An Approach to Classify Burned Areas Using Few Previously Validated Samples

Olga O. Bittencourt, Fabiano Morelli, Cícero A. S. Júnior, and Rafael Santos(✉)

INPE – National Institute for Space Research,
Av. dos Astronautas, 1758, São José dos Campos, SP 12227-010, Brazil
{olga.bittencourt,fabiano.morelli,cicero.junior,rafael.santos}@inpe.br
http://www.inpe.br

Abstract. Monitoring the large number of active fires and their consequences in such an extensive area such as the Brazilian territory is an important task. Machine Learning techniques are a promising approach to contribute to this area, but the challenge is the building of rich example datasets, whose previous examples are unavailable in many areas. Our aim in this article is to move towards the development of an approach to detect burned areas in regions for which there is no previously validated samples. We deal with that by presenting some experiments to classify burned areas through Machine Learning techniques that combine remote sensing data from nearby areas and it can distinguish between burned and non burned polygons with good results.

Keywords: Remote sensing · Burned forest classification keyword · Forest fire survey and monitoring

1 Introduction

Brazil recorded more than 197,000 active fires within its 851 million hectares of are between January and December of 2019 [6]. This represents an increase of around 48% compared to the number of active fires recorded throughout 2018. Among the six biomes in the country, the Cerrado (savannas and scrub forests), a biodiversity-rich region that occupies around 204 million hectares, was the most affected biome – it recorded more than 63,000 active fires in this period. It is estimated that it lost almost half of its original vegetation cover.

Monitoring this large number of active fires and their consequences in such an extensive area such as the Brazilian territory is an important task, and requires the involvement of policy makers, environmentalists and the scientific community. Several studies cover the fire-related aspects and their economic, social, and environmental impacts [3,9,12,18]. They agree with the need for reliable information about the extent and location (both in space and time) of the areas affected by the fire.

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12253, pp. 239–254, 2020.
https://doi.org/10.1007/978-3-030-58814-4_17

The National Institute for Spatial Research (INPE), part of the Brazilian Ministry of Science, Technology, Innovation and Communications, is the official institute responsible for monitoring forest fires and reporting information in Brazil. INPE's monitoring is done by remote sensing and it is developed in two independent ways, using different satellite images' resolutions. Low-resolution images (with pixels larger than 300 m) are used to generate daily data products. The active fire monitoring, for example, is done this way for the whole country. Medium-resolution images (with pixel' sizes around 30 m) are used for less frequent but more accurate studies. The burned areas' estimation is done this way, but only for the Cerrado biome. With these distinct views, it is possible to offer products such as estimation of burned area and the prediction of the risk of fire on vegetation. These products can be used to prevent, monitor, combat and create actions to analyze the impacts of burning, to estimate the emission of pollutants and to reduce the damage caused by fire. However, other biomes that also had a high number of active fires, do not have this more accurate monitoring of burned areas.

The general approach used by INPE's estimation of burned area with medium-resolution images is to compare consecutive images of the same regions. Figure 1 illustrates fragments of three consecutive remote sensing images. They are in a band combination applied to Landsat 8, displayed as a red, green, blue (RGB). Areas change over time and some are highlighted on the right side of the figure. The objective is to detect spectral changes and to determine the changes caused by fires. As part of the official data, it is necessary to present the quality of indication at least 95% of success. And, to ensure this quality, nowadays, it is necessary for manually verify the data before the publication.

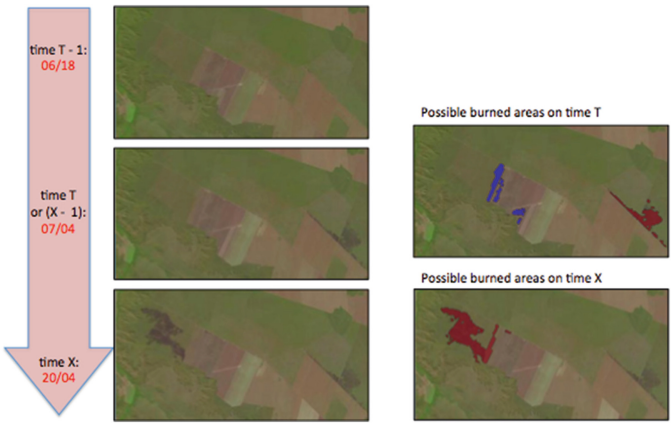


Fig. 1. Fragments of three consecutive Landsat images.

Recent studies [2,5,12] present challenges and new advances in fire monitoring. However, there is no generic automatic model for the problem of classifying burned data in continuous monitoring on a global scale, as the Brazilian

whole territory. A consensus is that to develop and evaluate more automatic approaches, it is important to have a rich knowledge database about previous occurrences of burned areas. Unfortunately, many Brazilian regions, and even biomes, have few validated burned area studies and datasets, such as Amazon and Caatinga (a dry shrub-land biome). The major challenge of this approach is the need for previously validated data from the same area – validated data is data that was evaluated by a specialist, which is very time consuming. Thus, to expand the analysis to the whole country, a new approach is essential.

