65	—	—
Burning Index	0.63	0.63	0.63	0.66	0.61	—	—
Ignition Component	0.65	0.66	0.63	0.67	0.61	—	—

Optimized scores of the different models. The hyperparameters tested for the ML models are presented, alongside the chosen values. A logistic regression fit based on each traditional index is presented for comparison.



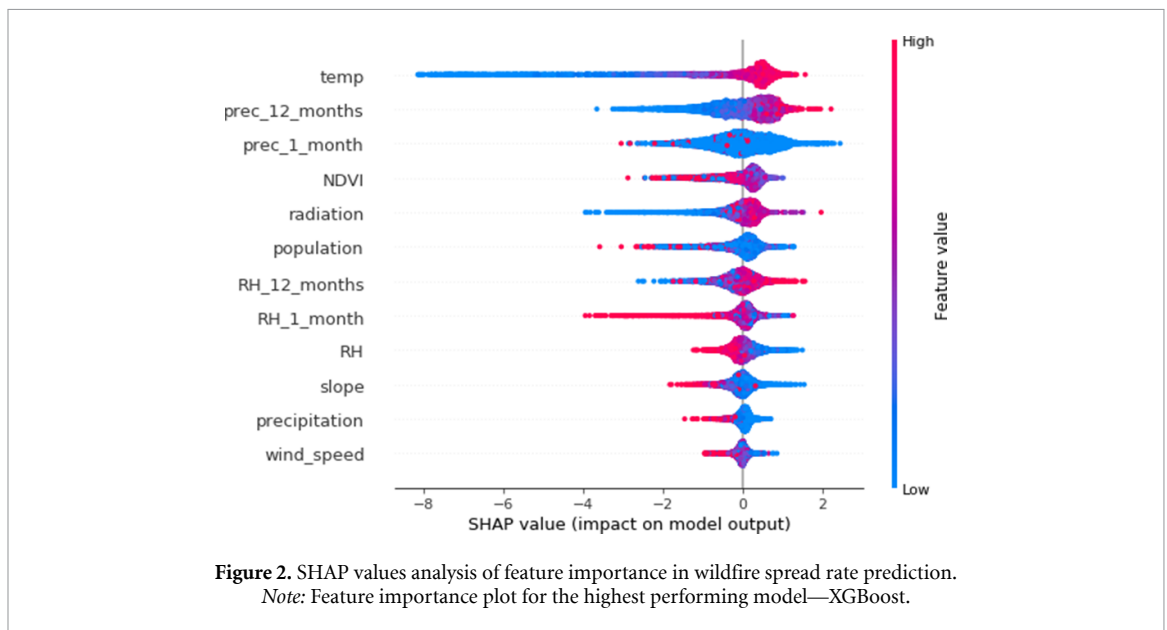
analysis and the PDPs are performed on the basic XGBoost model (MLFWI-1) which does not include traditional FWIs or regional wildfire history as predictors.

We present the prediction accuracies of the different MLFWI and LR models for the testing data in figure 4. The MLFWI-1 model (figure 4(a)) provided an almost perfect prediction in all relevant regions. The prediction accuracies of MLFWI-2 and MLFWI-3 are not presented, as they appear very similar. The LR models, in contrast, had substantially different prediction accuracies maps. The prediction accuracy of LR-1 (figure 4(b)) was relatively poor and did not significantly outperform a random classification. LR-2

**Table 3.** Summary of model performances including FWIs and regional wildfire history.

Model	Variables	ROC-AUC	PR-AUC	Accuracy	TPR	TNR
MLFWI-1	Temp, RH, precipitation, radiation, slope, NDVI, population, RH_1_month, prec_1_month, RH_12_months, prec_12_months, wind_speed	0.984	0.938	0.946	0.934	0.959
MLFWI-2	As in MLFWI-1+FWIs	0.988	0.940	0.959	0.951	0.968
MLFWI-3	As in MLFWI-2+fire_history	0.990	0.947	0.964	0.958	0.971
LR-1	As in MLFWI-1	0.680	0.598	0.510	0.510	0.720
LR-2	As in MLFWI-2	0.683	0.599	0.507	0.503	0.719
LR-3	As in MLFWI-3	0.754	0.766	0.730	0.765	0.704

Summary of the MLFWI and LR models' performances. All three MLFWIs are based on the XGBoost model. MLFWI-1 and LR-1 include the original data as in table 2. MLFWI-2 and LR-2 include the traditional FWIs in addition to the data of model #1. MLFWI-3 and LR-3 include regional wildfire history in addition to the data in model #2.

