Anuário do Instituto de Geociências
ISSN: 0101-9759
ISSN: 1982-3908
Universidade Federal do Rio de Janeiro

Loiola, Táscilla Magalhães; Fantinel, Roberta Aparecida; Santos, Fernanda Dias dos; Bastos, Franciele de; Schuh, Mateus Sabadi; Fernandes, Pablo; Oliveira, Bruna Andrieli Simões de; Pereira, Rudiney Soares

Use of Machine Learning Algorithms in the Classification of Forest Species

Anuário do Instituto de Geociências, vol. 46, 50490, 2023
Universidade Federal do Rio de Janeiro

DOI: https://doi.org/10.11137/1982-3908_2023_46_50490

Available in: https://www.redalyc.org/articulo.oa?id=695375062003

Anuário do Instituto de Geociências
Universidade Federal do Rio de Janeiro
https://revistas.ufrj.br/index.php/aigeo/

ISSN 0101-9759
e-ISSN 1982-3908

Use of Machine Learning Algorithms in the Classification of Forest Species

Uso de Algoritmos de Aprendizado de Máquina na Classificação de Espécies Florestais

Táscilla Magalhães Loiola[1], Roberta Aparecida Fantinel[1], Fernanda Dias dos Santos[1], Franciele de Bastos[1], Mateus Sabadi Schuh[1], Pablo Fernandes[1], Bruna Andrieli Simões de Oliveira[1], Rudiney Soares Pereira[1]

[1]Universidade Federal de Santa Maria, Centro de Ciências Rurais, Departamento de Ciências Florestais, Santa Maria, RS, Brasil

E-mails: tascillaloiola@gmail.com; fantinel.ar@gmail.com; fernandaengflor@gmail.com; franciele.debastos@gmail.com; mateuschu@gmail.com; pablofernandes30@yahoo.com.br; brunasimoeso@gmail.com; rudineyspereira@gmail.com

Corresponding author: Táscilla Magalhães Loiola; tascillaloiola@gmail.com

Abstract

Optimization in the process of managing forest resources seeks alternatives that make data collection possible. One of them alternatives is spectroradiometry, which consists of measuring the spectral response, having as product the response of the target in relation to the incident radiation along the electromagnetic spectrum, and that, using machine learning, with pre-selected models, makes it possible to identify. Given the above, the study aimed to use machine learning algorithms to classify species by vegetation indices from reflectance data. The study was developed at the Federal University from Santa Maria, working with the species *Ficus benjamina*, *Inga marginata*, *Handroanthus chrysotrichus*, *Psidium cattleianum*, *Salix humboldtiana*, *Corymbia citriodora* and *Myrcianthes pungens*, and spectral readings of the leaves were taken using the FieldSpec®3 spectroradiometer connected to RTS-3ZC3 integrating sphere. The reflectance values with wavelength ranged in amplitude from 350 ηm to 2,500 ηm and spectral resolution of 1 ηm. Vegetation indices were calculated using the software R Studio, being: NDVI, SAVI, RVI, GNDVI, NDWI, NDWI2, GEMI, DVI, TVI, RVI, MSAVI, WDVI. The algorithms used to develop machine learning were: Random Forest (RF), k-Nearest Neighbors (K-NN), Naive Bayes (NB) and Support Vector Machine (SVM). RF proves to be the most appropriate for data validation, with 85% global accuracy, followed by SVM, with 71%, K-NN with 64% and NB with 35%. The indices with the best performance to point the species were NDWI and SAVI.

