Where there's Smoke, there's Fire: Wildfire Risk Predictive Modeling via Historical Climate Data

Shahrzad Gholami [1], Narendran Kodandapani [2], Jane Wang [1], Juan Lavista Ferres [1*]

[1] AI for Good Research Lab, Microsoft, Redmond, USA
[2] Center for Advanced Spatial and Environmental Research, Bengaluru, India
[*] corresponding author: jlavista@microsoft.com

Abstract

Wildfire is a growing global crisis with devastating consequences. Uncontrolled wildfires take away human lives, destroy millions of animals and trees, degrade the air quality, impact the biodiversity of the planet and cause substantial economic costs. It is incredibly challenging to predict the spatio-temporal likelihood of wildfires based on historical data, due to their stochastic nature. Crucially though, the accurate and reliable prediction of wildfires can help the stakeholders and decision-makers take timely, strategic and effective actions to prevent, detect and suppress the wildfires before they become unmanageable. Unfortunately, most previous studies developed predictive models that suffer from some shortcomings: (i) in the evaluation phase, they do not take the temporal aspects into account precisely and they assume the independent and identically distributed random variables; (ii) they do not evaluate their approaches comprehensively, thus it is not clear if their proposed predictions and selected models remain reliable across different locations and years for practical deployment; and (iii) for the supervised learning models, they use predictor features and fire observations from the same time step in the training phase, which makes the inference task infeasible for future fire prediction. In this paper, we revisit the wildfire predictive modeling, explore the inherent challenges from a practical perspective and evaluate our modeling approach comprehensively via historical burned area, climate and geospatial data from three vast landscapes in India.

Introduction

A wildfire is an uncontrolled fire affecting areas that consist of combustible vegetation. Such fires threaten environmental resources and human lives significantly at a global scale. They also cause a substantial economic burden for governments and individuals (Thomas et al. 2017). For instance, the Australian bushfire and the Amazon rainforest wildfires in 2019 and 2020 affected nearly 44.5 million (18 million hectares)[1] and 2.3 million acres of area[2] , respectively. Also, in recent decades, there has been an increased forest fire activity across the western United States (Abatzoglou and Williams 2016). For instance as of September 2020, over 3.1 million acres in California has burned[3]. Although the major causes of wildfire ignitions are dry climate, lightning, and volcanic eruptions (Bowman et al. 2009), human mistakes also play an important role in such incidents (Balch et al. 2017). Besides arson and accidental ignitions, the uncontrolled use of fire for land-clearing and agricultural purposes by local communities (e.g., slash-and-burn farming in Southeast Asia) is also a major source of human-caused fires.

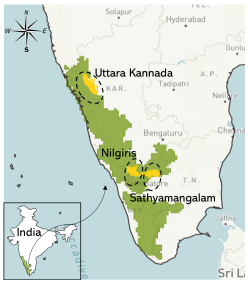


Figure 1: Southern Western Ghats is shown in green and three landscapes studied in this paper are shown in yellow. In total these landscapes cover 5306 sq. km.

Forest areas are vast and fire management resources are extremely scarce; thus managing resources strategically is a challenging task for stakeholders and decision-makers (i.e., forest and fire departments) in the field. In this paper, we study wildfire spatio-temporal risk prediction by exploiting the power of the artificial intelligence framework along with historical burned areas, climate, and geospatial data to assist decision-makers with identifying high-risk regions in protected areas. Such risk estimations are helpful to plan regarding where preventative actions and resources should be applied during high-risk seasons of wildfires.

Although wildfire risk predictive models based on historical data has been studied extensively in the past (Zhang, Lim, and Sharples 2016; Coffield et al. 2019; Catry et al. 2010), we believe it has to be revisited in terms of modeling details and extensive evaluation to assess accuracy, reliability, and practicality for actual deployment and decision-making in the field. The previous studies suffer

