ADVANCES IN FOREST FIRE RESEARCH 2022

Edited by
DOMINGOS XAVIER VIEGAS
LUÍS MÁRIO RIBEIRO

Empirical fire propagation potential from a balanced dataset

Ljiljana Šerić*; Antonia Ivanda; Marin Bugarić; Darko Stipaničev

University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, Rudera Boskovica 32, 21000 Split, Croatia,
{ljiljana.seric, antonia.senta.00, marin.bugaric, darko.stipanicev}@fesb.hr

**Corresponding author*

Keywords

Fire Danger, Fire Propagation Potential, Empirical Model, Machine Learning, Classification

Abstract

In order to assess fire and wildfire risk one must address various features and analyse the danger and vulnerability aspects. Besides fire ignition probability, one of the most important variables for addressing fire danger is fire propagation potential. Fire propagation potential (FPP) can be described as a quantitative description of the circumstances under which, if fire ignites, it leads towards propagation of fire. This means that not all ignitions cause propagation of significant fires. Some ignitions are easily extinguished and pose no danger to vulnerable assets. On the other hand, some ignitions result in large and mega fires, causing large, burned areas and huge casualties. Fire propagation potential (FPP) provides quantitative distinction between these two different circumstances.

Machine learning techniques are more and more applied in fire management tools as they provide us with techniques for learning from the past data and predicting the future outcomes. Majority of previous work is focused on analysis of the large fire events, their causes and development. However, when modelling the FPP, we should consider situations on both ends of the outcome spectrum - situations when fire ignites and propagates and situations when fire ignites and does not propagate. If one uses only data on fires that propagate, without considering the alternative situations data, results that are achieved can be incomplete.

In this paper we propose a novel and more full approach to fire danger assessment by analysing situations of both cases - high and low fire danger. We simplify the value of FPP and consider that in cases the fire propagates the value of FPP is one, and zero otherwise. We used data collected from the events of both cases. We obtained a balanced dataset and trained machine learning model with a data set having representatives of both ends of the FPP spectrum.

The research is demonstrated in the study area of Split and Dalmatia County. We consider past fires that are sensed by satellite and recorded in the EFFIS system as situations when FPP had value 1. To assess the situations when FPP was 0 we analysed the fire intervention database maintained by fire departments. We filtered fire interventions related to forest fires that lasted less than 2 hours and engaged 2 or less firefighters since these records represent time and place of the fire that did not propagate.

For these two cases of events, we collected Sentinel-2 imagery and weather data that consists of temperature and wind speed. Sentinel-2 imagery pixels were extracted for the area associated with both types of events. The dataset was split into train and test datasets, where classifiers were trained by using 80% of data and 20% of remaining data was used for testing the classifier performance. Experiments were conducted by training classifiers using commonly used classifiers - Decision Tree Classifier, K-Nearest Neighbors, Multi-layer perceptron, Random Forest Classifier, Naive Bayes Classifier and Logistic regression. The best performance, according to the R2 score and RMSE is measured on Decision Tree Classifier.

1. Introduction

Assessment of fire danger is done by assessing the probability of occurrence of hazardous fire event. It is obvious that fire poses a hazard or threat in case it ignites, but not all ignitions result in a hazard. In case the fire does not propagate it is not considered dangerous. That is the reason to study the fire propagation potential (FPP) as a limiting factor for fire danger assessment.

Fire danger is studied in literature and various approaches, among which physics-based methods, statistical methods, and machine-learning methods for fire danger predictions are proposed (Pourghasemi 2020). Machine learning methods have shown itself useful for analysis of past events and predicting the future events when appropriate data is collected (Pham, 2020). Thus, in order to apply Machine learning methods we must obtain a dataset that will be used in the training phase of building the machine learning based model and select the appropriate learning algorithm.

Previous attempts of building a machine learning based model for fire danger prediction mostly relied on data of high fire danger from historical fires (Dimuccio,2011;Tehrany 2019; Pourghasemi 2020) but are neglecting the cases with low fire danger.

In this research we created a balanced dataset covering both cases - fires with burned areas and fire ignitions that did not result in spread. This dataset is used for building a machine learning classifier that distinguishes between situations with high and low FPP.

2. Materials and methods

2.1. Study area

Study area of this research is Split and Dalmatia county, a central-southern county in Croatia with the population of 455,242 (2011) and the land area 4,540 km2.

This region often suffers from large fires especially during the summer season.