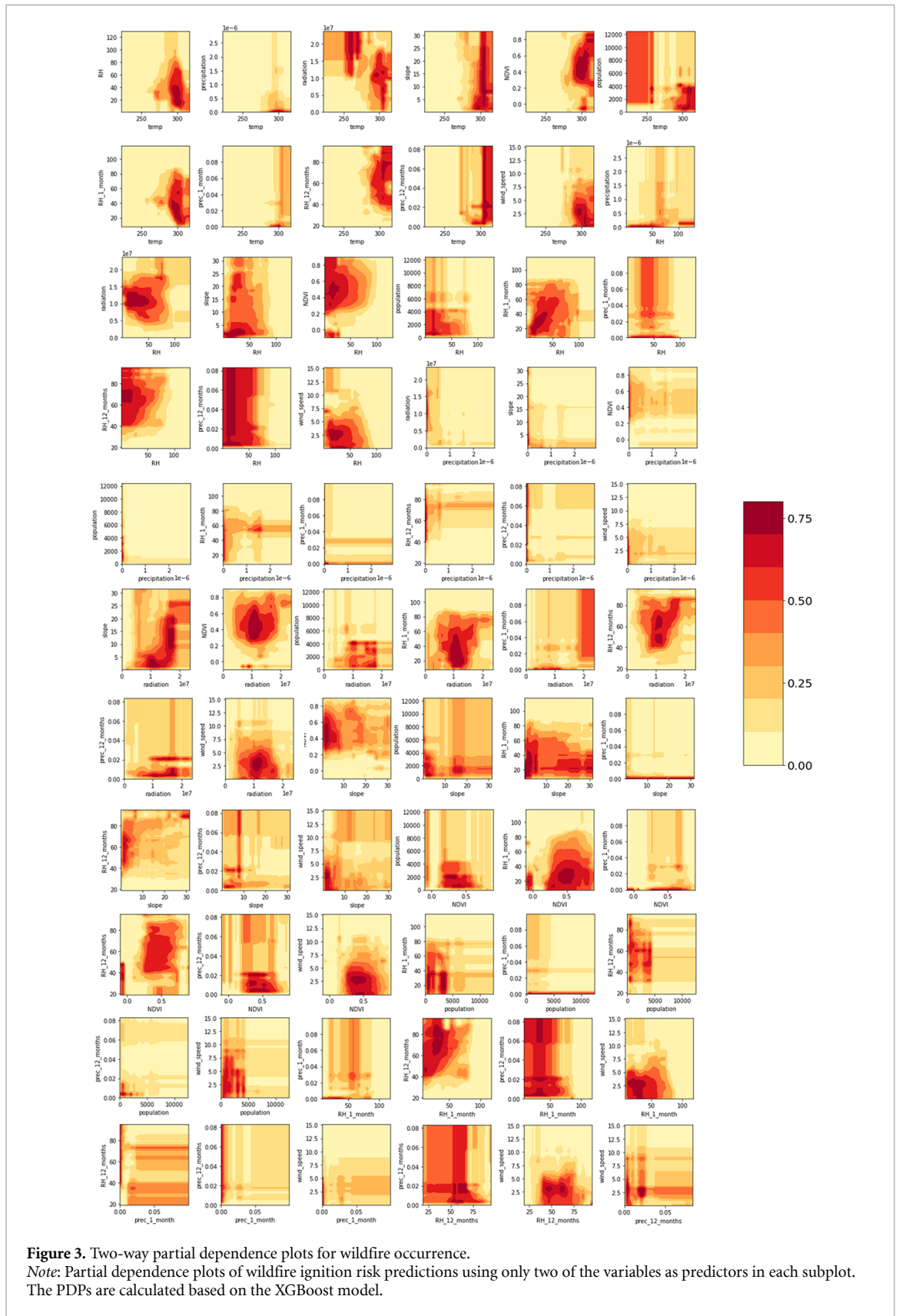


(figure 4(c)), which includes FWIs data, performed well in certain regions such as Africa, but performed as poorly as LR-1 in other regions. LR-3 (figure 4(d)), which includes data on regional wildfire history, achieved significantly higher accuracy scores in almost all regions globally; however, its accuracy was still substantially lower than even the basic MLFWI model (MLFWI-1)—ROC-AUC score of 0.984 compared to 0.754 and accuracy of 0.946 compared to 0.730 (table 3).

