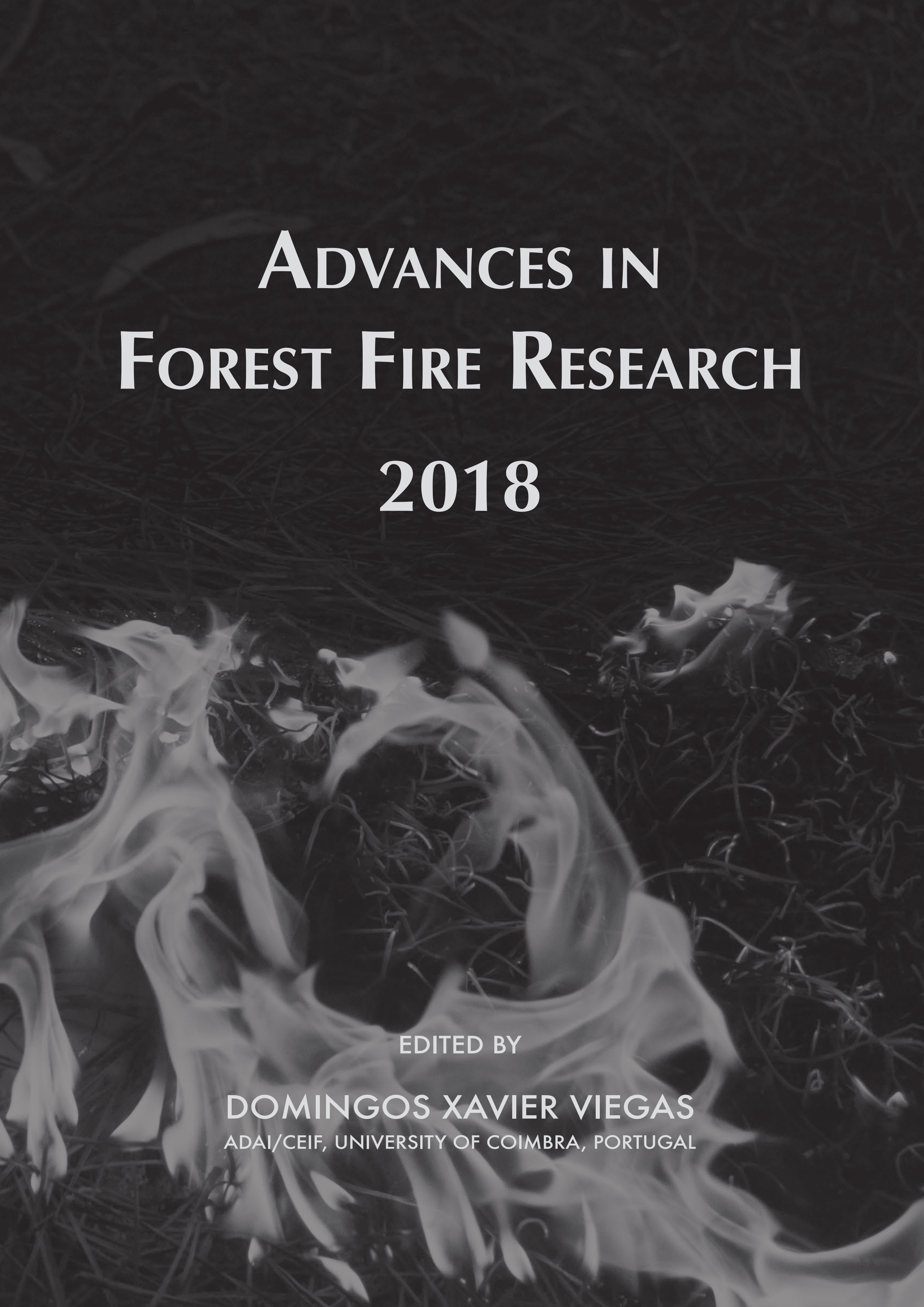


Advances in Forest Fire Research 2018

EDITED BY

DOMINGOS XAVIER VIEGAS
ADAI/CEIF, UNIVERSITY OF COIMBRA, PORTUGAL

https://doi.org/10.14195/978-989-26-16-506_112

Feature Selection for Burned Area Classification in the Castelo de Paiva Region

Carlos Viegas Damásio*[1]; Lourdes Bugalho[2]; Pedro Medeiros[1]; Susana Nascimento[1]; Luís Duque Santos[1]

