RANDOM FOREST AND SUPPORT VECTOR MACHINE APPLIED FOR MAPPING BURNED AREAS IN AMAZON

Nathália Silva de Carvalho[1], Igor José M. Ferreira[1], Thales Sehn Körting[1], Luiz Eduardo Oliveira e Cruz de Aragão[1], Liana Oighenstein Anderson[2]

National Institute for Space Research – INPE, Av. dos Astronautas, 1758 – Jardim da Granja, São José dos Campos – SP,12227-010, Brazil;
[1]{nathalia.carvalho, igor.malfetoni, thales.korting, luiz.aragao}@inpe.br; [2]liana.anderson@cemaden.gov.br

ABSTRACT

The use of fire for land management is one of the main anthropic activities that have led to the impoverishment of tropical forests. Therefore, mapping these areas is paramount for public policies implementation. Currently, machine learning techniques have shown very effective results in the classification of land cover on extensive areas. This paper aims to compare the Random Forest (RF) and Support Vector Machine (SVM) algorithms performance on burned areas mapping in Amazon. Using a multiresolution segmentation algorithm applied to a Landsat image, the training dataset included 300 objects of burned and non-burned areas. Additionally, 24 attributes were tested in both RF and SVM approaches. An overall classification accuracy of 91% was achieved by RF and SVM models using spectral and geometric attributes. Nonetheless, regarding the omissions and inclusion errors, SVM models had the best performance on burned areas mapping.

Key words* — *Pattern recognition, geobia, fire, degradation.

1. INTRODUCTION

Human activities have been one of the major causes of global climate changes such as the droughts regime frequency in tropical forests, which makes the ecosystems vulnerable to wildfires events [1, 2]. Since tropical ecosystems have global relevance by playing an important role in climate regulation and providing ecosystem services [3], forest fire events may result in imbalance due to negative impacts on these natural environmental. The use of fire for land management is one of the main anthropic activities that have led to the impoverishment of tropical forests by altering their functioning and sctructure. As a result the biodiversity and hydrological and energetics processes are affected, contributing for the carbon emission increasing [1]. Thus, it is important to evaluate the occurrence of fire as a way of understanding its dynamics.

In general, a decrease of carbon emission into the atmosphere is expected when the deforestation rate reduces [4]. However, areas with reduction in deforestation rates presented a 59% increase in fire occurrence [5]. These outcomes point to the need of considering fire as one of the major threats to forest systems, being critically important in Reducing Emissions from Deforestation and Forest Degradation (REDD) programs [5, 6].

The technological advances in remote sensing field and Geographic Information Systems (GIS) have significantly contributed to the amount of studies focusing on mapping and managing of fire affected areas [7, 8, 9, 10]. Since the mapping of burned areas is paramount to understand the spatial distribution of fires, as well as to evaluate the social, economic and environmental impacts of these events [11], several methodologies have been proposed to improve the accuracy of fire scar mapping [7, 8, 12].

In the literature, there is a growing trend for machine learning classification methodologies [7, 13]. Two of most popular approaches involving land cover tasks are Random Forest (RF) and Support Vector Machine (SVM) techniques [14, 15] due to their overall accurately results [16, 13].

These approaches apply robust methods based on the use of input data to perform the hierarchical training of the model, used as a reference for object based image classification [14]. These techniques use different layers of information of spectral and spatial features representation from the investigated targets in order to perform a classification, resulting in clustering of similar targets [17, 18]. In the case of RF algorithm, it creates several decision trees by combining different possibilities for the input data set (spectral and spatial features) and evaluates the lower entropy error between them [17]. Each tree contributes to a single vote for the most frequent class. Similarly, SVM algorithm is a binary based method of classifiers that creates an optimal linear hyperplane through maximizing the distance from the data points of each class [18].

In this context, this paper compares the learning performance of RF and SVM algorithms for burned areas mapping in the Amazon Forest by using an object based approach with a spectral and geometric feature sets. The outcomes may be used as subsidies for public policy actions for forest fire monitoring, contributing to the use of automatic techniques in the burned areas detection and mapping.

2. MATERIAL AND METHODS

2.1. Study Area

The studied area is located in the southern region between Novo Progresso and Altamira municipalities, in the State of Pará (Figure 1). With an area of 12.800 km², this region is predominantly composed of the Dense Ombrophilous Forest [19], and the anthropic activities is represented by small subsistence farmers, which are associated with governmental settlements projects [20].