[1] http://bitly.ws/9JYE

[2] https://cbsn.ws/2UPSvjd

[3] https://www.fire.ca.gov/daily-wildfire-report/

from at least one of the following gaps: (i) the proposed predictive models are not evaluated comprehensively in terms of the train/test sets and modeling approach. To propose a reliable solution to the domain experts, proposing the best model for a single scenario/testing set is not sufficient. As such, we demonstrate the performance of several different models on several different test sets, created with temporal considerations, to assess the consistency of the proposed machine learning models performance to variations in the experimental setups across three landscapes in India. (ii) most of the previous studies only focus on the spatial distribution of fire and not the temporal aspect of this event when creating data points (O'Connor, Calkin, and Thompson 2017) and splitting their datasets for the evaluation step (Rodrigues and de la Riva 2014; Safi and Bouroumi 2013; Castelli, Vanneschi, and Popovič 2015). However, we evaluate our models based on a temporal division of the data as opposed to the random division of the entire dataset—a less challenging task compared to our approach. (Cheng and Wang 2008) proposed a predictive model based on a dynamic recurrent neural network to address spatio-temporal data mining upon historical observations in Canada. In this study, the data was divided temporally, but it suffers from the shortcomings discussed in item iv. (iii) they have created the data points such that predictive features and observation labels correspond to the same time step. However, from a practical perspective, it is impossible to have access to future predictive features to forecast future wildfire incidents. As such we suggest a time lag between the predictor features and the wildfire observation labels as explained in the Dataset Transformation and Modeling section, to evaluate our models more realistically. (iv) some of the previous studies propose accuracy as their only metric to evaluate the performance of their models. However, for highly imbalanced datasets like wildfire, this metric could be misleading. In this paper, we pose the problem as a binary classification task and we address all of the above gaps in the existing literature. We use the historical burned areas, climate and geospatial data of three vast landscapes in India to train and evaluate predictive models (see Figure 1).

Related Work

This paper is related to data-driven studies to forecast wildfires risk using machine learning models (Jain et al. 2020). The major trend in literature is the use of satellite data, local meteorological sensors and wildfire records to predict the risk of fire occurrence. While there are some prior studies proposing unsupervised context-based models (Salehi et al. 2016), the mainstream of the literature proposes supervised classification including mostly binary classification and in a few cases multi-label classification approaches (Sakr, Elhajj, and Mitri 2011; Özbayoğlu and Bozer 2012). For example, (Özbayoğlu and Bozer 2012) proposed an estimation of the burned area based on historical data from Turkey using a multi-layer perceptron, radial basis function networks, and support vector machines.

Binary classification methods have been extensively explored in previous studies. For example, (Cortez and Morais 2007) used meteorological data detected by local sensors in weather stations (i.e., Canadian Fire Weather Index and four weather conditions) from the northeast region of Portugal to predict the burned areas modeled as a regression task via neural nets, SVM, decision tree and random forest techniques. In (Zhang, Lim, and Sharples 2016), the authors proposed using Moderate Resolution Imaging Spectroradiometer (MODIS) data and logistic regression models to understand the effect of environmental and socioeconomic variables on predicting the spatial distribution of wildfire in Australia. (Coffield et al. 2019) has proposed a decision tree model as the best-performing model to study the controls and predictability of final fire size at the time of ignition based on fires in the boreal forests of Alaska. (Catry et al. 2010) proposes using logistic regression models to predict the likelihood of ignition occurrences based on the data from Portugal.

A comparison between the performance of logistic regression and neural nets to predict the spatial distribution of wildfires ignition in central Portugal is studied in (De Vasconcelos et al. 2001). (Stojanova et al. 2006) studied the predictive performance of logistic regression, random forests, decision trees, bagging and boosting ensemble methods on the meteorological data, MODIS data, and geographic data, along with the historical records of data from Slovenia. A fire-fighting management tool including a predictive neural network model is proposed in (Alonso-Betanzos et al. 2003) based on weather data and historical fire data from Galicia in northwest Spain. (Jaafari et al. 2019) proposed a binary classification model based on an adaptive neuro-fuzzy inference system combined with different metaheuristic optimization algorithms using a dataset from Iran.

data sources	spatial res.	temporal res.	time span	missing years
IRS satellite	24m	yearly	**Nilgiris**	
			1996-2016	2003, 2008 (2010, 2011 no fire years)
			Sathyamangalam	
			1997-2016	2008 (2014 no fire years)
			Uttara Kannada	
			1999-2016	2003, 2008 (2010, 2015 no fire years)
TerraClimate	~ 5000m	monthly	**available for all landscapes**	
			1997-2017	-

Table 1: Climate and burned area datasets characteristics.

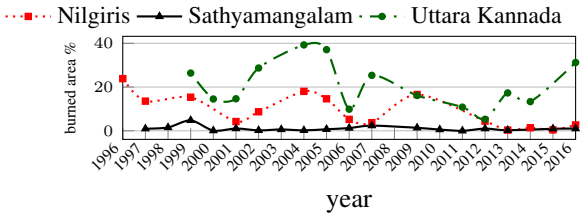


Figure 2: Burned area % across years for all landscapes.