Our aim in this article is to move towards the development of an approach to detect burned areas in biomes or regions for which there is no previously validated data. We deal with that by presenting some experiments to classify burned areas through Machine Learning techniques that combine remote sensing data from nearby areas and it can distinguish between burned and non burned polygons in areas without samples or with few previously validated remote sensing data. Figure 2 details some examples of polygons of the fragments highlighted in Fig. 1. The red polygons are burned examples and the blue polygons, caused by crops, are non burned areas.

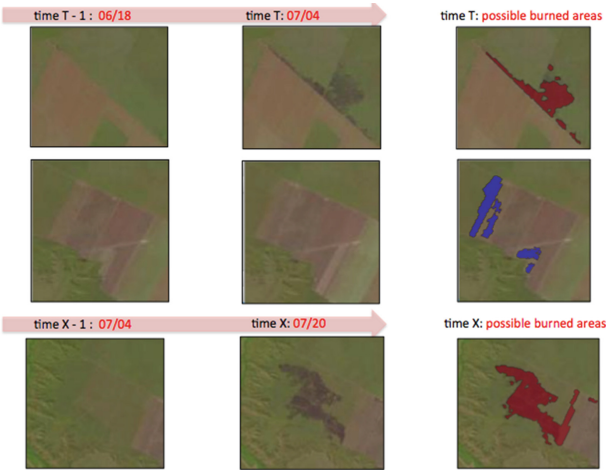


Fig. 2. Detailed examples of burned and non burned polygons. (Color figure online)

The proposed approach is validated over a large study area in the Brazilian Cerrado and Caatinga against reference data derived from classifications done by experts by INPE. The good results of the experiments contribute to propose using the approach to adapt data from nearby areas.

This paper is organized in the following sections: Section 2 shows related work and how remote sensing is used to monitor burned areas. Section 3 presents the proposed approach and explains the dataset. Section 4 presents the experiments for some classification models and discuss the results. Section 5 presents the conclusions and future work.

2 Burned Areas Mapping by Remote Sensing Monitoring

The use of remote sensing images collected in different wavelengths is the most efficient way to monitor fires in places with great territorial extension or areas of difficult access. The spectral data present different information for each type of target and for the different regions of the electromagnetic spectrum that are changed after vegetation burning. Furthermore, remote sensing images frequently provides recent information. In Brazil, for example, it is the most efficient fire monitoring means, with the lowest cost. In recent years, new generations of satellites (e.g. Landsat 8, Sentinel-2 and CBERS-4 – the China-Brazil Earth Resources Satellite) were developed to provide better resolution images and more precise georeferencing. Features and advances made possible by these new satellites generation, and some strategies to combine them to improve cloud-free observations may be found in [11].

There are some projects that monitor fire events and related information in many places around the world. The SERV-FORFIRE (Integrated Services And Approaches For Assessing Effects Of Climate Change) [8] project is one of them that presents collaborative efforts of the international community of remote sensing to deal with forest fires. For specific burned areas estimation, they employ and combine local information, such as soil, vegetation and risk management database to detect burned areas in extreme events with high success.

Inside the automatic mapping of burned areas literature, Chuvieco et al. [5] present a recent and complete burned area review with the main wavelengths, sensors, and satellites used for it. They also explore the physical basis for detecting burned areas from remote sensing data, describes the historical trends and summarizes some recent approaches to map burned areas.

Studies are applying the new advances of medium resolution data. Liu et al. [12] developed an algorithm for continuous monitoring of annual burned areas using Landsat time series. Different of us that use two images comparison and knowledge database to classify burned areas, their algorithm is based on burned pixel detection using harmonic model fitting with longer time series. It allows breakpoint identifications to detach burned pixels with overall accuracies at around 80%.

Pereira et al. [18] presents an approach for mapping burned areas using automated sample selection based on active fires and One-Class Support Vector Machine (SVM) supervised learning model. They present the advances of using automatic detection with good overall accuracies and some limitations such as the dependence of the images' contrast. This study shows the feasibility of using machine learning techniques for the burned areas mapping problem.

Andrade et al. [1] proposed a semiautomatic approach to classify burned areas through the use of neural networks. Previous results decreased the number of polygons wrongly classified and showed the viability of using Neural Networks in the classification process. Mithal et al. [16] employ machine learning approaches in the problem of burned areas mapping. They present a three-step approach to map burned areas in tropical areas which compares data from two low resolution (500 m) successive images from Moderate Resolution Imaging Spectroradiometer

(MODIS), feeds it into a proprietary algorithm to classify these data and perform a final processing step for detection of burned areas. According to the authors, although it has many errors in indicating correct burned areas, it is a promising global approach that brings a more comprehensive assessment of tropical fires.