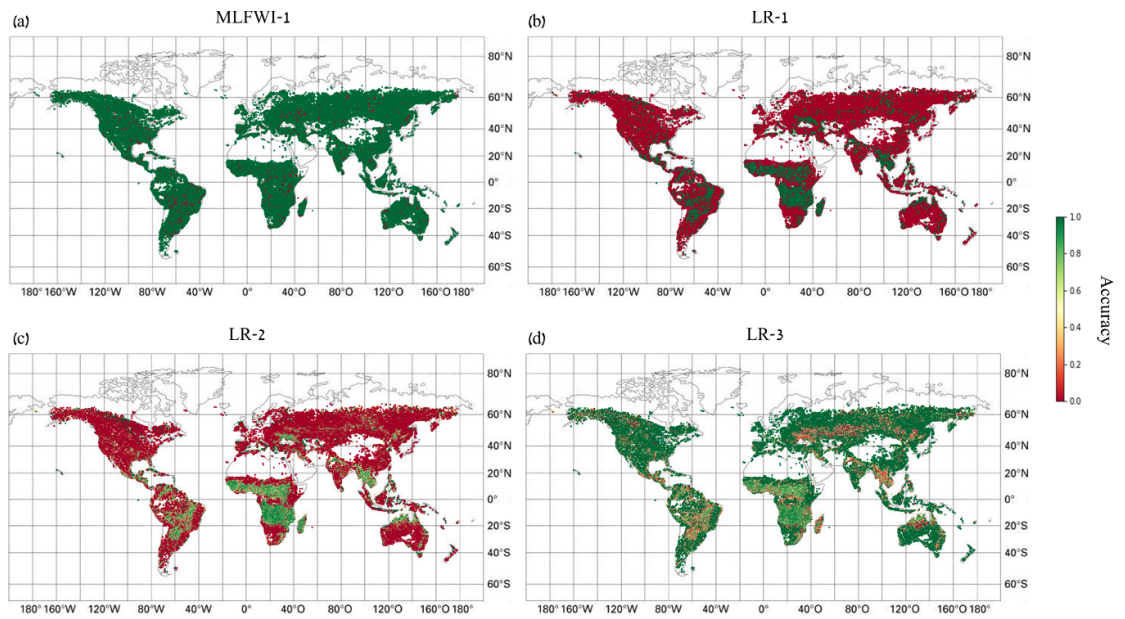
Figure 5 presents the predictions of the 100 largest wildfires in the dataset. We emphasize that these observations were removed from the training data to prevent data leakage. To enable a comparison between the models and indices, we present the values as percentiles of the full dataset. A perfect prediction would mean that these 100 observations would be ranked in the higher percentile, in the upper part of the graph. While none of the models obtain a perfect prediction, the values in the ML models are substantially more centered in higher percentiles, most of them above 80th percentile. The performance of the FWI and FFDI indices is substantially lower, and their observations are almost homogenously scattered around the vertical axis.

In figure 5(a) the danger ranking is calculated in comparison to all the observations in the dataset, including spatiotemporal observations in which no wildfires occurred. When predicting the probability of ignition for the days and locations of the 100 largest wildfires, the three ML models ranked the median