Motivating Domain and Datasets

In this study, we focus on datasets from three landscapes in the Western Ghats mountain range in India. These landscapes include Nilgiris, Sathyamangalam and Uttara Kannada, which cover 1545, 1495 and 2266 sq. km.,

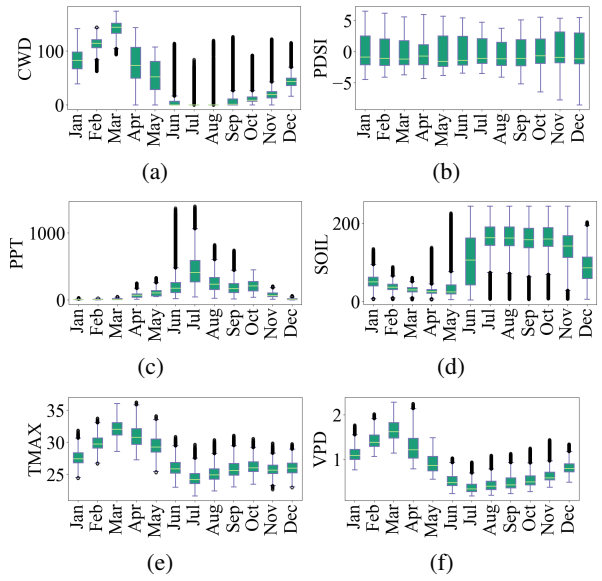


Figure 3: Spatio-temporal distribution of climate features across all months, Nilgiris.

respectively. Sathyamangalam and Uttara Kannada are referred to as Sathya. and UK. in some of the figures and tables in this paper. In this study, we develop predictive models based on datasets studied in (Kodandapani, Cochrane, and Sukumar 2008; Kodandapani 2013) where further information about the landscapes and delineation of forest fires from the Indian Remote Sensing (IRS) satellite is provided. Using IRS satellite imagery, we obtained annual maps of the landscapes categorized as burned or unburned with 24m spatial resolution. These maps indicate burned regions if a wildfire has ever affected that region during the fire season (i.e., January to March) in a specific year.

A summary of the characteristics of the datasets including spatial and temporal resolutions, time span and missing years information for all landscapes is shown in Table 1. Figure 2 demonstrates the percentage of areas burned across different years for each landscape. According to this historical data spanning almost two decades, among the three landscapes studied in this paper, the burned areas have had the greatest extent in the Uttara Kannada landscape and the lowest extent in the Sathyamangalam landscape.

To study the relationship between climate variations and burned areas caused by wildfire incidents, we use the TerraClimate dataset (Abatzoglou et al. 2018), which produces monthly climate and water balance data at about 5km spatial resolution. In particular, we use (i) climatic water deficit (CWD); (ii) Palmer drought severity index (PDSI); (iii) soil moisture (SOIL); (iv) maximum average temperature (TMAX); (v) vapor pressure deficit (VPD); and (vi) precipitation (PPT), as the climate metrics with demonstrated links to fire activity (Williams et al. 2015; Abatzoglou and Kolden 2013). The spatio-temporal distribution of monthly climate data used as predictor features is shown in Figure 3, as an example for Nilgiris. From January to May, the monthly average of the maximum average temperature and the vapor pressure deficit are higher compared to other months, and the soil moisture is lower than other months. However, the precipitation starts increasing from April. Thus, from January to March all of these landscapes experience the highest climatic water deficit and the lowest precipitation. This observation explains the high likelihood of the wildfire occurrences and burned areas during those periods captured in the IRS dataset used in this study. Unlike other features, the spatio-temporal average of the PDSI values across different months has the least variation.

Dataset Transformation and Modeling

The major challenge with this study is the coarse temporal resolution of the datasets and thus the limited number of burned area maps available to us (i.e., less than 20 maps (time steps) for each landscape where each map represents the burned areas for the entire fire season in each year). As such, we are unable to apply vision and deep learning methods or traditional time series models to our datasets. To remedy this issue, we discretize the landscapes into small grid cells and transform the maps into a tabular format for a supervised binary classification approach as illustrated in Figure 4. Each data point (each row in the data table) represents a pair of time and location. Through this discretization scheme, wildfires temporal and spatial aspects are implicitly captured. We denote the dataset for each landscape area as $\mathcal{D} = (\mathbf{X}, \mathbf{y})$ where $\mathbf{X} \in R^{TN \times f}$ is a matrix of f predictor features recorded at each of these T discrete time steps and N locations. Each time step $t \in [T]$ is one year and each location $n \in [N]$ is a small region in the landscape that we define through applying a grid of size N on the entire area for the discretization of the spatial dimension. $\mathbf{y} \in \{0,1\}^{TN}$ denotes the observation vector associated with all data points. Note that we assume the maps delineated from IRS data showing the burned areas are indicative of the locations affected by the wildfire incidents. $\mathbf{y}_{t,n}$ for $t \in [T], n \in [N]$ equals 1 if that location is marked as burned in that time step in the original maps. The f predictor features consist of monthly and aggregated climate data along with the geospatial data. The aggregated climate data is computed on a 3-month and 6-month basis as demonstrated in Figure 5.