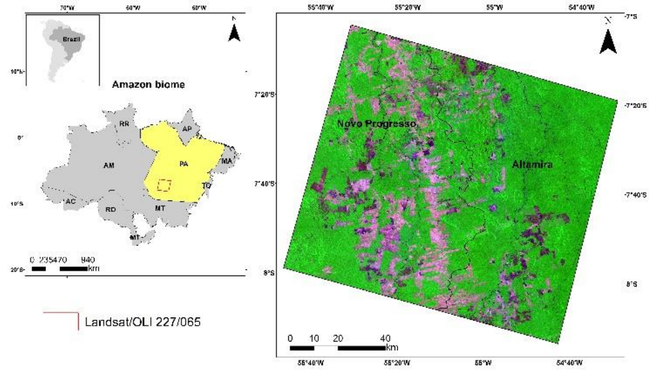

Figure 1. Location of the studied area in the State of Pará, Brazil.

2.2. Data acquisition

In order to identify a representative area for fire occurrence in Amazon forest, an annually search of the fire hotspots spatial distribution was carried out. The fire data was acquired from Fire Information for Resource Management System (FIRMS) as a MODIS derived product from Terra and Aqua satellites (MCD14ML, collection 6) with 1 km of spatial resolution. In this study, only active fire data with confidence level higher than 80% was considered. Then, the fire data was combined to the Landsat World Reference System in order to identify areas with a higher fire occurrence for the scene identification.

After the spatialization process, a Collection-1 and Level-2 Landsat/OLI imagery (227/065 path/row) acquired from U.S. Geological Survey (http://earthexplorer.usgs.gov) was used. In this data set is included the bands sensitive to the blue (0.435 µm - 0.451 µm), green (0.452 µm - 0.512 µm), red (0.636 µm - 0.673 µm), near infrared (0.851 µm - 0.819 µm) short wave infrared (1.566 µm - 1.651 µm) sections of the electromagnetic spectrum with 30m of spatial resolution. The acquired image was radiometrically, atmospherically and geometrically corrected and resampled to UTM projection zone 21 south.

In addition to the reflectance bands, the Normalized Burn Ratio (NBR), Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI) spectral indices were computed as inputs to the training phase.

2.3. Supervised object-based classification

The object based classification was performed by using a multiresolution segmentation algorithm applied to the Landsat image composition in order to delineate the image into homogeneous regions [21]. Here, the homogeneity threshold was defined as 0.4 for shape and 0.4 for compactness. Also, the segment size was established based on a scale parameter of 180 with equivalent weight for all 10 bands. As a result of the segmentation procedure, 3614 regions were obtained. The mean and standard deviation for each object in all bands were calculated as spectral features and shape index, index border, rectangular fit e degree of skeleton branching as the geometric features, totaling 24 features. Afterwards, 300 objects were taken as training samples of the model, divided in burned areas and non-burned areas.

In order to carry out the classification, two approaches were adopted. In the first one, the RF and SVM algorithms was performed using only spectral attributes. Secondly, the geometric attributes were added to the model.

RF classifier relies on predictions from an ensemble of decision trees created using a trained data set. Given the high dimensionality of remotely sensed data, RF classifier uses the lower entropy between the subsets in order to select the best features to describe the classes of interest. Hence, each tree accounts for one vote based on internal validation technique for estimating how well the resulting has been performed. The final classification is obtained by averaging all the probabilities calculated by all trees [17]. SVM algorithm works by creating a linear separating hyperplane which is able to distinguish two different target classes. The hyperplane is created by maximizing distance between two support vectors from the trained dataset. Similar to RF classifier, SVM can be successfully applied for large input dimensionality data due to its validation technique that relies on weighting function (kernel). Furthermore, kernel function allows nonlinear separating boundaries to be learned [18].

Finally, the burned area identification was performed by visual interpretation of the satellite imagery. Afterwards it was used to generate a reference map and validate the tested models. This reference map was compared to the outcomes produced by the RF and SVM models. Also, the confusion matrix and coefficient kappa were calculated for each classification and a z-test was carried out ($p < 0.05$).

3. RESULTS AND DISCUSSION

Table 1 shows the overall accuracy and kappa values produced by the classification using RF and SVM models. Also, the omission and inclusion errors are presented for all classifications.

In general, all the results for burned area classification showed overall accuracy higher than 90%. Also, of all approaches tested in this study, the SVM and RF classifiers presented similar overall accuracy, differing only by the lower kappa value when the geometric attributes were added to the model (Table 1). In this case, the kappa value for RF model was significantly lower than SVM model ($p < 0.05$).

Table 1 – Inclusion and omission errors, overall accuracy and kappa value for RF and SVM approaches.