Keywords: Remote sensing; Spectroradiometry; Vegetation indices

Resumo

A otimização no processo de gerenciamento de recursos florestais busca alternativas que viabilizem a obtenção de dados. Uma dessas alternativas é a espectroradiometria, que consiste em medir a resposta espectral, tendo como produto a resposta do alvo em relação à radiação incidente ao longo do espectro eletromagnético, e que, utilizando do aprendizado de máquina, com modelos pré-selecionados, permite identificar o mesmo. Diante do exposto, o estudo teve como objetivo utilizar algoritmos de aprendizado de máquina para classificar as espécies pelos índices de vegetação a partir de dados de reflectância. O estudo foi desenvolvido na Universidade Federal de Santa Maria, trabalhando-se com as espécies *Ficus benjamina, Inga marginata, Handroanthus chrysotrichus, Psidium cattleianum, Salix humboldtiana, Corymbia citriodora* e *Myrcianthes pungens*, sendo que foram feitas leituras espectrais das folhas por meio do espectrorradiômetro FieldSpec®3 conectado à esfera integradora RTS-3ZC3. Os valores de reflectância com comprimento de onda variaram na amplitude de 350 ηm a 2.500 ηm e resolução espectral de 1 ηm. Os índices de vegetação foram calculados por meio do Software R Studio, sendo: NDVI, SAVI, RVI, GNDVI, NDWI, NDWI2, GEMI, DVI, TVI, RVI, MSAVI, WDVI. Os algoritmos usados para desenvolver o aprendizado de máquina foram: Random Forest (RF), k-Nearest Neighbors (K-NN), Naive Bayes (NB) e Support Vector Machine (SVM). RF revela-se o mais apropriado para a validação dos dados, com 85% de acurácia global, seguido pelo SVM, com 71%, K-NN com 64% e NB com 35%. Os índices com melhor desempenho para apontar as espécies foram NDWI e SAVI.

Palavras-chave: Sensoriamento Remoto; Espectroradiometria; Índices de vegetação

Received: 28 February 2022; Accepted: 14 August 2022

1 Introduction

Species identification is essential for the preservation of forests, and for precise supervision of forest management, ensuring the maintenance of existing species and providing accurate forest inventories (Paula Filho 2013). Traditionally, this kind of activity is made in the field, demanding a considerable quantity of time, financial and human resources. Moreover, these activities are depended on the flowering and fruiting season of the species, among other factors that are observed to facilitate identification via morphological characteristics.

In order to decrease these disadvantages, optimizing the management of forest resources, alternatives that allow the obtention of this data more quickly and with a lower cost has been researched. Remote sensing techniques have been shown to be promising methodologies for the identification of forest species (Kovacs, Wang & Flores Verdugo 2005).

The constant evolution in technological and methodological development in relation to data worked on remote sensing, contribute to the accuracy of vegetation analysis, being possible adopt, for example, the spectroradiometry, which consists of measuring the spectral response in situ, that is, close to the target, in order to reduce the interference of environmental factors that are present in the readings of other sensors (Demarez & Gastellu-Etchegorry 2000).

The final product of the spectroradiometry approach is the design of the target's response to the incident radiation along the electromagnetic spectrum. Thus, it becomes possible to estimate a series of parameters on the general conditions of the variable studied, and also to obtain vegetation indices from the process of calculating these indices, being possible identify the most suitable for the type of data being analyzed.

However, for the recognition of the patterns present in the data coming from spectroradiometry, the use of machine learning is an alternative that consists of investigating computational techniques for learning and obtaining knowledge, assuming that computers learn from models (samples) provided by the researcher. Recent machine learning models are divided into supervised and unsupervised, and what differentiates them is the presence of labels in the data (Mitchell 1997; Rezende 2003). Supervised machine learning is based on a set of real data or training, where an answer is provided, in other words, based on the training model previously labeled there is the construction of a classifier (prediction model) that will be able to predict the new example's label (Mitchell 1997). Each machine learning technique has unique properties. The K-Nearest Neighbors (K-NN) is based on Instances calculating the Euclidean distance of a new example, where is a function for each Instance belonging to the database.

The Random Forest (RF) algorithm (Santacruz 2015) is a learning method that proposes to group data entry variables through several decision trees, built at the time of method training (TrainData) (Oshiro 2013). The algorithm creates multiple decision trees, which are trained from the random selection of a part of the data (two thirds), while the rest is used in the cross-validation of the generated tree (Breiman 2001). The final product of the classifier is given by the class that was returned as an answer by most of the trees belonging to the classifications (Tan, Steinbach & Kumar 2009). Random Forest uses prediction from different decision trees that arise from resampling the original data set and calculates an average from it (Inza et al. 2010).

Support Vector Machines (SVM) (Vapnik 1995), developed with the formulation that encompasses the principle of minimizing structural risk (SRM), involving the minimization of an upper limit for the generalization error. It is a technique that uses the Theory of Statistical Learning and builds a binary classifier based on a set of patterns (training examples). Considering Xi and Yi, where Xi is the input vector and Yi is the desired classification, the objective is to use the training examples so that there is a correct classification in the tests not used in the training. Thus, machine learning models based on the SRM principle tend to have a greater ability to generalize unobserved data, which is one of the main purposes of statistical learning (Vapnik 1995).