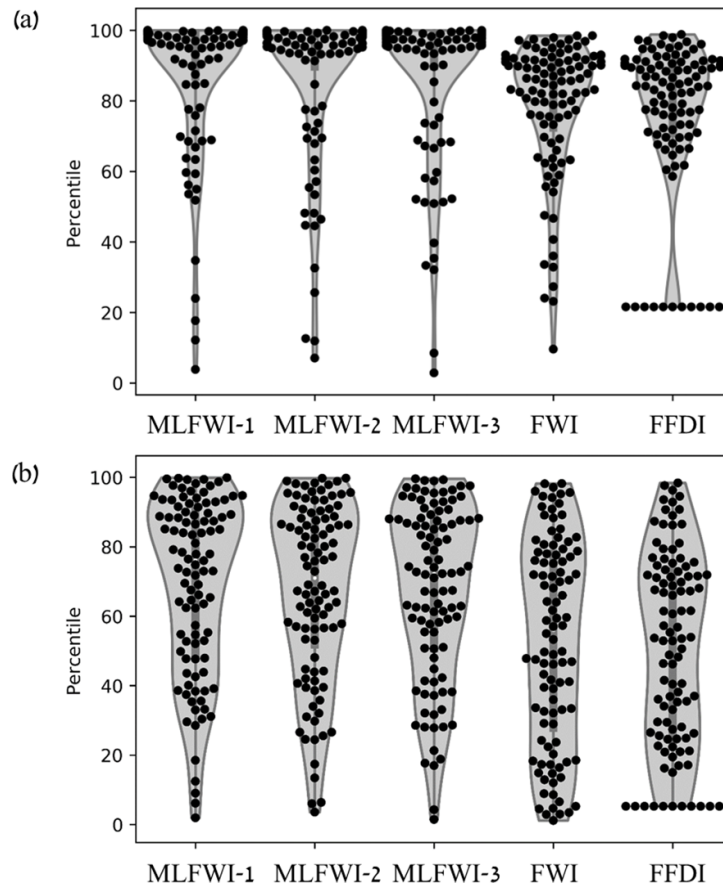




danger value of these 100 observations at 97.5, 97.0 and 97.4 percentiles (from left to right); the median FWI index of these observations was only 85th percentile and the FFDI value was at the 84th percentile. Similarly, in figure 5(b) which calculates the percentiles in comparison to spatiotemporal observations in which a wildfire ignited, the ML models scored median values of 76.2, 71.0 and 72.3 percentiles, in comparison to 56.0 and 53.6 percentiles of the FWI and FFDI indices.



**Figure 4.** Prediction accuracy maps—wildfire occurrence.  
*Note:* Global maps of wildfire occurrence classification models. The variables included in each model are detailed in table 3.



**Figure 5.** Prediction of the 100 largest wildfires in the dataset.  
*Note:* A violin plot of the value in percentiles for the spatiotemporal observations of the 100 largest wildfires in dataset, according to different ML models and two traditional indices. In figure 5(a) the percentiles are calculated in comparison to all the observations in the dataset. In figure 5(b) the percentiles are calculated in comparison to days and locations in which a wildfire ignited.

## 4. Discussion

In this study we developed ML-based fire weather indices and compared their performance to traditional FWIs. Previous studies which have mapped wildfires globally provided us with numerous observations to be used as training data for the ML models. By crossing these fire occurrences with meteorological factors, topography, fuel loads and anthropogenic factors at the time and place of the fires we were able to develop accurate predictions of wildfire ignition.

We compared the prediction performance of our models to 14 traditional fire weather indices and subindices. As these FWIs are commonly used by weather services, we consider their performance as the benchmark performance for wildfire risk estimation. The three ML models developed in this study significantly outperformed all 14 traditional FWIs in predicting wildfire ignition and obtained an ROC-AUC score of up to 0.99 compared to 0.62–0.80 scores of the traditional indices. Including a regional wildfire history variable improved the performance of the model, but even without it the MLFWI achieved an almost perfect 0.98 ROC-AUC score.

The most important feature for the prediction of wildfire occurrence was found to be temperature. The effect of temperature on the model's output is characterized by a long left 'tail', meaning low temperatures could significantly decrease the probability of wildfire ignition (figure 2). Precipitation in the previous year was also a significant variable, in line with previous studies which have found that high precipitation is associated with high fuel loads and increases the probability of wildfires in the long term (e.g. Verhoeven *et al* 2020). Precipitation in the shorter term (one month) had a mixed effect, as rains increase fuel moisture content in the short term. As expected, daily precipitation eliminated the probability of ignition almost entirely (figure 3).

Several terrain-related factors had a substantial impact on our models. NDVI had a significant yet non-linear effect on the model. Moderate positive values which reflect unhealthy vegetation increased the probability of wildfire occurrence. The effect of NDVI was even larger in combination with additional risk factors such as low RH (figure 3). When RH was low, high NDVI were also at a risk of ignition; however, when RH was medium or high, the risk of ignition at high NDVI regions was smaller. Topographic slope had a relatively small effect on the probability of ignition, as could be expected—its major effect is on wildfire propagation. High population density reduced the probability of wildfire occurrence, in line with previous studies (Andela *et al* 2017). This result holds even when examining the combined effect of population density with various additional factors (figure 3). Research suggests several mechanisms which reduce wildfire risk in populated areas: high suppression efficiency in densely populated areas (Wang and Wang 2019); changes in land use in developed and populated countries produce a sharp decline in burned areas compared to natural landscapes (Andela *et al* 2017); firebreaks, which reduce the potential of ignitions to become substantial wildfires, are more common in populated areas (Pinto *et al* 2020). It is important to note that several mechanisms operate in the opposite direction and increase wildfire risk as population density increases. The most important mechanism of this kind is human-related ignitions. However, studies have consistently shown that this effect is considerably smaller and the mechanisms reducing wildfire risk in populated areas have an overwhelmingly stronger effect (Knorr *et al* 2014, Andela *et al* 2017). At most, studies have found a non-monotonic relationship between population density and wildfire risk with the risk initially increasing up to a certain (relatively low) population threshold but then strongly declines as population density continues to increase (e.g. Bistinas *et al* 2013).