[1]*NOVA LINCS, Departamento de Informática, FCT/UNL, Portugal, {cd@fct.unl.p*t}*

[2]*Instituto Português do Mar e da Atmosfera, Portugal*

Abstract

Burned area mapping is a fundamental activity for the study of wildfires, but good quality data requires long in situ campaigns performed by specialized personnel. The use of machine learning algorithms and the availability of high-quality remote sensing data bring new opportunities to make the process more precise and expedite. This paper elicits the best combinations of features to be used in machine learning pixel-based burned area mapping, obtained by LASSO regression. The feature selection is applied to ground-truth obtained in the region of Castelo de Paiva, Portugal, in 2016.

The selected number of features is 9, from an initial set of 51 features, and includes besides a new index a mix of prefire, postfire as well delta differences of spectral indices (namely dNBR). The selected features can be used with Sentinel 2, Landsat, and MODIS imagery.

In parallel, the identified features have been fed to several classifiers, namely a multilayer neural network, gradient boost and extreme gradient boosting, support vector machines and K-nearest neighbour classifiers to validate the choice performed by the LASSO regression. Accuracies of 96% and kappa of 0.86, are consistently obtained for Sentinel 2 imagery, and Landsat 8 also scores very well; as expected, MODIS results in score reduction to the coarser spatial resolution. Confusion maps are presented to visualise the quality of the obtained results, as well as to pinpoint existing problems with the ground-truth.

Extreme Gradient Boosting shows to combine very high classification metrics with very efficient processing, possible via the Graphics Processing Units (GPU) implementation of XGBoost. A simple processing architecture is proposed to support an automatic classification system based on the publicly available satellite imagery, supported by the benchmarks obtained in this paper. A set of Sentinel 2 granules covering Portugal can be processed in less than 10 minutes in consumer hardware, for each satellite passage.

Keywords: Remote Sensing, Burned Area Indices, Machine Learning, Sentinel 2, Landsat 8, MODIS

1. Introduction

Portugal is a country with a recurrent history of wildfires. Every year thousands of hectares are burned as result of wildfires running rampant. These events have greater intensity in the center and northern regions of the country. A timely and accurate assessment of burned areas is important to have reliable data that can be used at the planning stage, and even at the operational level. The existence of ground-truth for the entire country provided by the ICNF (Instituto da Conservação da Natureza e das Florestas), as well as the availability of public remote sensing date fosters the study of machine learning classifiers/techniques to provide better identification of burned areas for the Portuguese territory. However, the use of pixel-based techniques requires the identification of the appropriate set of features both to reduce time as well as overfitting, and this the main concern of this paper.

Presently there are also other repositories focussed on fire related data as for examples the European Forest Fire Information System (EFFIS), which consists on web geographic information system that provides near-real time historical information on forest fires and their regimes. This service encompasses the European, Middle Eastern and North African regions. EFFIS is composed of five modules, Fire Danger Assessment, Rapid Damage Assessment (these last two have near real time availability), Emissions Assessment and Smoke Dispersion, Potential Soil Loss Assessment, and Vegetation Regeneration (COPERNICUS 2018).

For the present work we defined as study area the Castelo de Paiva region that had significant fire activity during August 2016. The area of study is situated in northern Portugal where the Douro and Tâmega rivers intersect. The area selected has 559 Km^2 of which by the end of the year nearly 118 Km^2 had burned (circa 20% of the area). Using burned area shapefiles provided by the ICNF (Instituto da Conservação da Natureza e das Florestas) as ground-truth, and feeding a variety of spectral indices over MODIS, LANDSAT 8, and Sentinel 2 imagery to machine learning classifiers, we evaluate and obtain a very accurate burned area classifier over a selected set of those indices. In this process, we have defined a new burned area index that provides high-quality discrimination between burned and non-burned areas.

The paper is organized as follows. First, in Section 2, we present the general methodology used for this work. Section 3, presents the results regarding the best combination of features to be used in the classification approach; these features are then used to determine the best performing classifier, both in terms of classification scores and execution times. Section 4 presents the classification results and analyses them, contrasting the use of standard burned area indices with the classification approach based on the ensemble of classifiers. Section 5 describes a proposed automatic system, and discusses its feasibility supported by the results previously obtained. Section 6, closes the papers with the main conclusions.