Bittencourt et al. [2] proposed a strategy to classify burned areas by using a Machine Learning approach. They used a one-year knowledge dataset to classify unknown data in the same region. Extending their work, we employ an approach through the use of Machine Learning (ML) classification models to classify which areas have changed in a comparison of two different acquisition moments. The main difference is that our work is focused on providing to extend strategies to INPE's machine learning classification process to classify areas with few historic datasets. Previous results show the difficulty of choosing one specific model to answer our burn classification problem. To contribute to the understanding of this problem, our work is related to the effort to develop automated approaches to classify burned areas with few previously validated data.

3 Proposed Approach

We hypothesize that knowledge databases with lots of previously validated data can characterize and classify regions in nearby areas. Our aim in this article is to contribute to the development of an approach to detect burned areas with no previously validated data. We treat this problem by proposing a strategy that employs machine learning techniques based on a knowledge database from classified regions to classify near regions. We defined the process to determine relevant features and experimented with the approach on an applicable dataset.

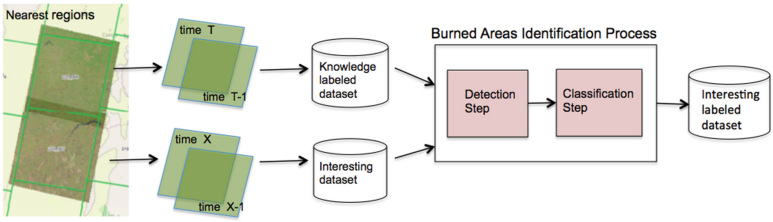


Fig. 3. General view of burned areas identification process.

The general idea of the whole burned areas identification process, illustrated in Fig. 3, is based on remote sensing data evaluation and classification, and it is composed by two steps: the first is to detect spectral changes in two consecutive images of the same region in a limited period. This first step has high reliability and it was developed and is being used at INPE.

The resulting database is composed of polygons extracted by the burned areas mapping algorithm [15]. This process detects candidate burned polygons every two weeks with high reliability for medium resolution images (30 m). It compare

images from the same region to different dates, with the standard temporal difference being a 16-day. The ones that present interference like noise or many clouds are discarded and an image before this one is used, with 32, 48 or at most 64 days difference. At intervals greater than this, the vegetation starts the recovering process and it is not possible to detect relevant changes.

However, some of these changed regions are deforestations, crops, and clouds, and, to be officially considered as burned areas and reported, it needs to be labeled in order to ensure the quality of at least 95% of success in the indication of burned areas. This is done in a second step that defines if the detected change is due to fire, based on a knowledge database of previously detected changes of the same region.

The most recent INPE's published official result of burned areas estimation in the Cerrado employed a further evaluation process to separate the confirmed burned areas dataset. The result of this process is a huge validated knowledge database from the Cerrado biome that is composed of a set of changed areas that can be caused by many factors, with some of them being burn occurrences. The results are good and reliable but the manual evaluation is expensive and time-consuming. Other strategies to create an automatic evaluation must ensure the quality of results.

3.1 Data

The dataset used in this paper's experiment was acquired through INPE's Fire Monitoring Program [7] that uses images from Landsat satellites. The data cover part of the Cerrado biome, as shown in Fig. 4.

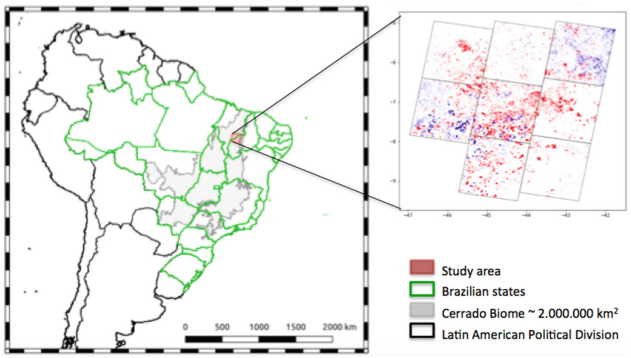


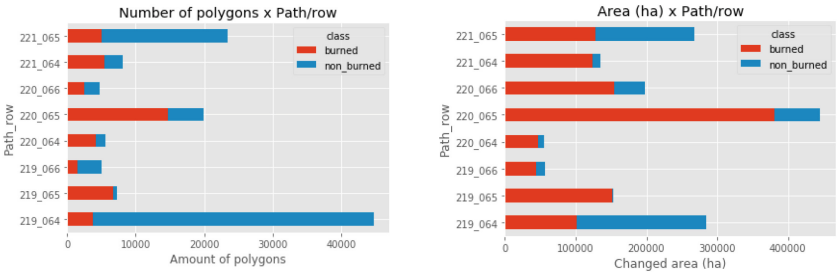
Fig. 4. Fragment of Brazilian territory covering the Cerrado biome and highlighting eight Landsat path/rows. (Color figure online)

The Landsat Program [22] stands out among Earth observation satellites that provide medium-resolution orbital images of the same area every 16 days since the 1980s. It provides environmental data with all corrections processed, such as spectral band values and spectral vegetation indices.