Wind is known to be one of the most important factors in determining wildfire dynamics (Cruz and Alexander 2019). Strong winds increase flame lengths and encourage a rapid and hazardous progress of wildfires (Abatzoglou *et al* 2018); strong winds cause spotting and support convective pre-heating which are associated with firestorms and fatalities (Pagni 1993); winds limit aerial fire suppression efforts and ground aircrafts beyond a certain wind velocity threshold (NWCG 2014). Winds are associated not only with the spread rate of existing fires but also with the probability of ignition, as they supply oxygen and encourage ignitions. Previous studies have shown that the probability of ignition is higher in the presence of moderate winds (approximately  $3 \text{ m s}^{-1}$ ) compared to little no winds ( $0\text{--}1 \text{ m s}^{-1}$ ) (Harrison 1970, Ellis 2000, Manzello *et al* 2006, Curt *et al* 2007, Ganteaume *et al* 2009, Schiks and Wotton 2015). The positive effect of wind on wildfire spread rate is reflected in the traditional FWIs (Nelson 1964, Deeming *et al* 1977). The impact of wind velocity on wildfire occurrence probability is also reflected in the models developed in the current paper. However, figure 4 reflects a non-monotonic effect of wind velocity on wildfire occurrence. While some wind is required for wildfire ignition, strong winds create an opposite effect and reduce the probability of ignition. Scholars have identified several mechanisms for this phenomenon. First, studies have shown that beyond a certain wind velocity, ignition sources such as cigarette butts are more likely to be extinguished (Satoh *et al* 2003, Koo *et al* 2010). Second, the wind's cooling effect reduces the transition from

smouldering to flaming beyond a certain threshold (Santoso *et al* 2019). Viegas *et al* (2021) developed a physical model which predicts a non-monotonic correlation between wind velocity and wildfire ignition probability. These effects have also been demonstrated in laboratory experiments. Sun *et al* (2018) performed 2500 ignition attempts with cigarette butts in the presence of varying winds; they found positive correlation between wind velocity and ignition probability up to a certain threshold and a negative correlation beyond this threshold. Plucinski and Anderson (2008) reached a similar conclusion in an experiment in which they ignited cotton balls. The traditional FWIs include a monotonic relation between wind velocity and wildfire ignition probability and do not reflect the complex relation described above. While a monotonic relation may be appropriate when assessing the progress rate of existing wildfires, a non-monotonic relation may be more appropriate in assessing the probability of ignition. As winds are one of the most important factors in wildfire science, we believe this issue warrants further research.

In addition, we demonstrated the effects of additional hazardous combinations such as low RH and high temperatures, moderate wind and low RH, and more. While some of these combined effects are already reflected in existing FWIs, the more complex interactions are best reflected in the full ML models.

The results of this study could have several important implications. First, we find that the improvement in prediction accuracies compared to the LR models and the traditional FWIs warrants examining the application of the MLFWIs in actual fire weather services. Accurate prediction of wildfire occurrence is an essential tool for almost any fuel management and firefighting strategy (Taylor *et al* 2013, Trucchia *et al* 2022). The models can be used to alert population, to provide firefighting forces with daily wildfire risk maps, or to allow forest management to effectively plan fuel reduction. Our models were developed on a global dataset and were only based on simple widely available variables, making them easily applicable wherever needed. More importantly, the results indicate that the high performances of the models are obtained throughout the world and not restricted to a specific region or continent. Future studies can build on this work to develop ML-based wildfire risk models on a regional scale; while limiting the scale of the model to a specific region could reduce the amount of available data, it could allow the model to address region-specific phenomena more accurately; for example, different models could be developed in regions where lightning-caused wildfires are more frequent.

## 5. Conclusions

As extreme fire weather and mega fires are rising in frequency in the face of global climate change, accurate wildfire risk estimation is becoming increasingly important for forest management and firefighting. Wildfire danger estimation depends on non-linear interactions between multiple factors including meteorological factors, topography, and anthropogenic factors. Various FWIs have been developed for the purpose of estimating wildfire risk. In this study we developed a ML-based FWI which provides wildfire danger estimation in a daily resolution in all relevant regions worldwide. The MLFWI obtained extremely high accuracy scores and significantly outperformed various traditional FWIs and LR models which were based on the same data. The MLFWI is based on simple and widely available factors and is therefore available to any fire weather service that would be interested in applying it. We propose to build upon this study to gradually replace the existing fire weather indices with ML-based indices, which have the potential of substantially improving fire weather alerts.

## Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

## Acknowledgments

We would like to thank the two anonymous reviewers and the editors for providing us with insightful comments which substantially improved the manuscript.