2. Methodology

The difference between pre and post-fire imagery displays a difference between near and mid infrared values, namely, the near values are very high and the mid values are very low in pre-fire, the opposite is verified in a post-fire scenario. It is the relationship between these two bands that the several Burned Area Reflectance Classification (BARC) methods exploit. The areas with the largest differences are the best candidates for being severely burnt while remaining areas are likely to be unburned or lightly burned (USFS 2017). Indices are used to capture specific changes of the earth's surface through the use of spectral bands. The indices chosen for burned area identification in our experiments are known vegetation indices, namely Normalized Difference Vegetation Index (NDVI) (Pettorelli et al. 2005), Soil Adjusted Vegetation Index (SAVI) (Rondeaux et al. 1996), Moisture Stress Index (MSI) (Welikhe et al. 2017), and the BARC specific indices, like the Mid Infrared Burn Index (MIRBI) (Trigg et al. 2001), Burn Ratio (BR) and Normalized Burn Ratio (NBR) (Key et al. 2006), two variations (NBR2, NBR3) and a new proposed index (NBR4), as well as the differences (deltas) of these indices from the post-fire scenario to the pre-fire one. The rationale behind NBR4, is based on the fact that the green spectrum is greatly influence by fire activity, thus being able to contribute in separation between burned and unburned areas, since the areas with greater amounts of green will trend toward the lower end of the index's range and vice versa. The equations for determining these indices can be found in Figure , where Green and Red are the visible light satellite bands, NIR, LSWIR, and SSWIR are respectively the Near Infrared, Longer short wave infrared and Shorter short wave infrared spectral ranges. The exact band ranges are slightly distinct from satellite to satellite, but here those differences are ignored.

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$$

$$SAVI = (1 - L) * \frac{(NIR - Red)}{(NIR + Red + L)}, \textbf{ with L = 0.5}$$

$$MSI = \frac{SSWIR}{NIR}$$

$$MIRBI = 10 * SSWIR - 9.8 * LSWIR$$

$$BR = \frac{NIR}{LSWIR}$$

$$NBR = \frac{(NIR - LSWIR)}{(NIR + LSWIR)}$$

$$NBR2 = \frac{(2 * NIR)}{(SSWIR + LSWIR)}$$

$$NBR3 = \frac{(SSWIR - LSWIR)}{(SSWIR + LSWIR)}$$

$$NBR4 = \frac{(NIR - LSWIR)}{(((NIR + LSWIR) * Green) + 1)}$$

Figure 1 - Remote sensing indices used for burned-area classification

In this study Sentinel 2, Landsat 8 and MODIS satellite imagery were used. Sentinel 2 is characterized by its great resolution (10m x 10m) generating the same granule every 5 days. While Landsat 8 presents a three-fold reduction in resolution (30m x 30m) in comparison with Sentinel 2 and generates the same granule every 16 days. Finally, MODIS, with a 250m x 250m, yet, a daily granule generation.

The Sentinel 2 raster images from July 29th (pre-fire) and September 27th were processed with a DOS1 (Dark Object Subtraction 1) correction and then clipped to the size of the study area. Sentinel 2 was chosen due to the fact of it possessing largest quantity of samples, providing a larger amount of information about the land cover. The dataset was then compiled using all the pre and post-fire Sentinel 2 bands, indices and deltas.

This dataset was then subjected to a LASSO (Least Absolute Shrinkage Selection Operator) regression: "The LASSO is a shrinkage method like ridge, with subtle but important differences" (Hastie et al. 2008). LASSO is a regression method used in both variable selection and regularization in order to enhance the prediction accuracy of the classification process. As such, the LASSO method was used to select a small subset out of a considerably larger set of the provided features with the intent of reducing the dimension of the dataset and improving their prediction capabilities. The LASSO regression method constrains the regression coefficients sizes, encouraging the simplification of the models. This method is useful for automating feature selection and to better deal with models where the correlation levels between variables are high. Weights are the normalization factor associated with this regression. These weights, define the amount of severity applied in the “punishment” of the outliers. By using various weights to select the best features and subsequently averaging the importance attributed to each feature across all weight, a simple heuristic created. This heuristic not only makes our choice of the amount of features less biased, since the number of features is not being directly chosen, but ensures that the most useful features are used in the classification process in addition to making the model generation more efficient.