To create the data points for a supervised learning approach, we associate the fire observations in each year to the climate data from the previous year, to assure that our models can be used for the wildfire risk prediction in future times based on the "existing climate data (i.e., predictor features)". In other words, for each specific year, we use the individual and aggregated climate data (i.e., CWD, PDSI, PPT, SOIL, TMAX, VPD) from the **previous year** as part of the predictor features for the risk of fire in the **following year**. As an example, the entire monthly climate data from January to December of 2015 along with the aggregated climate data are used as the predictor features X that corresponds to the binary observations of fire y during the fire season in 2016. To boost the predictor feature set and incorporate more spatial information, we

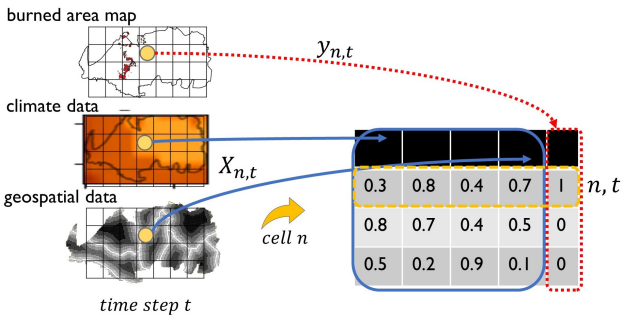


Figure 4: Spatio-temporal data transformation schema; reshaping data from map to tabular format via discretization.

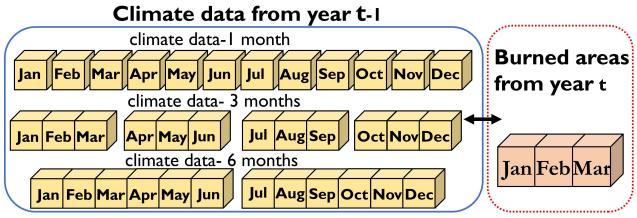


Figure 5: Data modeling approach: For each cell in each landscape, climate data from previous year t-1 and other static features are used as covariates corresponding to the binary observations of burned areas during the fire season in the current year t. This data point generation scheme holds for all locations N and all years T in $\mathcal{D}$.

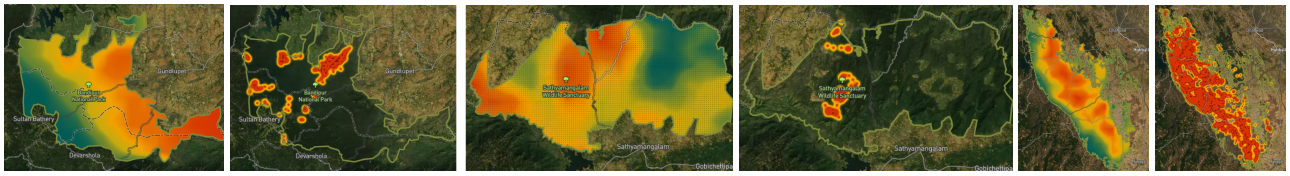
also augment the datasets with geospatial data by including two distance features, i.e., distance to road (as illustrated in Figure 4 where darker shades show further distances to roads within the landscape) and distance to boundaries to incorporate the human accessibility characteristics of each location since many of the wildfires are due to human mistakes (Kodandapani, Cochrane, and Sukumar 2008). The geospatial data are static and they remain similar across different years. We include 110 features in total.

We conduct our analysis on 500m x 500m granularity to define the size of the grid cell $n \in [N]$. The burned areas and climate data are smoothly projected from 24m and 5000m spatial resolutions to 500m resolution. We did not use a 24m resolution to mitigate the extreme imbalance between positive and negative classes in the dataset. The geospatial data are also created based on 500m resolution from area boundary and roads maps for each landscape to capture spatial information.

For each landscape and for each testing year, we train and assess models' performances, separately. Thus, to train and test the predictive models, we propose to divide the data into two parts considering the temporal dimension—this approach is proposed to overcome the shortcoming of the previous work where they randomly divide the data without any temporal considerations. We show the results for two cases here to examine the reliability and consistency of the results and we compare it with the approach proposed by prior studies to highlight the caveats with their assumptions. For **case I**, we train all of the machine learning models based on data from the first available year to 2015 and evaluate the models on an unseen test dataset, which is the entire data of 2016. For **case II**, we train based on data from the first available year to 2014 and evaluate the models on an unseen test dataset, which is the entire data of 2015 for the Nilgiris and Sathyamangalam landscapes. Note that since the data from 2015 is not available for the Uttara Kannada landscape, we train the model on the data from 1999 to 2013 and evaluate it on an unseen test dataset, which is the entire data of 2014. Our temporal division of the data results in significantly imbalanced training and testing sets and adds another challenge to this study. We will demonstrate later that this evaluation scheme is a more difficult problem compared to the random partitioning of the entire dataset proposed by previous studies due to the stochastic and spatio-temporal nature of the fires. Through the above proposal for data transformation, we aim to take into account the practical and realistic challenges associated with the future wildfires prediction based on the limited data to assist the forest and fire departments to manage resources in the real fields via exploiting the power of the historical data. Such practical aspects require attention from the Artificial Intelligence for Social Impact community.