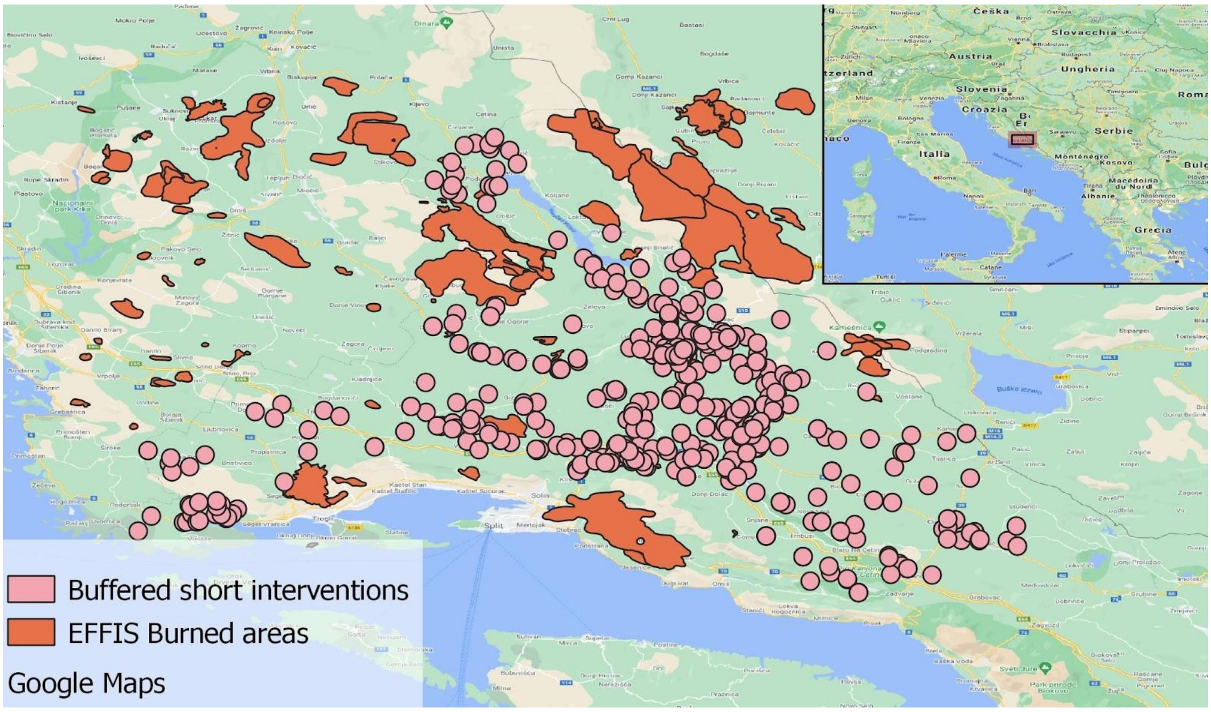

Figure 1- Study area -Split and Dalmatia County with burned areas and locations of interventions without fire spread we used in our research.

2.2. Dataset

In order to observe two cases on both ends of fire propagation potential values we collected data from two different sources.

First, we collected burned area data from the European forest Fire information system (EFFIS) that described burned areas in the study area from 2015 till the end of 2021.

Second dataset was fire departments intervention records from the 2017 till the end of 2021.

These two datasets were used to observe two cases:

Case 1: fire that ignites and propagates to the surrounding area. We observed such situations resulting in burned area noticed by EFFIS system.

Case 2 fire that ignites but is easily extinguished. We observed fire interventions with the description of forest fire, where intervention lasted less than 2 hour and less than 3 firefighters were engaged.

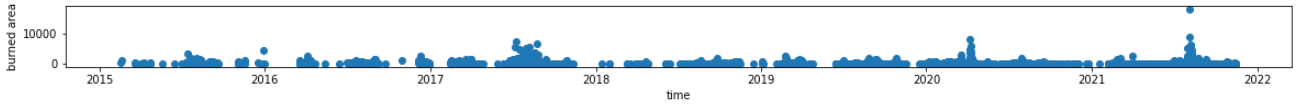

Figure 2- Temporal distribution of burned areas during the period 2015-2022

Reflecting on the fire triangle, the mostly used tool for fire behaviour understanding, propagation of fire is limited by the presence of three variables: fuel, heat, and oxygen. If ignition has happened, heat is present, thus we must determine the presence of fuel and oxygen to enable fire propagation. In our research we used two sources of data that we use as a proxy for quantifying the presence of fuel and oxygen - Sentinel-2 imagery and weather data. Sentinel-2 imagery is collected as Level-2A reflectances in 12 available wavelength bands for the study area. We also collected weather archived data for the period of interest. For each burned area we collected Sentinel-2-L2A pixels value with 60m resolution for each of the 12 bands, representing atmospherically corrected surface reflectances of the 12 wavelength bands, from the imagery taken between 1 and 3 days before the fire.

For the Case 2 events, interventions were described not as an area but as the single point. We collected the same Sentinel-2-L2A pixel values with 60m resolution for the 1 km buffer area surrounding the ignition incident.

This resulted with a dataset each row representing a 60x60m pixel described with 12 values of reflectances for each of the Sentinel-2 band at the time of the event, air temperature for the date of the event, wind speed at the time of the event and value 0 or 1 indicating whether the pixel is burned or not burned.

48604 data items representing burned pixels, and 216376 pixels representing non burned pixels, with a total of 264980 items. However, this dataset is not balanced. We used a common balancing method - down sampling to create a balanced dataset that has a similar number of representatives of both classes. We randomly selected only 25% of majority class samples and created a balanced dataset that consists of total 102698 data items.