Using the selected indices as features, SVM (Support Vector Machine), Gradient Boosting (GB), Artificial Neural Networks (ANN) and K Nearest Neighbours (KNN) classifiers were trained with using different training set percentages in a pixel-based approach, and confusion maps were calculated as well as overall accuracy and kappa statistics for accuracy assessment. We will take special attention in the use of GPU-accelerated classifiers, mostly due to the magnitude of data that needs to be processed. This is the precise scenario GPUs excel, since they are optimized for mass parallel execution of operations, since the major caveat of traditional approach is that some operations take too long to execute when the dataset exceeds a certain dimension. All tests of this study were run on a laptop machine with an Intel Core i7 6700HQ, 16GB of RAM and an NVIDIA GTX950M with 4GB, running on Windows 10 version 1709.

3. Feature selection and classifier evaluation

The objective of the work reported in this section is the selection of remote sensing based features, and the corresponding most appropriate classifier to be used for burned area classification. The result is the definition of the classification setting to be used.

3.1. Feature selection

In order to perform feature selection, we applied a LASSO regression to the dataset containing all Sentinel 2 pre and post-fire scenario band data and indices, as well as the indices' variation between the sensing dates of July 29th and September 27th, 2016 (deltas), and all the original satellite bands with a total of 51 features per pixel. The LASSO regression was run several times with different values for the weight parameter, as choosing a single weight value and opting for the resulting feature selection would turn our choice biased. A solution found for the problem was to define a heuristic to help us choose the most relevant features, in this case an average. The mean value for each feature across all the different weight values being greater than zero equates to that particular feature having some relevance. As a result of the LASSO regression, the pre-fire NDVI and NBR4, the post-fire MIRBI and NBR as well as the delta values of the MSI, MIRBI, BR, NBR and NBR4 indices, were considered to be the features that better help in solving the presented classification problem. The remaining features had a mean score of zero, meaning that they are not as relevant as the ones with non-zero values. We used the Sentinel 2 imagery due to it having the best spatial resolution of the three satellites considered, which equates to a greater amount of data points to assist in finding the best combination of features for our models.

3.2. Classifier evaluation

Classification processes were run using five different classifiers for Sentinel 2 data, and for each the best training set size combinations for this dataset were collected. The criteria chosen for ranking the classifiers, by order, were Accuracy, followed by Kappa score and lastly its total execution time.

The used classifiers were off the shelf MLP, GB, and KNN, from Python's scykit-learn library, as well as XGBoost and LiquidSVM. XGBoost. XGBoost is short for "Extreme Gradient Boosting", where the term "Gradient Boosting" is proposed in the paper Greedy Function Approximation: A Gradient Boosting Machine, by Friedman 1999, and is a variant of the original Gradient Boosting algorithm implemented in scykit-learn. It has a plugin that adds the option for GPU accelerated tree construction and prediction algorithms. LiquidSVM is an implementation of SVMs whose key features are fully integrated hyperparameter selection and extreme speed on both small and large data sets. XGBoost, MLP and LiquidSVM are GPU accelerated, while GB and KNN are not.

Table 1. The top 5 classifiers for Sentinel 2 data

Classifier	Training Set (percentage)	Accuracy	Kappa	Total Time (s)
XGBoost	5%	0.96	0.86	6.68
MLP	20%	0.96	0.86	77.42
GB	10%	0.96	0.86	215.78
LiquidSVM	5%	0.96	0.86	898.94
KNN	10%	0.95	0.85	126.98

The obtained experimental results are present in Table 1 and it is clear that very high classification accuracy and kappa values are obtained, independently of the classifier. The best results for the classification process were obtained with the GPU XGBoost library (version 0.6), with 10% training data (559311 pixels) using a stratified split on the dataset in order to maintain the ratio between labels. This combination of parameters obtained an accuracy score of 0.96, a kappa score of 0.86 and a total execution time of 6.68 seconds.