Evaluation Results and Discussions

Predictive Models: To train the machine learning models, we conduct a 10-fold cross-validation on the training set over a range of possible parameters for each type of model. According to Figures 7 and 2, the distribution of the data varies significantly across different years. Thus, while our test sets consist of the entire data of their corresponding year (i.e., entire data of that year completely unseen to the training set), we allow the cross-validation step to be done based on the random sampling (rather than creating validation sets on a yearly basis) to minimize the impact of over-fitting based on a specific year in the training phase. We present results for both linear models (i.e., logistic regression denoted as LR) and nonlinear models, including SVM classifiers (SVC) and the tree-based techniques. For the tree-based models, we show results for a single decision tree (DT) as well as methods for bagging and boosting of decision trees including XGBoost (XGB), bagging ensemble of decision trees (BDT), gradient boosting (GRB) and random forest (RF). We also present the predictive performance of a simple but almost effective approach, which only uses the burned area observation labels to provide the predicted likelihood of future fires. We denote this approach as TA in the tables, which stands for Temporal Average (the average number of times that each location has been burned across the years of the study in the training set). Given that the burned area labels, which indicate fire observations, are binary, the outcome is between 0 and 1 for this TA approach. In the following equation $p(n)$ indicates the probability of wildfire observation in location (or grid cell) n. $p(n) = \sum_{t=0}^{T}, \frac{y_{t,n}}{T} \; \forall n \in [N]$.



(a) Nilgiris Pred. (b) Nilgiris GT (c) Sathya. Pred. (d) Sathya. GT (e) UK. Pred. (f) UK. GT

Figure 6: Visual comparison between predictions (Pred.) of LR and ground truth (GT) data in 2016 at 500m granularity.

	Test Case I- 500m							
Landscape	**Nilgiris-Train 1996-2015, Test 2016**							
Model	**TA**	**LR**	**RF**	**XGB**	**DT**	**BDT**	**GRB**	**SVC**
AUC	0.57	**0.54**	0.59	0.52	0.51	0.67	0.52	0.53
A-Prec.	0.033	0.03	0.046	0.039	0.027	0.053	0.027	0.028
F1	0.06	0.07	0.07	0.06	0.05	0.1	0.05	0.07
Landscape	**Sathyamangalam-Train 1997-2015, Test 2016**							
Model	**TA**	**LR**	**RF**	**XGB**	**DT**	**BDT**	**GRB**	**SVC**
AUC	0.64	**0.77**	0.69	0.32	0.6	0.38	0.66	0.7
A-Prec.	0.02	0.022	0.017	0.007	0.014	0.007	0.015	0.027
F1	0.07	0.04	0.03	0.01	0.04	0.01	0.03	0.02
Landscape	**Uttara Kannada-Train 1999-2015, Test 2016**							
Model	**TA**	**LR**	**RF**	**XGB**	**DT**	**BDT**	**GRB**	**SVC**
AUC	**0.79**	0.63	0.55	0.59	0.48	0.45	0.58	0.53
A-Prec.	0.56	0.43	0.34	0.38	0.32	0.27	0.39	0.33
F1	0.61	0.46	0.43	0.45	0.34	0.35	0.44	0.38

Table 2: Comparing all models' performances, grid cells granularity 500m, first test case.

	Test Case II- 500m							
Landscape	**Nilgiris-Train 1996-2014, Test 2015**							
Model	**TA**	**LR**	**RF**	**XGB**	**DT**	**BDT**	**GRB**	**SVC**
AUC	0.28	**0.74**	0.27	0.33	0.28	0.38	0.28	0.12
A-Prec.	0.004	0.007	0.003	0.004	0.003	0.003	0.003	0.002
F1	0	0.02	0	0	0	0	0	0
Landscape	**Sathyamangalam-Train 1997-2014, Test 2015**							
Model	**TA**	**LR**	**RF**	**XGB**	**DT**	**BDT**	**GRB**	**SVC**
AUC	0.53	**0.54**	0.34	0.53	0.53	0.3	0.53	0.51
A-Prec.	0.01	0.013	0.009	0.01	0.01	0.009	0.01	0.01
F1	0.03	0.02	0.01	0.03	0.03	0.01	0.03	0.54
Landscape	**Uttara Kannada-Train 1999-2013, Test 2014**							
Model	**TA**	**LR**	**RF**	**XGB**	**DT**	**BDT**	**GRB**	**SVC**
AUC	**0.82**	0.83	0.77	0.82	0.78	0.8	0.81	0.82
A-Prec.	0.37	0.4	0.32	0.39	0.3	0.32	0.39	0.39
F1	0.42	0.45	0.39	0.43	0.41	0.42	0.43	0.44

Table 3: Comparing all models' performances, grid cells granularity 500m, second test case.