Landsat imagery is separated by paths and rows. The red parallelogram in Fig. 4, the study area of this work, illustrates a fragment of an image position with a set of eight path/rows illustrated in Fig. 4. Light gray highlights indicate the Cerrado delimitation and dark gray indicates protected areas. Every single pixel in the image corresponds to a square cell of 30 m × 30 m on the terrain and each complete image represents a coverage area of 18,500 × 18,500 ha. As each original Landsat image contains the reflectance spectra for each pixel in a digital image, the polygons are composed of a set of pixels.

The parallelograms in Fig. 4 illustrates the eight path/rows and the polygons detected as change between two consecutive images. The central parallelogram corresponds to path 220 and row 065, named by 220/065 path/row, part of the Cerrado biome in the states of Maranhão and Piauí, in Brazilian territory. The complete dataset is composed of positive and negative examples of burned areas occurrences, in a total of 118881 polygons. Confirmed burned areas (positive examples) are represented in red and non burned areas (negative examples) are represented in blue. These latter are polygons detected by the reference algorithm and that were discarded by experts because they were not caused by fires, such as clouds, crops, and deforestations.

The polygons' set was developed by the evaluation of a set of 80 images of the 2018 year. The data ranging is from April to October. Other months correspond to the rainiest period in that region. This makes it difficult to compare consecutive images without a huge portion of clouds and because that period is less prone to the existence of active fires. There are 10 images of each path/row on the following sets: 219/064, 219/065, 219/066, 220/064, 220/065, 220/066, 221/065 and 219/066. In this work, the area division is indicated by path/rows because it is simple to simulate a real process and to facilitate data visualization process.



(a) Polygon's amount (b) Changed area (ha)

Fig. 5. Polygon's amount and changed area by path/row

The study area contains 118881 polygons that cover around 1594211 ha (around 14761155 soccer fields). From those, 75060 correspond to confirmed burned polygons totaling 1133059 ha and 43821 are areas changed due to other

factors, which we call non burned polygons, and which total area is 461151 ha. Figure 5 illustrates the total number of polygons and the area occupied by changed areas by month in each year.

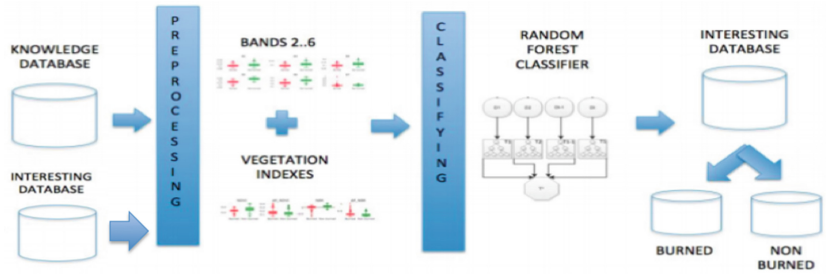


Fig. 6. General view of classification step.

Figure 6 illustrates the general Burned Classification Process. It starts with the comparison of an image I on a time T with an image J on a previous time (T-1). Adjacent pixels on the image I detected as change are grouped in a polygon. For each polygon on the image I, it is computed the medians of each spectral band. After that, it is computed the medians of each spectral band on the same polygon's area on the image J and the differences between values on time T related to time T-1 are estimated.

We obtained the dataset available in early December to build our knowledge base and run experiments from that release. However, as work is in operation and with continuous development, as advances are made within the INPE's research, some advances are introduced into the system, for example, improvements in the data collected by the satellites and also enhancements in clouds and smoke masks. Besides, because the data is used by real users and there is input from other information systems, some omission or detection errors are reported. In this way, the data is reprocessed, re-evaluated and sometimes the official results are updated in the official system and made available to the user.

3.2 Features' Understanding

The knowledge database is composed of features for the polygons, their spectral bands values and spectral vegetation indices. The features can be subdivided into four categories based on this source: original values of bands, original values of spectral indices, values computed of bands' values differences and values computed from spectral indices differences. In this study, b4 denotes data from Landsat band 4, b5 corresponds to data from Landsat band 5 and so on. There are two direct sets of features based on the value of bands. The first set is composed of the median values of each pixel that make up the polygon. The features indicated without suffixes correspond to the median value observed on the day of the satellite's passage.

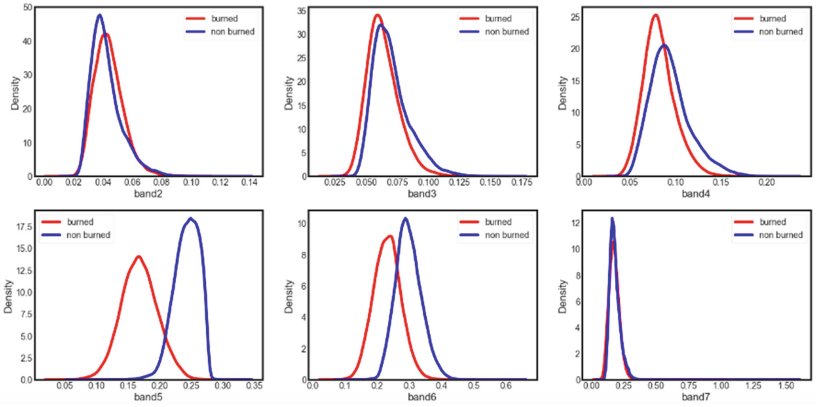


Fig. 7. Histogram of spectral bands b2, b3, b4, b5, b6 and b7 original values, for polygons labeled as burned and non burned.