The classification processes were subsequently repeated using different combinations of training set size and satellite data for the XGBoost classifier, obtaining:

- Sentinel 2 with 0.96 Accuracy and 0.86 Kappa in 6.68 seconds with 5% of sample as training data;
- Landsat 8 with 0.95 Accuracy and 0.83 Kappa in 1.41 seconds with 5% sample;
- MODIS with 0.89 Accuracy and 0.61 Kappa in 0.98 seconds with 20% sample.

Mark again that these timings were obtained with a standard consumer laptop, and therefore for professional hardware better results are expected. These findings support the use of XGBoost as a top choice classifier for burned area identification classification problems.

4. Classification Results

In order to visually analyse the resulting classifications, confusion maps were created in order to provide a visual, spatial representation of a regular confusion matrix. An example of the map for XGBoost is depicted in Figure .

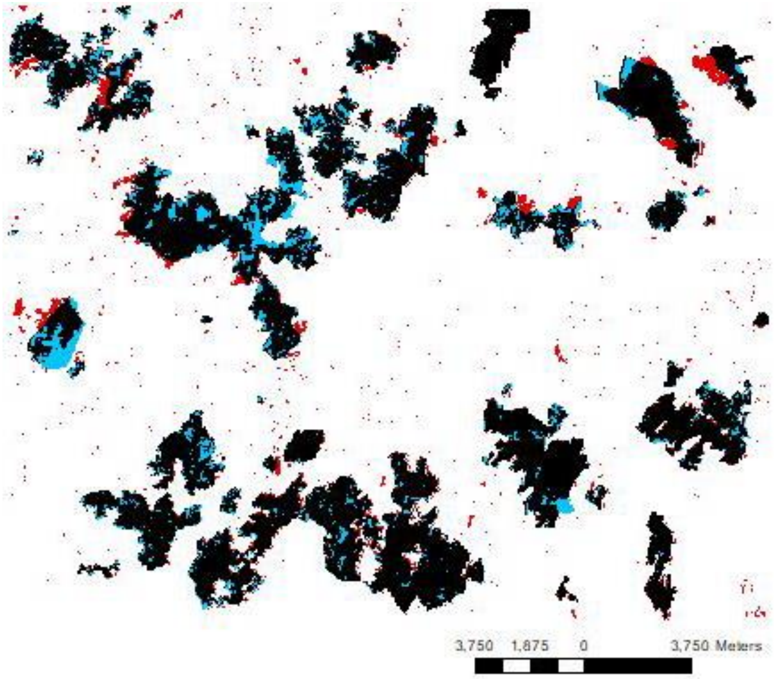
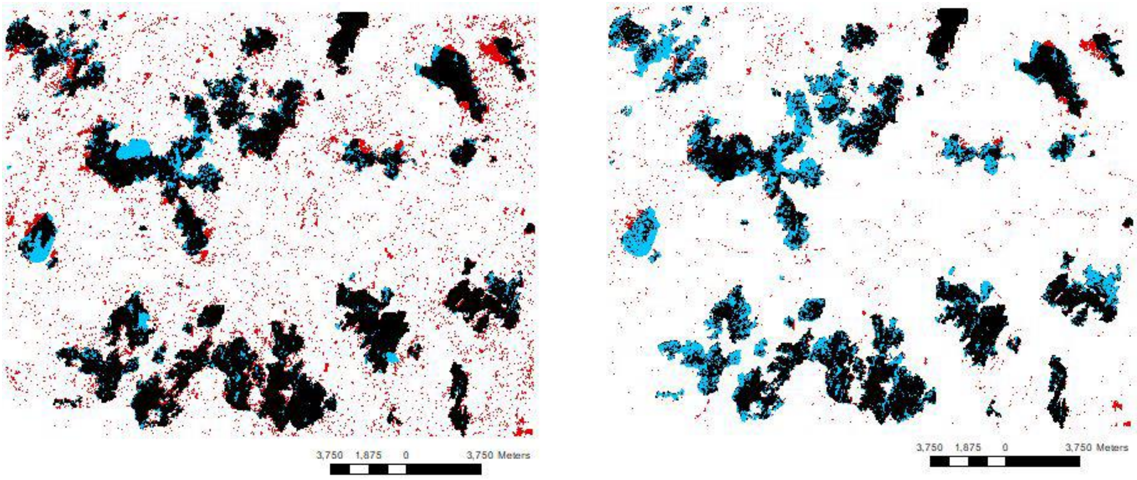