Naive Bayes (NB) is a classification technique based on Bayes' theorem that completely disregards the correlation between variables (features). In simple theory, a Naive Bayes classifier assumes that the presence of a particular characteristic in a class is not related to the presence of any other resource. Each training example can decrease or increase the probability that a hypothesis is correct, using a probabilistic model to describe the data set (Santos 2016).

The aim of the study was to use the machine learning algorithms to classify the species by the vegetation indices from reflectance data for 7 forest species commonly found in the study region. used wood for this purpose in Brazil, the *Pinus* spp. and the *Eucalyptus* spp, for this a complete analysis of the raw materials technological characteristics and their beheiver in the kraft pulping processes were carried out.

2 Methodology and Data

The study was conducted on the campus of the Federal University of Santa Maria. According to the Köppen classification, the climate is humid subtropical - Cfa, with an

average annual temperature of 19.2 °C, and well-distributed rainfall throughout the year, with average annual rainfall ranging from 1.400 to 1.900 mm (Alvares et al. 2013) whose location is shown in Figure 1.

The campus is located in southern Brazil, in a transition zone between the Central Depression and the sandstone-basalt cliff of the Southern Brazilian Plateau, with an average altitude of 113 m (INMET 2018). Variations in soil classes are accentuated in the region, with Typic Hapludalf (USDA 2003) being the predominant class in the study area. The vegetation in the region is formed by clean fields and seasonal deciduous forest, escarpments of the Serra Geral and several testimony hills (Longhi et al. 2000).

The material was collected on August 15, 2018, between 7 am and 8 am. The temperature varied from 12 ºC to 13 ºC and the relative humidity of the air was 73 to 85% (INPE 2013). The choice of forest species was random. Adult leaves were collected, visibly free from pests and diseases of seven species, which are: Ficus benjamina, Inga marginata, Handroanthus chrysotrichus, Psidium cattleianum, Salix humboldtiana, Corymbia citriodora and Myrcianthes pungens. The methodological procedures are summarized in the flowchart (Figure 2).

The spectral readings of the leaves were performed in the Remote Sensing laboratory at UFSM using the FieldSpec®3 one spectroradiometer connected to the integrating sphere RTS-3ZC3, to perform the spectral readings. After the optimization and calibration of the sensor system with Spectralon plates, the samples (isolated sheets) were positioned with the adaxial face inside the equipment's integrating sphere. Five pages were read (one reading each), totaling 35 readings. The spectra were stored on the microcomputer and recorded in a text file for further processing.

The resulting data were reflectance values with a wavelength in the range of 350 ηm to 2,500 ηm and spectral resolution of 1 ηm. These data in the ".txt" format were converted into ".csv" so that they could be processed statistically in the R software through the Rstudio interface (R Core Team 2014). The vegetation indexes were calculated using a programming script in the R Studio software. This script contains functions and equations based on Table 1. As a result, a new table "ML Indexes" in the ".csv" format was generated. The table contains fields composed of the 12 vegetation indices and the field of the species analyzed, while the lines contain the data. To evaluate the vegetation indexes, the equations described in Table 1 were used.