Class	RF spectral		RF Spectral/geometric	
	Inclusion Error	Omission Error	Inclusion Error	Omission Error
Burned area	32.63	21.18	40.31	24.88
Others	2.93	5.15	3.48	6.85
Overall Accuracy		0.93		0.91
Kappa		0.69		0.61
	SVM spectral		SVM Spectral/geometric	
Burned area	35.57	14.78	40.27	13.05
Others	2.09	6.35	1.88	7.91
Overall Accuracy		0.93		0.91
Kappa		0.69		0.66

It is noted that both RF and SVM tested methods had their overall accuracy decreased from 0.93 to 0.91 when the geometric features were added to the model. Similarly, the addition of geometric features also influenced the kappa value decrease.

The results for the SVM classification using only spectral features demonstrated that the algorithm omitted 14.78% of the burned areas, which is a better performance if compared to the RF method (21.18%). Regarding the inclusion error, even though the RF method classifies non burned area as burned areas less than the SVM algorithm, this difference was only 3%.

According to [14], misclassification errors may be related to the object-based segmentation process by including objects with information from both burned and non-burned classes. It is commonly observed in forest dense that was just burned and lands with sparsely shrubs. In this case, the use of an image with finer spatial resolution may contribute to better results in the segmentation phase once it improves the discrimination of regions (based on the homogeneity threshold), and consequently, it improves the separability of the target classes.

Comparing the performance of the algorithms with and without the geometric features, it is noticed that, unlike the RF, SVM produced a slight lower omission error from 14.78% to 13.05%. Conversely, in both cases RF and SVM the inclusion errors produced were increased. To sum up, the geometric features addition led to a decrease of the machine learning algorithm capacity for discriminating burned area. It is evidenced by the increase of the inclusion error in both RF and SVM classifier.

The non-metric multidimensional scaling (NMDS), evidenced that there was a low separability between the targets regarding the shape index, index border, rectangular fit e degree of skeleton branching (Figure 3). Because the burned area occurrence was varied in the landscape, identifying a specific geometric pattern was difficult. This reason may have contributed to a greater confusion in the training of the classifiers, resulting in an increase of inclusion error when the geometric attributes were added to the model. As suggested by [22], the algorithm performance may negatively be affected when new features that does not provide representative information about the target class are included. From this perspective, the spectral features provided more discriminating information on burned affected areas.

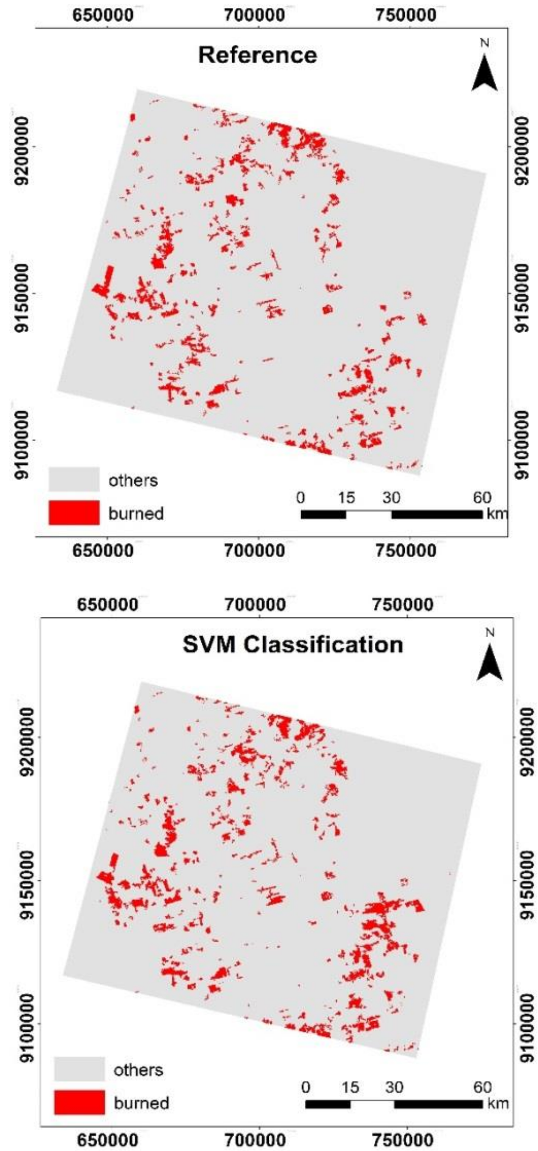

Figure 2. Comparison between the reference map and SVM classification using only spectral attributes.

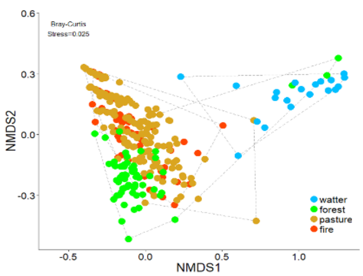

Figure 3.Analysis of the dissimilarity between the geometric attributes.