Figure 2 - Confusion map generated for the XGBoost classification with Sentinel-2 imagery. Black and red areas are the burned areas identified by the classifier, and black plus blue areas form the ground-truth

There is a very good agreement of the ground-truth with the classifier results, however some differences are clear. Blue areas correspond to burned areas marked in the ground-truth but not identified as burned by the classifier (false negatives); by visual inspection of the images these correspond mostly to villages or rock formations inside the perimeter of fire. Reversely, red areas are classified as burned according to our classifier and may correspond to errors in the ground-truth (false positives), namely due to the fact that till recently ICNF did not record rural fires. This shows that using a pixel-based machine learning approach might obtain finer burned area maps inside the perimeter of fires and show that the methods are robust to noise that may exist in the ground-truth.

To assess the behaviour of the machine learning based approach, we compared the XGBoost results to the differences (or delta) of the selected indices. For instance, the dNBR is one of the standards for burn severity classification but can be simplified in order to map burned areas, according to Lutes et al. 2006, burned areas correspond to values of dNBR ≥ 100. However, a threshold of 200 is used since it can vary ±100 and should be chosen according to the existing data (USGS 2004). Our proposed index NBR4 marks burned areas when its value is greater than zero. By using these thresholds burned area confusion maps were generated (see Figure).



a) dNBR ≥ 200 b) dNBR4 ≥ 0

Figure 3 - burned area classification confusion maps for delta indices over a threshold

The dNBR presents a moderate amount of false positives in its classification. Our dNBR4 index has far less false positives, but more false negatives concentrated in some regions. By contrasting with Figure , it can be easily concluded the superior performance of the classification-based approach.

To be able to quantify the above empiric result, the delta indices are compared to the result with the XGBoost classification, as well as the majority vote of the top 5 distinct classifiers described in Section 3. A weighted confusion map was determined, in order to observe the overall opinion of an ensemble composed by the different classifiers, voting on which category should be attributed to a specific pixel. For this intent, the Sentinel 2 dataset was chosen due its greater resolution, as well as its overall results in previous tests. The results are summarised on Table .

Table 2. Comparison of classification scores

Index	Accuracy	Kappa
dNBR	0.91	0.73
dNBR4	0.92	0.72
XGBoost	0.96	0.86
Majority	0.96	0.86

When comparing the dNBR with the dNBR4, one can verify an increase in accuracy, but kappa reduces; the exact determination of the most appropriate discriminating threshold is ongoing work.

This was expected and can be associated in the reduction of the "salt and pepper" effect present in dNBR4. The classification approaches are better, with no quantitative clear distinction between XGBoost and Majority. However, the generated confusion map based on from the ratios of the majority vote, enables us to visualize how the overall classifier ensemble consensus on the label outcome and shows the benefits of using an ensemble of classifiers. It is clear the reduction of the salt-and-pepper effect with respect to Figure and Figure .