## Funding

This research received no external funding.

## Conflicts of interest

The authors declare no conflict of interest.

## ORCID iDs

Assaf Shmuel  <https://orcid.org/0000-0002-1794-9381>

Eyal Heifetz  <https://orcid.org/0000-0002-3584-3978>

## References

- Abatzoglou J T, Balch J K, Bradley B A and Kolden C A 2018 Human-related ignitions concurrent with high winds promote large wildfires across the USA *Int. J. Wildland Fire* **27** 377–86
- Amatulli G, Domisch S, Tuanmu M-N, Parmentier B, Ranipeta A, Malczyk J and Jetz W 2018 A suite of global, cross-scale topographic variables for environmental and biodiversity modeling *Sci. Data* **5** 180040
- Andela N, Morton D C, Giglio L, Chen Y, van der Werf G R, Kasibhatla P S and Randerson J T 2017 A human-driven decline in global burned area *Science* **356** 1356–62
- Andela N, Morton D C, Giglio L, Paugam R, Chen Y, Hantson S and Randerson J T 2019 The global fire atlas of individual fire size, duration, speed and direction *Earth Syst. Sci. Data* **11** 529–52
- Artés T, Oom D, De Rigo D, Durrant T H, Maianti P, Libertà G and San-Miguel-Ayanz J 2019 A global wildfire dataset for the analysis of fire regimes and fire behaviour *Sci. Data* **6** 1–11
- Biau G and Scornet E 2016 A random forest guided tour *Test* **25** 197–227
- Bistinas I, Oom D, Sá A C, Harrison S P, Prentice I C and Pereira J M 2013 Relationships between human population density and burned area at continental and global scales *PLoS One* **8** e81188
- Bjånes A, De La Fuente R and Mena P 2021 A deep learning ensemble model for wildfire susceptibility mapping *Ecol. Inform.* **65** 101397
- Boyd K, Eng K H and Page C D 2013 Area under the precision-recall curve: point estimates and confidence intervals *Joint European Conf. on Machine Learning and Knowledge Discovery in Databases* (Berlin : Springer) pp 451–66
- Bradley A P 1997 The use of the area under the ROC curve in the evaluation of machine learning algorithms *Pattern Recognit.* **30** 1145–59
- Cao Y, Wang M and Liu K 2017 Wildfire susceptibility assessment in Southern China: a comparison of multiple methods *Int. J. Disaster Risk Sci.* **8** 164–81
- Castelli M, Vanneschi L and Popović A 2015 Predicting burned areas of forest fires: an artificial intelligence approach *Fire Ecol.* **11** 106–18
- Center for International Earth Science Information Network—Columbia University 2018 Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC) *Population Density, Revision 11* (Accessed 16 October 2021) (<https://doi.org/10.7927/H49C6VHW>)
- Chen T and Guestrin C 2016 Xgboost: a scalable tree boosting system *Proc. 22nd Acm Sigkdd Int. Conf. on Knowledge Discovery and Data Mining* pp 785–94
- Chuvieco E, Pettinari M L, Koutsias N, Forkel M, Hantson S and Turco M 2021 Human and climate drivers of global biomass burning variability *Sci. Total Environ.* **779** 146361
- Chuvieco E, Pettinari M L, Lizundia-Loiola J, Storm T and Padilla Parellada M 2018 ESA fire climate change initiative (fire\_cci): MODIS fire\_cci burned area pixel product, version 5.1 *Centre for Environmental Data 2018* (<https://doi.org/10.5285/58f00d8814064b79a0c49662ad3af537>)