B01.tif	B02.tif	B03.tif	B04.tif	B05.tif	B06.tif	B07.tif	B08.tif	B09.tif	B11.tif	B12.tif	B8A.tif	temp	wspd	burn
0.157250	0.231404	0.342539	0.389541	0.585679	1.000000	1.000000	1.000000	1.000000	1.000000	0.662658	1.000000	26.250000	4.000000	1
0.254563	0.214879	0.321657	0.397304	0.686704	0.979826	1.000000	1.000000	1.000000	1.000000	0.851914	1.000000	28.750000	12.750000	1
0.259033	0.302021	0.375166	0.541772	0.635147	0.760511	0.828962	0.901695	0.974500	1.000000	0.977706	0.922368	12.333333	4.666667	0
0.126821	0.193180	0.231912	0.311580	0.462847	0.622117	0.726232	0.784196	0.865924	1.000000	0.760598	0.846687	13.500000	17.500000	0
0.256889	0.237518	0.282364	0.302973	0.405287	0.606758	0.695203	0.741671	0.907718	0.937644	0.646878	0.794378	3.000000	17.333333	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Figure 3- A sample of the balanced dataset used for classification

2.3. Classification Algorithms

The dataset described in the previous section was randomly splitted into two subsets - train dataset consisting of 80% of the data and test dataset consisting of the remaining 20% of the data. We used the train set to train a classifier that would predict which case the pixel belongs to - case 1 - the pixel will burn if fire ignites, or case 2 the fire will not propagate over the pixel.

The following machine learning algorithm were tested for classifying Case 1 and Case 2 pixels

- Logistic regression
- Naive Bayes Classifier
- K-Nearest Neighbours with 3 neighbours
- Decision Tree Classifier
- Random Forest Classifier with maximum tree depth 5
- Multilayer perceptron with 10 hidden layers

For each classifier we calculated two evaluation measures - R-squared - R^2 score (Chicco, 2021) and Root Mean Square Error (RMSE) (Jierula, 2021). The R^2 and RMSE were calculated both on train and test set in order to check for overfitting.

3. Results

The selected classifiers were trained to predict two classes of propagation potential based on Sentinel 2 imagery and weather features. The classifiers performance measures, as measured on our dataset are shown in Table 1.

Table 1. Performance evaluation measures of classifiers used in the experiment

ML ALGORITHM	TRAIN R^2 SCORE	TEST R^2 SCORE	TRAIN RMSE	TEST RMSE
K-Nearest Neighbors	0.999528	0.999056	0.021728	0.030727
Decision Tree Classifier	1.000000	0.999056	0.000000	0.030727
Multi Layer Perceptron	0.990676	0.990244	0.096559	0.098773
Random Forest Classifier	0.943704	0.938159	0.237269	0.248679
Logistic Regression	0.835045	0.828639	0.406147	0.413958
Gaussian Naive Bayes	0.839293	0.828009	0.400882	0.414717

Ideal value of R^2 measure should be 1. The closer the R^2 value to 1, better the performance of the classifier. We can see that tested algorithms predict the propagation potential that is corelated with our label. The RMSE value should be as small as possible. All tested classifiers have small value of RMSE.

4. Conclusion

From the results shown in Table 1 we can observe that non-linear classifiers result in better performance on our data set. Thus, we can conclude that the relationship between fire propagation potential and land surface reflectances and weather is not linear and straightforward, but rather complex. This was expected since we did not consider different types of land cover, vegetation, and roads, but merely used Sentinel-2 imagery as raw input of low-level land cover features. The resulting empirical propagation potential classifier gives promising results and performs well on test dataset. In future work we will investigate the results on larger study area and cross compare the results on areas with different climate and topographic features. Also, in future work we will investigate the usage of semantic data about the land and weather features to obtain a more general model.

5. References

Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci. 2021 Jul 5;7:e623. doi: 10.7717/peerj-cs.623.

Copernicus Sentinel data 2015, processed by ESA.

Dimuccio, Luca Antonio, et al. "Regional forest-fire susceptibility analysis in central Portugal using a probabilistic ratings procedure and artificial neural network weights assignment." International Journal of Wildland Fire 20.6 (2011): 776-791.

Jierula A, Wang S, OH T-M, Wang P. Study on Accuracy Metrics for Evaluating the Predictions of Damage Locations in Deep Piles Using Artificial Neural Networks with Acoustic Emission Data. Applied Sciences. 2021; 11(5):2314. https://doi.org/10.3390/app11052314.

Pourghasemi, Hamid Reza, et al. "Application of learning vector quantization and different machine learning techniques to assessing forest fire influence factors and spatial modelling." Environmental research 184 (2020): 109321.

Tehrany, Mahyat Shafapour, et al. "A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data." Theoretical and Applied Climatology 137.1 (2019): 637-653.

Pham, Binh Thai, et al. "Performance evaluation of machine learning methods for forest fire modeling and prediction." Symmetry 12.6 (2020): 1022.

Weather Underground - https://www.wunderground.com/