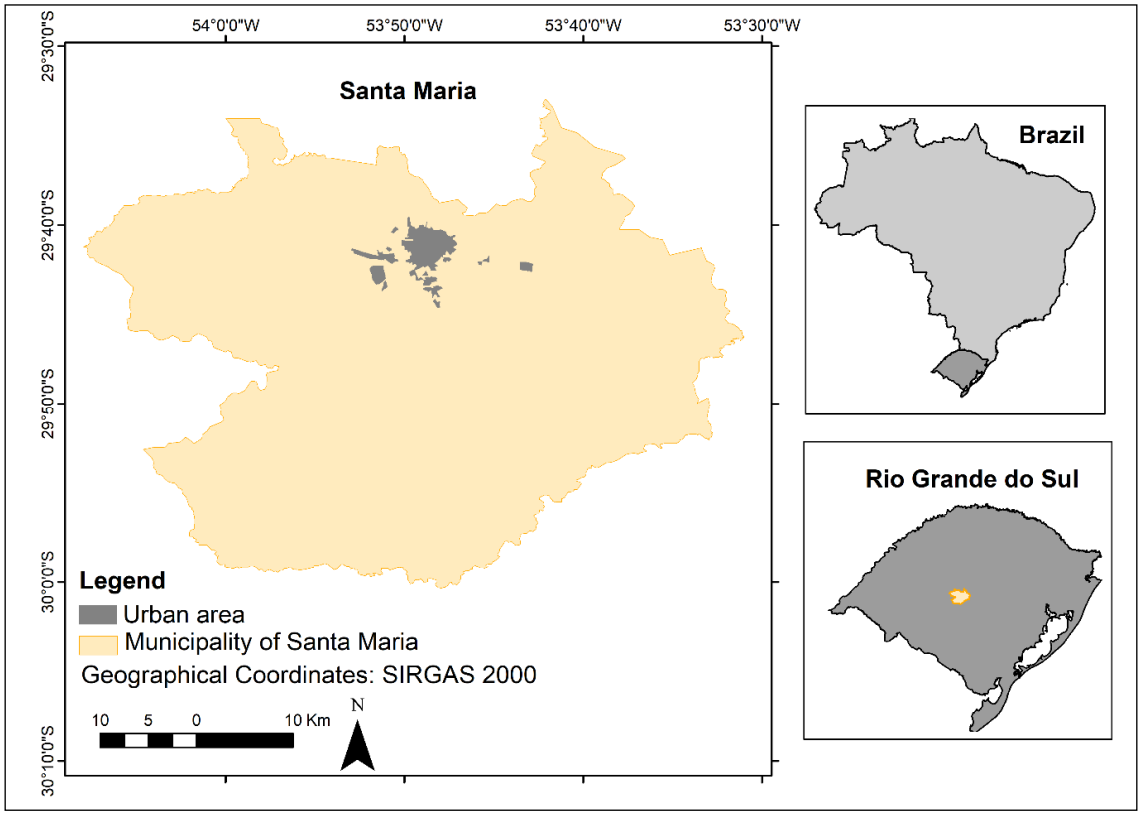

Figure 1 Location of the study area.