- Cruz M G and Alexander M E 2019 The 10% wind speed rule of thumb for estimating a wildfire's forward rate of spread in forests and shrublands *Ann. For. Sci.* **76** 1–11
- Curt T, Ganteaume A, Alleaume S, Borgniet L, Chandiooux O, Jappiot M and Martin W 2007 Vegetation flammability and ignition potential at road-forest interfaces (southern France) *Proc. 4th Int. Wildland Fire Conf.* pp 14–18
- Deeming J E, Burgan R E and Cohen J D 1977 *The National Fire-danger Rating System, 1978* vol 39 (Ogden, UT: Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station)
- Di Giuseppe F, Vitolo C, Krzeminski B, Barnard C, Maciel P and San-Miguel J 2020 Fire weather index: the skill provided by the European Centre for medium-range weather forecasts ensemble prediction system *Nat. Hazards Earth Syst. Sci.* **20** 2365–78
- Didan K, Munoz A B, Solano R and Huete A 2015 *MODIS Vegetation Index User's Guide (MOD13 Series)* (Arizona, University of Arizona: Vegetation Index and Phenology Lab)
- Ellis P F 2000 The aerodynamic and combustion characteristics of eucalypt bark: a firebrand study
- Fire Danger Indices Historical Data from the Copernicus Emergency Management Service 2021 Fire danger indices historical data from the Copernicus emergency management service—user guide (available at: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-fire-historical?tab=overview>) (Accessed 17 May 2021)
- Forkel M *et al* 2019 Emergent relationships with respect to burned area in global satellite observations and fire-enabled vegetation models *Biogeosciences* **16** 57–76
- Ganteaume A, Lampin-Maillet C, Guijarro M, Hernando C, Jappiot M, Fonturbel T and Vega J A 2009 Spot fires: fuel bed flammability and capability of firebrands to ignite fuel beds *Int. J. Wildland Fire* **18** 951–69
- Hanley J A and McNeil B J 1982 The meaning and use of the area under a receiver operating characteristic (ROC) curve *Radiology* **143** 29–36
- Harrison R 1970 *Danger of ignition of ground cover fuels by vehicle exhaust systems* US Forest Service, Equipment Development Center, ED&T Project 1337 San Dimas, CA
- Hasanin T and Khoshgoftaar T 2018 The effects of random undersampling with simulated class imbalance for big data *2018 IEEE Int. Conf. on Information Reuse and Integration (IRI)* (IEEE) pp 70–79
- Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J and Thépaut J N 2020 The ERA5 global reanalysis *Q. J. R. Meteorol. Soc.* **146** 1999–2049
- Jain P, Coogan S C, Subramanian S G, Crowley M, Taylor S and Flannigan M D 2020 A review of machine learning applications in wildfire science and management *Environ. Rev.* **28** 478–505
- Jurdao S, Chuvieco E and Arevalillo J M 2012 Modelling fire ignition probability from satellite estimates of live fuel moisture content *Fire Ecol.* **8** 77–97
- Knorr W, Kaminski T, Arneth A and Weber U 2014 Impact of human population density on fire frequency at the global scale *Biogeosciences* **11** 1085–102
- Kondylatos S, Prapas I, Ronco M, Papoutsis I, Camps-Valls G, Piles M and Carvalhais N 2022 Wildfire danger prediction and understanding with deep learning *Geophys. Res. Lett.* **49** e2022GL099368
- Koo E, Pagni P J, Weise D R and Woycheese J P 2010 Firebrands and spotting ignition in large-scale fires *Int. J. Wildland Fire* **19** 818–43