	Train 75% - Test 25%						
Landscape	**Nilgiris**						
Model	**LR**	**RF**	**XGB**	**DT**	**BDT**	**GRB**	**SVC**
AUC	0.84	0.95	0.94	0.81	0.95	0.93	0.88
A-Prec.	0.33	0.7	0.65	0.43	0.7	0.61	0.43
F1	0.35	0.55	0.53	0.34	0.57	0.5	0.39
Landscape	**Sathyamangalam**						
Model	**LR**	**RF**	**XGB**	**DT**	**BDT**	**GRB**	**SVC**
AUC	0.92	0.94	0.97	0.77	0.95	0.91	0.79
A-Prec.	0.24	0.49	0.44	0.16	0.49	0.3	0.052
F1	0.08	0.15	0.16	0.05	0.15	0.09	0.05
Landscape	**Uttara Kannada**						
Model	**LR**	**RF**	**XGB**	**DT**	**BDT**	**GRB**	**SVC**
AUC	0.81	0.91	0.88	0.8	0.91	0.88	0.79
A-Prec.	0.48	0.74	0.66	0.52	0.74	0.64	0.52
F1	0.52	0.66	0.61	0.55	0.66	0.61	0.54

Table 4: Comparing all models' performances based on random partitioning of the dataset.

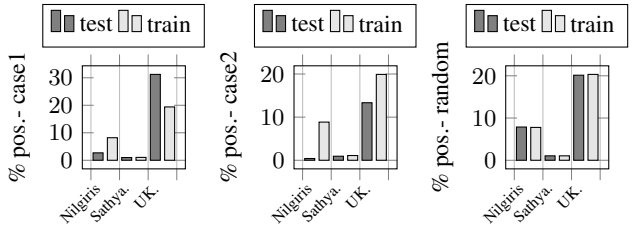


Figure 7: Imbalance across observation labels for different partitioning approaches. Case 1 & 2: temporal division; random: random partitioning with no time consideration.

Predictive Performance Analysis: To evaluate the predictive power of the machine learning models, we report standard metrics including AUC, average precision (denoted as A-Prec. in tables) and F1. Table 2 and Table 3 demonstrate results for the two test case scenarios discussed before, where each test set is formed based on the future year's data and consists of the data points that are temporally disjointed with the ones in the training set. Table 4 shows the results for random partitioning of the dataset proposed by the existing literature highlighted in the Related Work section. For each landscape, the train and test set information are shown above the corresponding part of the table. The train/test set size for Nilgiris, Sathyamangalam and Uttara Kannada are as follows, respectively. (i) test case 1: (80262, 6174), (95584, 5974) and (108468, 9039); (ii) test case 2: (74088, 6174), (89610, 5974) and (99429, 9039); and (iii) for random partitioning of the data: (64827, 21609), (76168, 25390) and (88130, 29377). The training and testing of the models based on the random partitioning outperforms the temporal partitioning approach as obviously demonstrated in Table 4, compared to Tables 2 and 3. However, most notably, this good performance does not reflect properly on the forecasting capability of the models if they are deployed in the real field to improve future decision making approaches. In real-world scenarios like wildfire prediction, historical data used as a training sample does not perfectly represent the entire event subject to modeling, due to the presence of confounding variables.

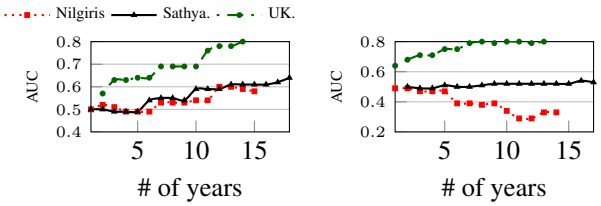
Figure 7 shows that when we do splits based on time (i.e., leftmost and middle plots), the training set class imbalance and test set class imbalance are significantly different due to the stochastic nature of the natural hazards across years. However, for evaluation based on random division, the class imbalance remains nearly similar across training and test

sets. This even distribution of data between training and test sets in the latter case, partly explains the good performance of the models trained and evaluated based on random division. Since none of future years data is available to the model during the training phase, such good performance of the latter case does not hold in the wild where stakeholders would use ML models for decision making for future years. Thus, evaluation based on random division can be unrealistic and misleading in real-world situations. Moreover, we observed that for the random division of data, nearly all of the machine learning models perform well. However, their performances vary dramatically when data is divided based on time, which is due to over-fitting of the more complex models on the training set data and stochastic nature of the wildfire incidents across both spatial and temporal dimensions.