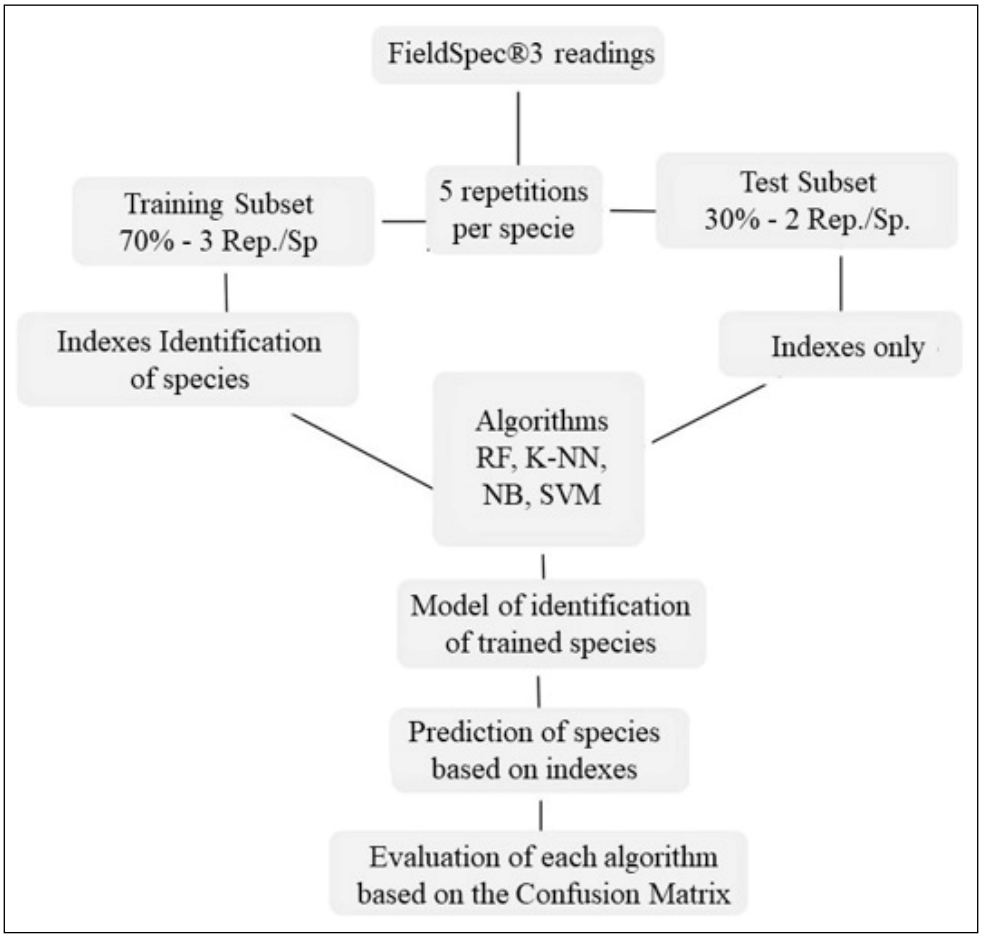

Figure 2 Flowchart of data processing.

Table 1 Vegetation Indexes.

Index	Equations	References
NDVI	(NIR – Red)/(NIR + Red)	Rouse et al. (1974)
SAVI	((1 + L) * (NIR- Red)) / (NIR+ Red+ L)	Huete (1988)
RVI	NIR/Red	Jordan (1969)
GNDVI	(NIR−G)/(NIR+G)	Gitelson, Kaufman and Merzlyak (1996)
NDWI	(Green - NIR)/(Green + NIR)	McFeeters (1996)
NDWI2	(NIR - Swir2)/(NIR + Swir2)	Gao (1996)
GEMI	(((NIR^2 - Red^2) * 2 + (NIR * 1.5) + (Red * 0.5))/(NIR + Red + 0.5)) * (1 - ((((NIR^2 - Red^2) * 2 + (NIR * 1.5) + (Red * 0.5))/(NIR + Red + 0.5)) * 0.25)) - ((Red - 0.125)/(1 - Red))	Pinty and Verstraete (1992)
DVI	(S * (NIR - Red))	Richardson and Wiegand (1977)
TVI	((NIR - S) *(Red))	Deering and Rouse (1975)
RVI	(Red/NIR)	Rouse, et al. (1974)
MSAVI	NIR + 0.5 - (0.5 * Sqrt ((2 * NIR + 1)^2 - 8 * (NIR - (2 * Red))))	Qi et al. (1994)
WDVI	((NIR - S) * (Red))	Richardson and Wiegand (1977)

NDVI = Normalized Difference Vegetation Index, SAVI = Soil Adjusted Vegetation Index, RVI = Ratio Vegetation Index, GNDVI = Green Normalized Difference Vegetation Index, NDWI = Normalised Difference Water Index, NDWI2 = Normalised Difference Water Index, GEMI = Global Environmental Monitoring Index, DVI = Difference Vegetation Index, TVI = Transformed Vegetation Index, RVI = Ratio Vegetation Index, MSAVI = Modified Soil Adjusted Vegetation Index, WDVI = Weighted Difference Vegetation Index.

To develop machine learning and evaluate the efficiency of each vegetation index, the Random Forest, k-Nearest Neighbors, Naive Bayes and Support Vector Machine algorithms were used, implemented in the R packages presented in Table 2. For the analysis of training samples in the classification process, 70% of the data were drawn, while for the training of classifiers, 30% were used for testing.

In the Random Forest method, an importance ranking graph was generated for each vegetation index in the species classification. Regarding the analysis of machine learning methods, the confusion matrix of the test analyzes was taken into account, which aims to generate a matrix of real values and values predicted by its classifier, indicating the amount of data classified correctly. Finally, the machine learning performance table was generated, demonstrating the capacity of the methods to learn automatically from the available data.

3 Results

The indices that showed the best efficiency to distinguish forest species, when submitted to analysis by the Random Forest algorithm, were the Normalized Difference Water Index (NDWI) and the Soil Adjusted Vegetation Index (SAVI) (Figure 3).

With the machine learning validation process using the test samples, the Random Forest algorithm obtained 85% of the global accuracy, being the most appropriate among the tested algorithms (Table 3).

In similar studies Gaiaad et al. (2017), Random Forest presented 95.3% of global accuracy. When analyzing a multi-temporal classification on the dynamics of land use and occupation, the best performance with the Random Forest algorithm was also the best result (Monteiro 2015). However, when comparing five different machine learning algorithms for mapping three different coffee areas, the

Table 2 Machine learning algorithms used to classify forest species.