Figure 7 shows the frequency distributions of original values of spectral bands data from b2 to b7 for all 118881 evaluated polygons. The distributions, except for b5 and b6, are unimodal and similar for burned and non burned polygons. The variability of data is within known specifications limits on the literature [5] for each band. We highlight bands b5 and b6 to differentiate between burned and non burned data because the curves and the mean of burns and non-burns are different.

The differences that appear in the bands, in general, are small with the highest concentrations around 0. The plots show some outliers that focused on the positive part of most of the sets. They are common to the data and that will be kept on to avoid tendency results. Except for the plots of bands 5 and 6, the shape of the burning curve is similar to the format of the non burning graph. This indicates the difficulty of performing linear separation of these sets using these data.

Features indicated with the suffix "_dif" correspond to differences between the median value of the pixels on each polygon in the passage and the medium value of pixels of the region bounded by that polygon in the previous date used in the comparison. Figure 8 shows an overview of the different values of spectral bands data from b2 to b7.

All of these histograms are unimodal and, except for data from bands 2 and 3, the curves are asymmetric. The curves on burned and non burned sets are more similar than original features and outliers are presented in bands b5, b6, and b7. The mean values are concentrated around 0, which demonstrates small differences between data on the previous date.

Another set of attributes of each polygon is composed of eight spectral vegetation indices already used in burned literature. They are: Burn Area Index (BAI) [4], Char Soil Index (CSI) [13], Global Environment Monitoring Index (GEMI) [19], Mid-Infrared Burn Index (MIRBI) [21], Normalized Burn Ratio

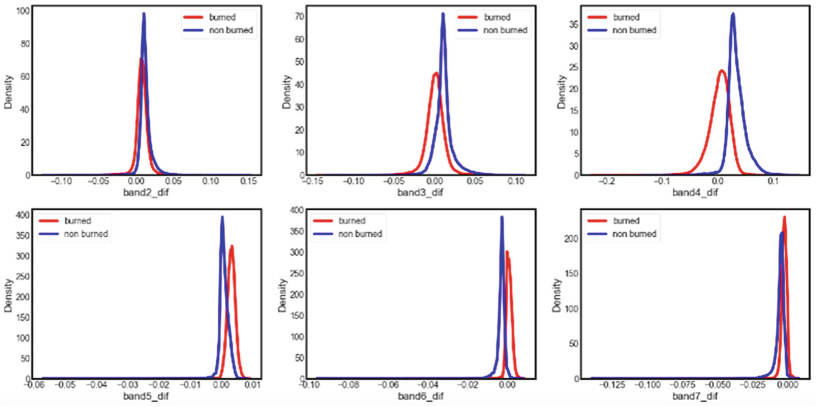


Fig. 8. Histogram of difference between spectral bands b2, b3, b4, b5, b6 and b7 original values and spectral bands previous values, for polygons labeled as burned and non burned.

(NBR) [10], Normalized Difference Vegetation Index (NDVI) [20], and Normalized Difference Water Index (NDWI) [14].

A summary of the vegetation indices applied in this approach is described above.

$$
\begin{aligned}
NDVI &= \frac{b5 - b4}{b5 + b4}, \\
NBR &= \frac{b5 - b7}{b5 + b7}, \\
MIRBI &= 10 * b7 - 9.8 * b6 + 2, \\
NDWI &= \frac{b3 - b6}{b3 + b6}, \\
BAI &= \frac{1}{(0.1 - b4)^2 + (0.06 - b5)^2}, \\
CSI &= \frac{b2}{b6}, \\
GEMI &= \frac{n * (1 - 0.25 * n) * (b5 - 0.125)}{1 - b5} \\
with\, n &= \frac{2 * b7^2 - b5^2 + 1.5 * b7 + 0.5 * b5}{b7 + b5 + 0.5}, \\
GEMIL &= \frac{n * (1 - 0.25 * n) * (b5 - 0.125)}{1 - b5}, \\
with\, n &= \frac{2 * b6^2 - b5^2 + 1.5 * b6 + 0.5 * b5}{b6 + b5 + 0.5}
\end{aligned}
$$

They present good results in preliminary experiments, which are not presented in this article.