- Lever J, Krzywinski M and Altman N 2016 Logistic regression. *Nat. Methods* **13** 541–2
- Ling C X, Huang J and Zhang H 2003 AUC: a better measure than accuracy in comparing learning algorithms *Conf. Canadian Society for Computational Studies of Intelligence* (Berlin: Springer) pp 329–41
- Liu Z, Eden J M, Dieppois B and Blackett M 2022 A global view of observed changes in fire weather extremes: uncertainties and attribution to climate change *Clim. Change* **173** 1–20
- Lizundia-Loiola J, Otón G, Ramo R and Chuvieco E 2020 A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data *Remote Sens. Environ.* **236** 111493
- Mangalathu S, Hwang S H and Jeon J S 2020 Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach *Eng. Struct.* **219** 110927
- Manzello S L, Cleary T G, Shields J R and Yang J C 2006 Ignition of mulch and grasses by firebrands in wildland–urban interface fires *Int. J. Wildland Fire* **15** 427–31
- McArthur A G 1967 *Fire behaviour in Eucalyptus Forests* Department of National Development Forestry and Timber Bureau, Canberra
- Nelson R M 1964 *The National Fire Danger Rating System: Derivation of Spread Index for Eastern and Southern States* vol 13 (Asheville, NC: US Department of Agriculture, Forest Service, Southeastern Forest Experiment Station)
- NWCG 2014 *Interagency Single Engine Air Tanker Operations Guide. National Wildfire Coordinating Group PMS 506 NFES 001844* (Boise, ID: National Interagency Fire Center)
- Pagni P J 1993 Causes of the 20 October 1991 Oakland hills conflagration *Fire Saf. J.* **21** 331–9
- Parks S A, Holsinger L M, Miller C and Nelson C R 2015 Wildland fire as a self-regulating mechanism: the role of previous burns and weather in limiting fire progression *Ecol. Appl.* **25** 1478–92
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O and Duchesnay E 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30
- Pimont F, Dupuy J L and Linn R R 2012 Coupled slope and wind effects on fire spread with influences of fire size: a numerical study using FIRETEC *Int. J. Wildland Fire* **21** 828–42
- Pinto G A S J, Rousseu F, Niklasson M and Drobyshev I 2020 Effects of human-related and biotic landscape features on the occurrence and size of modern forest fires in Sweden *Agric. For. Meteorol.* **291** 108084
- Plucinski M P and Anderson W R 2008 Laboratory determination of factors influencing successful point ignition in the litter layer of shrubland vegetation *Int. J. Wildland Fire* **17** 628–37
- Ramchoun H, Idrissi M A J, Ghanou Y and Ettaouil M 2016 Multilayer perceptron: architecture optimization and training IJIMAI **4** 26–30
- Santoso M A, Christensen E G, Yang J and Rein G 2019 Review of the transition from smouldering to flaming combustion in wildfires *Front. Mech. Eng.* **49**
- Satoh K, Zhong Y L and Yang K T 2003 Study of forest fire initiation due to lit cigarettes—measurement and observation of flaming probability of dried leaves *Proc. 6th ASME-JSME Thermal Engineering Joint Conf. (Hawaii, USA, 16–20 March 2003)* Paper TED-AJ03-551
- Schiks T J and Wotton B M 2015 Assessing the probability of sustained flaming in masticated fuel beds *Can. J. For. Res.* **45** 68–77
- Shmuel A and Heifetz E 2022 Global wildfire susceptibility mapping based on machine learning models *Forests* **13** 1050
- Shwartz-Ziv R and Armon A 2022 Tabular data: deep learning is not all you need *Inf. Fusion* **81** 84–90
- Son R, Ma P L, Wang H, Rasch P J, Wang S Y, Kim H and Yoon J H 2022 Deep learning provides substantial improvements to county-level fire weather forecasting over the western united states *J. Adv. Model. Earth Syst.* **14** e2022MS002995
- Srock A F, Charney J J, Potter B E and Goodrick S L 2018 The hot-dry-windy index: a new fire weather index *Atmosphere* **9** 279
- Sun P, Zhang Y, Sun L, Hu H, Guo F, Wang G and Zhang H 2018 Influence of fuel moisture content, packing ratio and wind velocity on the ignition probability of fuel beds composed of Mongolian oak leaves via cigarette butts *Forests* **9** 507
- Taylor S W, Woolford D G, Dean C B and Martell D L 2013 Wildfire prediction to inform fire management: statistical science challenges *Stat. Sci.* **28** 586–615
- Troccoli A 2020 Solar Radiation—Variable Fact Sheet *Copernicus Climate Change Service*
- Trucchia A, Izadgoshasb H, Isnardi S, Fiorucci P and Tonini M 2022 Machine-learning applications in geosciences: comparison of different algorithms and vegetation classes' importance ranking in wildfire susceptibility *Geosciences* **12** 424
- Van Beusekom A E, Gould W A, Monmany A C, Khalyani A H, Quiñones M, Fain S J and González G 2018 Fire weather and likelihood: characterizing climate space for fire occurrence and extent in Puerto Rico *Clim. Change* **146** 117–31
- Van Wagner C E and Forest P 1987 Development and structure of the Canadian forest fire weather index system *Can. For. Serv., Forestry Tech. Rep.*
- Verhoeven E M, Murray B R, Dickman C R, Wardle G M and Greenville A C 2020 Fire and rain are one: extreme rainfall events predict wildfire extent in an arid grassland *Int. J. Wildland Fire* **29** 702–11
- Viegas D X, Oliveira R, Almeida M and Kim D 2021 Ignition of fuel beds by cigarettes: a conceptual model to assess fuel bed moisture content and wind velocity effect on the ignition time and probability *Fire* **4** 35
- Wang S C and Wang Y 2019 Predicting wildfire burned area in South Central US using integrated machine learning techniques *Atmospheric Chem. Phys.* **2019** 1–25
- Wood D A 2021 Prediction and data mining of burned areas of forest fires: optimized data matching and mining algorithm provides valuable insight *Artif. Intell. Agric.* **5** 24–42
- Worsnop R P, Scheuerer M and Hamill T M 2020 Extended-range probabilistic fire-weather forecasting based on ensemble model output statistics and ensemble copula coupling *Mon. Weather Rev.* **148** 499–521
- Wotton B M and Beverly J L 2007 Stand-specific litter moisture content calibrations for the Canadian Fine Fuel Moisture Code *Int. J. Wildland Fire* **16** 463–72
- Zhang G, Wang M and Liu K 2021 Deep neural networks for global wildfire susceptibility modelling *Ecol. Indic.* **127** 107735