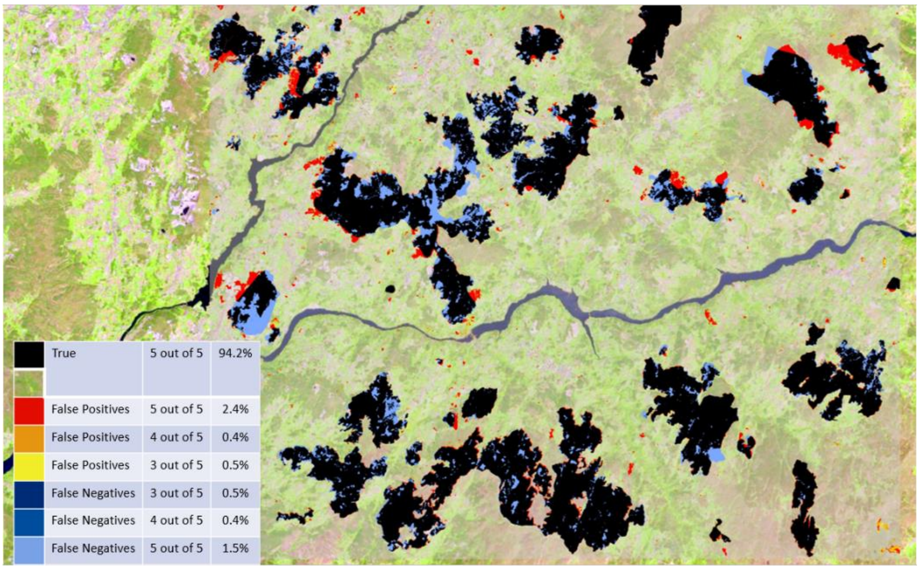


Figure 4 - Confusion map for the majority vote ensemble of classifiers

When analysing the images it became evident which areas where or where not controversial as well as some previously "hidden" terrain features became evident such as roads for example.

Figure depicts some aspects of the generated map.

5. Fire Season Classification

The purpose of this experiment is to assess the feasibility to perform burned area identification for all the days in the fire season (May 1st to October 31st), using the selected features. For this intent we defined our "pre-fire scenario" to the day before the official start of the season (April 30th). The model chosen was a XGBoost classifier trained with 30% of the MODIS dataset, referring to the 1st of August, and the following parameters: max_depth = 4, min_samples_split = 2 and n_estimators = 100. The choice of MODIS is mainly based on its daily availability of products, The differences between days were calculated using the the August 1st date and the day we intended to differentiate. In order to minimize the overall variation of the classification, this process was executed 5 times and the classification for each day was submitted to a majority vote. The metrics obtained from this process are present in Table .

Table 3 - XGBoost burned-area classification timings for seasonal MODIS data

Operation	Time (s)
Load train data	0.06
Load Fire Season Data	340.33
Total Classification Time	4119.27
Write Classifications	42.90
Total Time	4539.00

5.1. Granule Classification

Granules from Sentinel 2, Landsat 8 and MODIS were submitted to a classification process in order to obtain performance metrics. These granules where trained with a small portion of the granule and then the training classifier was used to classify the entire granule, and the obtained timings are present in Table .

Table 4 - Train and Classification times for each satellite tile/granule

Granule	Train (s)	Classification (s)
Sentinel 2	3.02	37.52
Landsat 8	1.04	20.06
MODIS	0.92	2.89

The classification time for the entire fire season can be approximately extrapolated from the classification time for each granule and the fact that 16 Sentinel 2 granules, 8 Landsat 8 tiles, and 2 MODIS tiles cover Portugal. This would equate to a total time of about 603 seconds with Sentinel 2 data, 162 seconds with Landsat 8 and 5.78 seconds with MODIS, for each satellite's passage. For half year of data the computation time required would be circa 6 hours for Sentinel 2, 32 minutes for Landsat 8, and 18 minutes for MODIS data. Note again that Table 1 timings for MODIS were obtained from 5 executions and some dates were discarded because of cloud contamination, justifying the differences of our estimation of MODIS data to the data in Table .

5.2. Architecture

The processing architecture is supported from publicly data available in the Web, and is depicted in Figure .