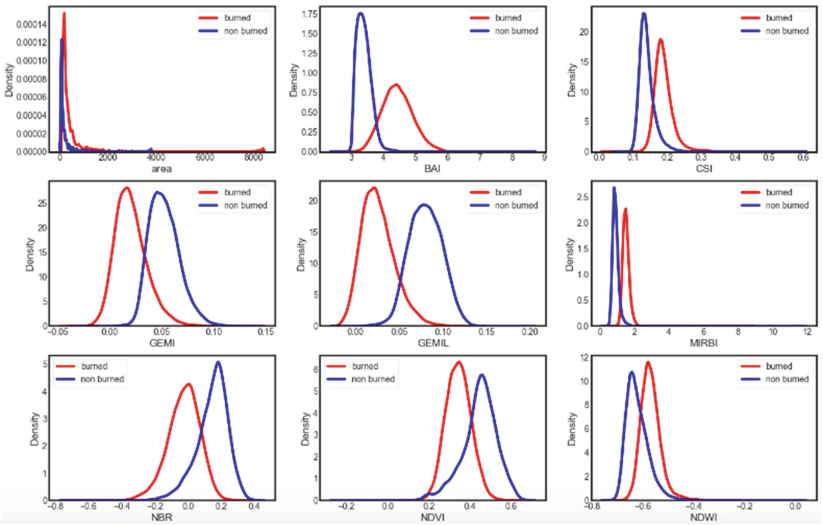


Fig. 9. Histogram of area (ha) and spectral indices original values, for polygons labeled as burned and non burned.

Figure 9 presents the histograms of the spectral vegetation indices and the area of each polygon. Some features present outliers and attributes, GEMIL, NBR and NDVI present differentiation between burned and non burned set but it is still not simple to do a linear cut between the sets. These curves present larger ranges between data. Features AREA and BAI present larger ranges. Curves on these histograms present mean values with more distinction between burn and non burned sets. No feature is alone sufficient to characterize the burn set but it can contribute to correctly distinguish the two sets.

The last set of features is based on the difference between the current indices and the values of indices on the previous images. After this set we have the following set of 29 features: AREA, b2, b3, b4, b5, b6, b7, b2_dif, b3_dif, b4_dif, b5_dif, b6_dif, b7_dif, BAI, CSI, GEMI, GEMIL, MIRBI, NBR, NDVI, NDWI, BAI_dif, CSI_dif, GEMI_dif, GEMIL_dif, MIRBI_dif, NBR_dif, NDVI_dif, and NDWI_dif.

3.3 Data Processing

The strategy in the data processing is to have a smaller set of relevant attributes to reduce the volume of analyzed data and the processing time while maintaining a low error rate improving the prediction performance of the classifiers.

The first step is the normalization of the set of features to have different features on the same scale. That will accelerate the learning process.

After that, the second step is the evaluation of the correlation between each pair of features. We consider correlations are greater than 97,5% represents a

strong value where both features point out about the same knowledge, a multicollinearity situation. In this case, we remove the redundant feature. The rest were maintained in our analysis since the relationships found may not be trivial. Thus, they remain in our analysis, along with the other bands.

At the end of this process, the complete database on this experiment is composed by the following 28 features: AREA, b2, b3, b4, b5, b6, b7, b2_dif, b3_dif, b4_dif, b5_dif, b6_dif, b7_dif, BAI, CSI, GEMI, MIRBI, NBR, NDVI, NDWI, BAI_dif, CSI_dif, GEMI_dif, GEMIL_dif, MIRBI_dif, NBR_dif, NDVI_dif, and NDWI_dif.

Processed data preserves the original distribution characteristics, decreases the effect of the outliers and allows a better interpretation of the set.

4 Classification Experiments

To explore the effectiveness of the approach we created validation tests that simulates a real process of using a recent knowledge dataset to predict the classification to a nearest regions dataset. Our approach is a classification problem solved through a supervised learning class of problems, where a model is trained with a labeled dataset. We explored our approach in five experiments that alternate training and testing datasets. The features of the training dataset are used to create a model. In the testing set, the features excluding the labeled class, are used to predict the class based on the created model. We employed all polygons from some path/rows to test all polygons from other path/rows. No data is repeated in the training and the testing dataset.

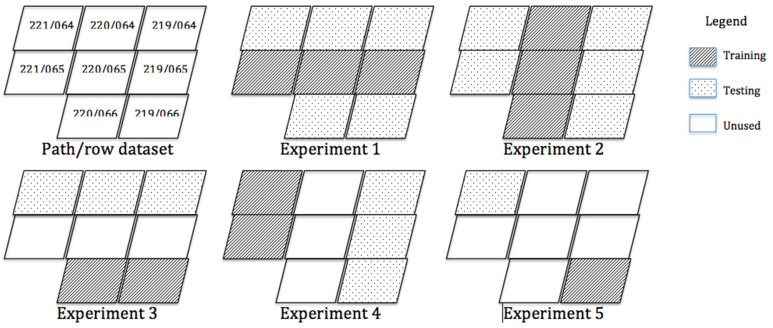


Fig. 10. Complete path/rows positions and each experiment training and testing sets

Figure 10 illustrates the complete path/rows positions and detaches the training and testing sets in each experiment. Experiments 1 and 2 show a global view of the set and they are done with all adjacent path/rows. In experiments 3 and 4, the idea is to test if more distant data than adjacent path/rows would affect the results. Experiment 5 simulates the more distant path/rows available in this dataset.