Even though some complex nonlinear models outperform less complex ones in some cases (e.g., see RF for the Nilgiris landscape, the test set of 2016 in Table 2), their performance fluctuates dramatically and they do not demonstrate consistency across different test sets (e.g., RF performs extremely poorly with AUC of 0.27, which is even worse than the random prediction with AUC of 0.5 for the Nilgiris landscape test set of 2015, in Table 3).

In Figure 8, we show the performance of the temporal averaging model vs. the number of previous years data available to compute probabilities by TA. For Uttara Kannada, the performance improves as we increase the number of years for both test cases. However, for Sathyamangalam and Nilgiris, the AUC remains mostly around or below the performance of random guess (i.e., AUC of 0.5). Based on AUC values reported in Table 2 and 3, for Nilgiris landscape, LR is the only model, which results in performance above random guess across both test cases. TA's performance is not consistent across both test cases and it varies significantly based on the extent of the historical data available (i.e., number of past years as shown in Figure 8). Similarly for the Sathya. landscape, LR outperforms all other models. TA's performance is also below LR for both test cases. Unlike Nilgiris and Sathya. landscapes where predictor features provide additional benefits for the UK. landscape, the TA model is outperforming all machine learning models, thus using predictor features and ML approaches do not provide additional predictive power. As shown in Figure 8, as we increase the number of past years used in TA, the AUC improves consistently as well.

We emphasize the performance consistency in this study since the main goal of this paper is to develop a practical and reliable tool for decision-makers and stakeholders to leverage machine learning prediction to improve prevention, detection, and suppression of wildfires in an efficient way for future years. Figure 6 shows the prediction of the LR model in 2016 along with the ground truth map of the wildfire that occurred in 2016 for all three landscapes. The visual inspection of the results demonstrates that even though it is challenging to pinpoint the location of future fires perfectly, there exists a significant overlap between the fire locations in the ground truth maps and machine learning predictions. Therefore data-driven approaches have the potential to



(a) AUC, Test case I (b) AUC, Test case II

Figure 8: AUC vs. length of the previous-year data is used to compute the temporal average predictions for all landscapes.

provide insights for decision-makers to improve their future actions in the field. Our modeling framework and evaluation scheme can be used for any landscapes where historical burned areas or wildfire incidents are available.

Conclusion

Wildfire incidents can result in devastating outcomes. While predicting this phenomenon is extremely challenging, we demonstrated that the historical data and machine learning models provide important insights to the stakeholders and decision-makers to direct their resources more efficiently. Such proactive actions can help minimize potential damage caused by the ignition and the spread of future fires. Notably, wildfires are stochastic events with a spatio-temporal nature. The evaluation of predictive models requires meticulous assessments to assure reliability and consistency for practical deployment in the field. We demonstrated that the independent and identically distributed assumption of random variables in the evaluation phase, proposed by prior studies, may lead to a misleading and unrealistic performance analysis. A successful deployment scale-up in the field depends on the level and extent of active collaborations between stakeholders (e.g., forest and fire departments) and the AI community. We believe that deployment challenges vary across different regions of the world and depend on resources and constraints that concern local forest managers. Thus, the AI solutions and the data preparation pipeline may need modifications in each new location to incorporate realistic constraints, data, and resource availabilities into AI solutions to advance over current practices. We also believe data-driven solutions should be used intelligently for reliable resource allocations with the human expert in the loop to monitor, assess and reiterate to ensure a positive impact in the long run.

Acknowledgments. We gratefully acknowledge the support of: Council of Scientific and Industrial Research, Government of India, Forest Departments of Karnataka, Kerala, and Tamil Nadu, and the Indian Space Research Organization.

References

Abatzoglou, J. T.; Dobrowski, S. Z.; Parks, S. A.; and Hegewisch, K. C. 2018. TerraClimate, a high-resolution

global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific data* 5: 170191.

Abatzoglou, J. T.; and Kolden, C. A. 2013. Relationships between climate and macroscale area burned in the western United States. *International Journal of Wildland Fire* 22(7): 1003–1020.

Abatzoglou, J. T.; and Williams, A. P. 2016. Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences* 113(42): 11770–11775.

Alonso-Betanzos, A.; Fontenla-Romero, O.; Guijarro-Berdiñas, B.; Hernández-Pereira, E.; Andrade, M. I. P.; Jiménez, E.; Soto, J. L. L.; and Carballas, T. 2003. An intelligent system for forest fire risk prediction and fire fighting management in Galicia. *Expert systems with applications* 25(4): 545–554.

Balch, J. K.; Bradley, B. A.; Abatzoglou, J. T.; Nagy, R. C.; Fusco, E. J.; and Mahood, A. L. 2017. Human-started wildfires expand the fire niche across the United States. *Proceedings of the National Academy of Sciences* 114(11): 2946–2951.