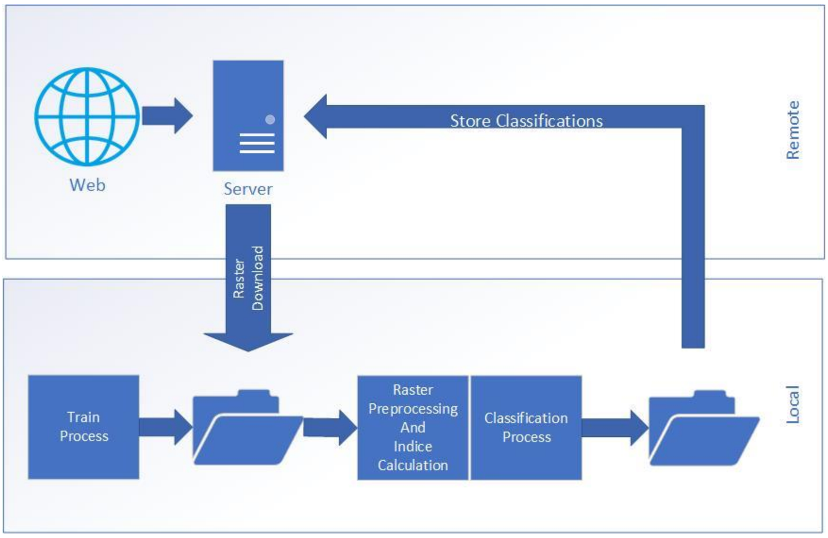


Figure 5 - Processing architecture

The system daily checks if new satellite products are available, downloads them and calculates the input indices using a bash/python script using the numpy library. The classifiers are previously trained with ICNF ground-truth over a small predefined area. This trained classifier can then be repeatedly used in the classification of the entire satellite granule. For the classification of imagery with substantial size, like Sentinel 2, the granule can be too big to fit in GPU's memory. This fact requires the granule to be split into smaller sized regions that are then classified independently since the approach is pixel-based. The classification results are then updated in a summary raster, containing the burned area since the start of the fire season period. From the results of this work, this system can be constructed with consumer hardware.

6. Conclusions

The reported experiments, show that very high classification scores can be obtained for artificial neural networks, support vector machine, gradient boosting, and k-nearest neighbour classifiers, using selected features by the LASSO regression. The gradient boosting algorithm, XGBoost, obtained excellent results for burned area classification with Sentinel 2 imagery taken in two dates, by combining pre-fire NDVI and our proposed NBR4, the post-fire MIRBI and NBR as well as the delta values of the MSI, MIRBI, BR, NBR and NBR4 indices. By employing GPU accelerated algorithms, a single Sentinel 2 granule can be classified in less than 40 seconds in a consumer laptop. This shows the appropriateness of the selected algorithm for burned area identification at the national and continental levels.

Based on the analysis of the collected metrics and scores, visual comparison between the dNBR, one of the standards in burned area assessment and the proposed index (dNBR4) its selection by the LASSO regression, makes it is clear that the reduction in the amount of misclassifications present in dNBR4 may result in more efficient burned area classifications. Yet, it still falls behind both the Majority Votes and XGBoosts in terms accuracy. More work is required to compare the proposed index with the other existing indices, for larger areas and different periods of time. This is possible, due to the existence of Landsat imagery and ICNF ground-truth, as well as the efficient classifiers identified.

An interesting outcome of the work, is the fact that the classifier are shown to be robust regarding noise in the ground-truth, and in practice almost perfect classification results may be obtained with the discussed techniques and algorithms, for the cases of burn scars visible from space. Therefore, obtaining very accurate ground-truth will be very important to validate these findings.

Finally, the proposed architecture will be made available to the community in a public web site.

7. References

Copernicus and European Committee, emergency management services, 2018

D. Lutes, R. Keane, J. Caratti, C. Key, N. Benson, S. Sutherland, L. Gangi, Integration of Standardized Field Data Collection Techniques and Sampling Design With Remote. Sensing to Assess Fire Effects. USDA Forest Service Gen Tech. Rep RMRS-GTR-164-CD. 2006

Friedman J. H., Greedy Function Approximation: A Gradient Boosting Machine, Stanford university, February 1999

Key, C., Benson, N.: Landscape Assessment (LA) Sampling and Analysis Methods, USDA Forest Service, Gen Tech. Rep, 2006

Pettorelli, N., Vik, J., Mysterud, A., Gaillard, J., Tucker, C., Stenseth, N. , Using the satellite-derived NDVI to assess ecological responses to environmental change. TRENDS in ecology and Evolution, Vol.20 No.9 September 2005

Rondeaux, G., Steven, M., Baret, F.: Optimization of Soil-Adjusted Vegetation Indices. REMOTE SENS. ENVIRON. 55:95-107 (1996)

S. Trigg and S. Flasse (2001) An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah, International Journal of Remote Sensing, 22:13, 2641-2647

T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Data Mining, Inference and Prediction, August 2009.

US Forest Service: Burned Area Reflectance Classification

USGS, The Normalized Burn Ratio (NBR) - Brief Outline of Processing Steps. FIREMON BR Cheat Sheet V4, June 2004

Welikhe, P., Essamuah, Quansah, J., Fall, S., McElhenney, W.: Estimation of Soil Moisture Percentage using LANDSAT-based Moisture Stress Index. Journal of Remote Sensing and GIS 2017, 6:2