- Experiment 1: Three central path/rows on the same row compose the training set. Other rows compose the testing set.
  - Training: 219/065, 220/065, 221/065.
  - Testing: 219/064, 219/066, 220/064, 220/066, 221/064.
- Experiment 2: Three central path/rows on the same path compose the training set. Other paths compose the testing set.
  - Training: 220/064, 220/065, 220/066.
  - Testing: 219/064, 219/065, 219/066, 221/064, 221/065.
- Experiment 3: The 066 row is the training and the 064 row is the testing.
  - Training: 219/066, 220/066.
  - Testing: 219/064, 220/064, 221/064.
- Experiment 4: The 221 path is the training and the 219 path is the testing.
  - Training: 221/064, 221/065.
  - Testing: 219/064, 219/065, 219/066.
- Experiment 5: The 219/066 path/row is the training data and the 221/064 path/row is the testing set.
  - Training: 219/066.
  - Testing: 221/064.

Table 1. Number of polygons in each experiment divided by training and testing sets.

Experiment	Training			Testing		
	Burned	Non Burned	Total	Burned	Non Burned	Total
1	21327	8870	30197	22494	66190	88684
2	24473	24079	48452	17348	50981	68329
3	13342	45144	58486	4006	5837	9843
4	12021	45114	57135	10473	21076	31549
5	1566	3501	5067	5409	2690	8099

Table 1 illustrates the number of polygons in each experiment. All of them have distinct amounts of data and some of them are more balanced than others. However, it is possible to see on previous histograms the different sets, on the majority of plots, do not present large variations within the same class.

4.1 Random Forest Model Classifier

This model consists of an ensemble of simple tree decision classifiers used to determine the outcome. Each simple tree employs a set of decision rules to separate the classes and, the ensemble votes for the most popular class. Random Forest is a robust algorithm that minimizes the errors in specific trees.

We chose this classifier because it showed good results in previous works and it has simple interpretability. We performed experiments using the Scikit-learn

environment [17]. The main parameters used on the model are a set of 64 trees in the forest, 5 as the minimum number of instances in leaves and the indication to the classifier to balance the number of examples in each class.

To evaluate the performance of the classifier in the problem of predicting unlabeled data, we analyze the following metrics: the accuracy (the proportion of true results among the total number of cases examined), the precision (the proportion of predicted positives related to real positive), the recall (the proportion of actual positives correctly classified) and the F1 score (the harmonic mean between precision and recall).

Table 2. General overview of each experimental result

Experiment	Accuracy	Precision	Recall	F1 score
1	98	99	94	97
2	98	99	94	97
3	97	93	99	96
4	99	97	98	98
5	97	96	98	97

Table 2 shows a general overview of the results. All analyzed metrics' values are higher than 93%. This indicates that it is possible to predict the class of the polygon with high accuracy, near to reaching our initial aim of 95% of success in each metric.

The training dataset has a different distribution than the test/validation dataset and population. As a real problem, we consider wide margins between training and test, but the results outperform previous experiments. In addition to this challenge, our database is not error-free. Some areas are ambiguous and even experts have doubts in certain places. In such cases, if the experts have not sure in certain areas, they are classified as non burned, further increasing the variability within the class.

These results indicate that for these sets, the approach was able to recover most of the burned areas and to indicate that areas identified as burned had few false positives. It is possible to note that, in these experiments, adjacent path/rows show results close to the results of the most distant orbits. New experiments with more orbits and more distant sets will be necessary to better analyze this result.

5 Conclusions and Results

This work aims to show directions and to add value for the construction of an automated high-performance environment to be able of dealing with the extensive Brazilian territory and the complexity of generating data to build knowledge bases for the classification of fire data and non burned.

We know it is a challenging task to perform this kind of test for large scale areas. Besides, there is the complexity, and, in some cases, the impossibility to generate train datasets. Then, we must advance in creating a fast and accurate automated classifier with little manual interference that can detect fires by combining data from more than one biome. It may treat and classify areas with sparse data. And with each new set of data generated by a satellite image can be autofitted to classify new data from nearby path/rows with the same accuracy.

In this article, we presented the approach, applied it in nearby regions and obtained accuracy close to 95%. Our results show that it is feasible to use the strategy of the near set to help characterize the sets in places with poor data or missing data. We believe that this can be improved with more tests in different areas to propose the minimum values of path/rows and polygons that validate the approach.

To the next steps, it is important to continue investigating whether data closer has a better answer and how far it is possible to apply the strategy with high accuracy. So, to this approach to be incorporated into the standard procedure to test new areas is needed.