Bowman, D. M.; Balch, J. K.; Artaxo, P.; Bond, W. J.; Carlson, J. M.; Cochrane, M. A.; D'Antonio, C. M.; DeFries, R. S.; Doyle, J. C.; Harrison, S. P.; et al. 2009. Fire in the Earth system. *science* 324(5926): 481–484.

Castelli, M.; Vanneschi, L.; and Popovič, A. 2015. Predicting burned areas of forest fires: an artificial intelligence approach. *Fire ecology* 11(1): 106–118.

Catry, F. X.; Rego, F. C.; Bação, F. L.; and Moreira, F. 2010. Modeling and mapping wildfire ignition risk in Portugal. *International Journal of Wildland Fire* 18(8): 921–931.

Cheng, T.; and Wang, J. 2008. Integrated spatio-temporal data mining for forest fire prediction. *Transactions in GIS* 12(5): 591–611.

Coffield, S. R.; Graff, C. A.; Chen, Y.; Smyth, P.; Foufoula-Georgiou, E.; and Randerson, J. T. 2019. Machine learning to predict final fire size at the time of ignition. *International Journal of Wildland Fire* .

Cortez, P.; and Morais, A. 2007. A data mining approach to predict forest fires using meteorological data. In *Proceedings of the 13th EPIA*. Portuguese Conference on Artificial Intelligence.

De Vasconcelos, M. P.; Silva, S.; Tome, M.; Alvim, M.; and Pereira, J. C. 2001. Spatial prediction of fire ignition probabilities: comparing logistic regression and neural networks. *Photogrammetric engineering and remote sensing* 67(1): 73–81.

Jaafari, A.; Zenner, E. K.; Panahi, M.; and Shahabi, H. 2019. Hybrid artificial intelligence models based on a neuro-fuzzy system and metaheuristic optimization algorithms for spatial prediction of wildfire probability. *Agricultural and forest meteorology* 266: 198–207.

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. *arXiv preprint arXiv:2003.00646* .

Kodandapani, N. 2013. Contrasting fire regimes in a seasonally dry tropical forest and a savanna ecosystem in the Western Ghats, India. *Fire Ecology* 9(2): 102–115.

Kodandapani, N.; Cochrane, M. A.; and Sukumar, R. 2008. A comparative analysis of spatial, temporal, and ecological characteristics of forest fires in seasonally dry tropical ecosystems in the Western Ghats, India. *Forest Ecology and Management* 256(4): 607–617.

Özbayoğlu, A. M.; and Bozer, R. 2012. Estimation of the burned area in forest fires using computational intelligence techniques. *Procedia Computer Science* 12: 282–287.

O'Connor, C. D.; Calkin, D. E.; and Thompson, M. P. 2017. An empirical machine learning method for predicting potential fire control locations for pre-fire planning and operational fire management. *International journal of wildland fire* 26(7): 587–597.

Rodrigues, M.; and de la Riva, J. 2014. An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling & Software* 57: 192–201.

Safi, Y.; and Bouroumi, A. 2013. Prediction of forest fires using artificial neural networks. *Applied Mathematical Sciences* 7(6): 271–286.

Sakr, G. E.; Elhajj, I. H.; and Mitri, G. 2011. Efficient forest fire occurrence prediction for developing countries using two weather parameters. *Engineering Applications of Artificial Intelligence* 24(5): 888–894.

Salehi, M.; Rusu, L. I.; Lynar, T.; and Phan, A. 2016. Dynamic and robust wildfire risk prediction system: an unsupervised approach. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 245–254. ACM.

Stojanova, D.; Panov, P.; Kobler, A.; Dzeroski, S.; and Taskova, K. 2006. Learning to predict forest fires with different data mining techniques. In *Conference on data mining and data warehouses (SiKDD 2006), Ljubljana, Slovenia*, 255–258.

Thomas, D.; Butry, D.; Gilbert, S.; Webb, D.; and Fung, J. 2017. The costs and losses of wildfires. *Spec. Publ. NIST SP-1215* .

Williams, A. P.; Seager, R.; Macalady, A. K.; Berkelhammer, M.; Crimmins, M. A.; Swetnam, T. W.; Trugman, A. T.; Buenning, N.; Noone, D.; McDowell, N. G.; et al. 2015. Correlations between components of the water balance and burned area reveal new insights for predicting forest fire area in the southwest United States. *International Journal of Wildland Fire* 24(1): 14–26.

Zhang, Y.; Lim, S.; and Sharples, J. J. 2016. Modelling spatial patterns of wildfire occurrence in south-eastern Australia. *Geomatics, Natural Hazards and Risk* 7(6): 1800–1815.