5. CONCLUSION

In summary, comparing the two classifiers, the results demonstrated that SVM classifier was less exclusive than RF classifier for mapping of burned area once it presented a lower omission error value. Also, adding geometric features to the models decreased the performance of both RF and SVM algorithms. Further studies should consider using temporal attributes in order to investigate improvement of the models.

6. ACKNOWLEDGEMENTS

This study was financed in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processes 309247/2016-0; 140379/2018-5, by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and grant #2017/24086-2, São Paulo Research Foundation (FAPESP).

7. REFERENCES

[1] Anderson, L.O. et al. “Disentangling the contribution of multiple land covers to fire-mediated carbon emissions in Amazonia during the 2010 drought”, *Global Biogeochemical Cycles*, v. 29, n. 10, p. 1739–1753, 2015.

[2] Corlett, R.T. “The Impacts of Droughts in Tropical Forests”, *Trends In Plant Science*, v. 21, n. 7, p.584-593, 2016.

[3] Myers, N. et al. “Biodiversity hotspots for conservation priorities”, *Nature*, v. 403, n. 6772, p. 853–8, 2000.

[4] Gullison, R.E. et al. “Tropical Forests and Climate Policy”, *Science*, v. 316, n. 5827, p. 985–986, 18, 2007.

[5] Aragão, L.E.O.C. e Shimabukuro, Y.E. “The Incidence of Fire in Amazonian Forests with Implications for REDD”, *Science, v. 328, n. 5983, p. 1275–1278,. 2010.*

[6] Barlow, J. et al. “The critical importance of considering fire in REDD+ programs”, *Biological Conservation*, v. 154, p.1–8,. 2012.

[7] Cao, X.; et al. “An automatic method for burn scar mapping using support vector machines”, *Int. J. Remote Sens*, 30, 577–594, 2009.

[8] Zammit, O.; Descombes, X. e Zerubia, J. “Burnt area mapping using support vector machines”, *For. Ecol. Manag*, 234, 2006..

[9] Castelli, M.; Vanneschi, L. e Popovic, A. “Predicting Burned Areas of Forest Fires: An Artificial Intelligence Approach*”, Fire Ecology*, v. 11, n. 1, p.106-118,2015.

[10] Giglio, L; et al. “Global estimation of burned area using MODIS active fire observations”, *Atmospheric Chemistry and Physics*, v. 6, p. 957-974, 2006.

[11] Chuvieco, E. et al. “Integration of ecological and socio-economic factors to assess global vulnerability to wildfire”, *Global Ecology and Biogeography*, v. 23, n. 2, p. 245–258, 2014.

[12] Anderson, L.O.A. et al. “Detecção de cicatrizes de áreas queimadas baseada no modelo linear de mistura espectral e imagens índice de vegetação utilizando dados multitemporais do sensor MODIS/TERRA no estado do Mato Grosso, Amazônia brasileira”, *Acta Amazônica*, v. 35, p. 445-456. 2005.

[13] Ramo, R. e Chuvieco, E. “Developing a Random Forest Algorithm for MODIS Global Burned Area Classification”, *Remote Sensing*, v. 9, n. 11, p.1193-1215, 2017.

[14] Dragozi, E. et al. “Burned Area Mapping Using Support Vector Machines and the FuzCoC Feature Selection Method on VHR IKONOS Imagery”, *Remote Sensing*, v. 6, n. 12, p.12005-12036, 2014.

[15] Belgiu, M. e Draguń, L. “Random forest in remote sensing: A review of applications and future directions”, *ISPR Journal Of Photogrammetry And Remote Sensing*, v. 114, p.24-31, 2016.

[16] Adam, et al.” Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers”, *International Journal Of Remote Sensing*, v. 35, n. 10, p 3440-3458,2014.

[17] Breiman, L. “Random Forests”, *Machine Learning*, v. 45, n. 1, p.5-32, 2001.

[18] Cortes, C.; Vapnik, V. “Support-vector networks”, *Machine Learning*, v. 20, n. 3, p.273-297, 1995.

[19] Vieira, S. et al. “Forest structure and carbon dynamics in Amazonian tropical rain forests”, *Oecologia*, v. 140, n. 3, p.468-479, 2004.

[20] Pinheiro, T. F. et al. “Forest Degradation Associated with Logging Frontier Expansion in the Amazon: The BR-163 Region in Southwestern Pará, Brazil”, *Earth Interactions*,, v. 20, n. 17, p.1-26, 2016.

[21] ECOGNITION DEVELOPER. Using eCognition developer in quickmap mode. In: *ECOGNITION DEVELOPER*. User guide. Munchen: Trimble Germany GmbH, 2001. p. 37-39.

[22] Witten, I. H.; Frank, E. e Hall, M. A. “Data mining: Practical machine learning tools and techniques*”*, p. 629, 2011.