We know that different types of vegetation, soil, seasonal effects, and other local characteristics may enrich the evaluation. For future work, we suggest incorporating other data products related to the fire risk, soil models and vegetation to generate a more precise characterization of fires. However, we show that the available data on LandSat is able of taking a step towards evaluating data in regions with little validated data using readily available data. Work continues to be improved. We are testing the most appropriate strategy to treat doubtful cases and working on the model's adaptability to all other Brazilian biomes.

Acknowledgements. This study was supported by National Council for Scientific and Technological Development (CNPq)/Coordination of Associated Laboratories (COCTE/INPE) (no. 300587/2017-1).

References

1. de Andrade, R.N., Bittencourt, O., Morelli, F., Santos, R.: Classificação semi-automática de áreas queimadas com o uso de redes neurais. In: XVIII Brazilian Symposium on Geoinformatics - GeoInfo 2017, pp. 92–97 (2017)
2. Bittencourt, O.O., Morelli, F., dos Santos Júnior, C.A., Santos, R.: Evaluating classification models in a burned areas' detection approach. In: Misra, S., et al. (eds.) ICCSA 2019. LNCS, vol. 11622, pp. 577–591. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24305-0_43
3. Bowman, D., et al.: Fire in the earth system. Science **324**, 481–484 (2009)
4. Chuvieco, E., Martín, M.: Cartografí de grandes incendios forestales en la península ibérica a partir de imágenes noaa-avhrr. Serie Geográfica **7**, 109–128 (1998)
5. Chuvieco, E., et al.: Historical background and current developments for mapping burned area from satellite earth observation. In: Remote Sensing of Environment (2019)
6. Instituto Nacional de Pesquisas Espaciais (INPE): Programa de monitoramento de queimadas. http://www.inpe.br/queimadas/portal. Accessed 28 Jan 2018

7. Instituto Nacional de Pesquisas Espaciais (INPE): Programa de monitoramento de queimadas, área queimada, resolu cão 30m. https://prodwww-queimadas.dgi.inpe.br/aq30m/. Accessed 28 Jan 2018
8. JPI Climate and European Union: Serv-for fire integrated services and approaches for assessing effects of climate change and extreme events for fire and post fire risk prevention. https://servforfire-era4cs.eu/
9. Katagis, T., Gitas, I., Toukiloglou, P., Veraverbeke, S., Goossens, R.: Trend analysis of medium- and coarse-resolution time series image data for burned area mapping in a Mediterranean ecosystem. Int. J. Wildland Fire **23**, 668–677 (2014)
10. Key, C., Benson, N.: Landscape assessment: Ground measure of severity, the composite burn index; and remote sensing of severity, the normalized burn ratio. In: FIREMON: Fire Effects Monitoring and Inventory System, pp. 1–51 (2006)
11. Li, J., Roy, D.: A global analysis of sentinel-2a, sentinel-2b and Landsat-8 data revisit intervals and implications for terrestrial monitoring. Remote Sens. **9**, 902 (2017)
12. Liu, J., Heiskanen, J., Maeda, E.E., Pellikka, P.K.: Burned area detection based on Landsat time series in savannas of southern Burkina Faso. Int. J. Appl. Earth Obser. Geoinf. **64**, 210–220 (2018)
13. Smith, A.M.S., Drake, N.A., Wooster, M.J., Hudak, A.T., Holden, Z.A., Gibbons, C.J.: Production of Landsat ETM+ reference imagery of burned areas within Southern African savannahs: comparison of methods and application to MODIS. Int. J. Remote Sens. **28**, 2753–2775 (2007)
14. McFeeters, S.: The use of normalized difference water index (NDWI) in the delineation of open water features. Int. J. Remote Sens. **17**, 1425–1432 (1996)
15. Melchiori, E., Setzer, A.W., Morelli, F., Libonati, R., Cândido, P.d.A., Jesús, S.C.d.: A Landsat-TM/OLI Algorithm for Burned Areas in the Brazilian Cerrado: Preliminary Results, pp. 1302–1311. Imprensa da Universidade de Coimbra (2014)
16. Mithal, V., Nayak, G., Khandelwal, A., Kumar, V., Nemani, R., Oza, N.C.: Mapping burned areas in tropical forests using a novel machine learning framework. Remote Sens. **10**, 69 (2018)
17. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
18. Pereira, A.A., et al.: Burned area mapping in the Brazilian savanna using a one-class support vector machine trained by active fires. Remote Sens. **9**(11), 1161 (2017)
19. Pinty, B., Verstraete, M.: GEMI: a non-linear index to monitor global vegetation from satellites. Vegetation **101**, 15–20 (1992). https://doi.org/10.1007/BF00031911
20. Rouse Jr., J.W., Haas, R.H., Schell, J.A., Deering, D.W.: Monitoring Vegetation Systems in the Great Plains with ERTS. NASA Special Publication, vol. 351, p. 309 (1974)
21. Trigg, S., Flasse, S.: An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah. Int. J. Remote Sens. **22**, 2641–2647 (2001)
22. United States Geological Survey (USGS): Science Data Lifecycle. https://earthexplorer.usgs.gov. Accessed 18 Oct 2018