Algorithm	R Package
K-nearest neighbor (K-NN)	class (Venables & Ripley 2002)
Random Forest (RF)	randonForest (Liaw & Wiener 2002)
Support Vector Machine (SVM)	e1071 (Meyer et al. 2018)
Naive Bayes (NB)	e1071 (Meyer et al. 2018)

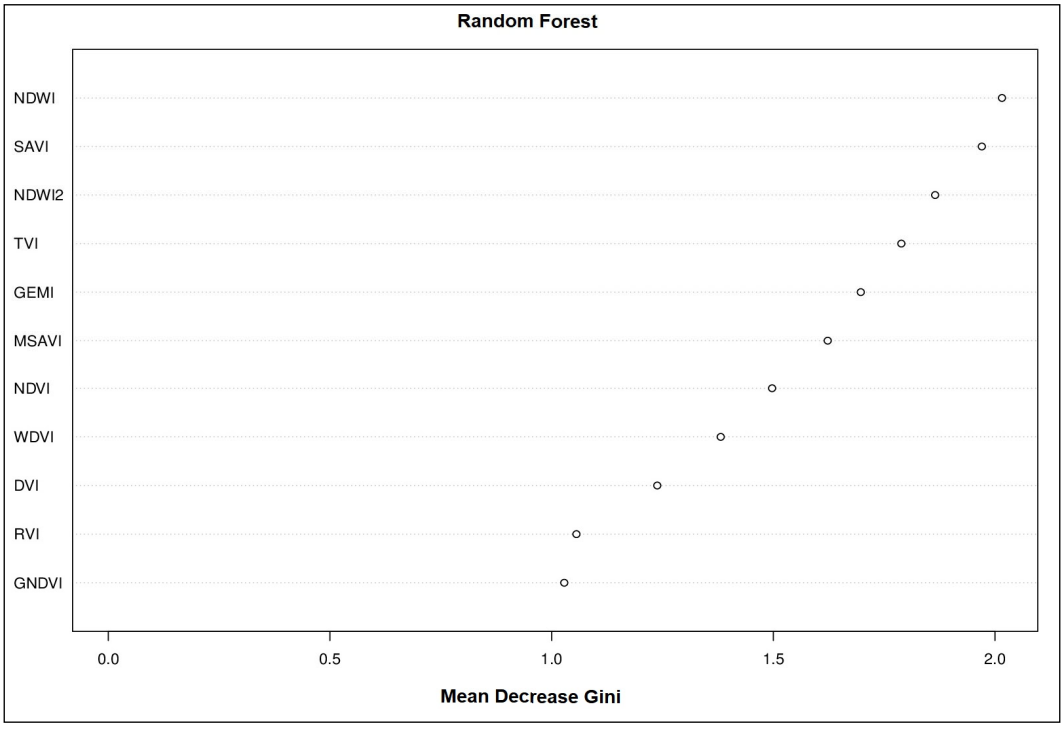

Figure 3 Efficiency of vegetation indexes.

Table 3 Global accuracy of the algorithms.

Algorithm	Accuracy
RF	85%
SVM	71%
K-NN	64%
NB	35%

worst results were found with the Random Forest algorithm, with an overall accuracy of 76.7% (Souza et al. 2016).

The evaluation of the SVM resulted in a total accuracy of 71%, indicating the smallest errors when compared to K-NN with 64% and NB, in which it had only 35% of accuracy in identifying species. In the classification of two forest types from a forest inventory associated with bands 3, 4 and 5 of the Landsat 5 TM satellite through SVM (Gonçalves, De Sá & Ribeiro 2017), an overall accuracy of 86% was obtained.

In similar studies, when evaluating six decision tree algorithms (Gaiaad et al. 2017), the algorithm that obtained the best performance was SVM, with 98.3%. When evaluating the performance of two algorithms based on SVM machine learning and the Multi Layer Perceptron (MLP) for the classification of land use and land cover in the Caatinga biome (Souza et al. 2010), the global accuracy values were 86.03% and 82, 14% respectively, demonstrating the best performance of machine learning methods.

When implementing the Naive Bayers algorithm in the classification of 10 species from the Atlantic Forest leaf database (Souza & Kai 2014), there was a 70.6% accuracy rate. When evaluating different algorithms in three different municipalities (Souza et al. 2016), the algorithm that had the best accuracy was SVM, with an overall accuracy of 85.3%, 87% and 88.3% respectively. The worst results were found with the Random Forest (76.6%) and Naive Bayes (76% and 82%) algorithms, respectively.

4 Conclusions

The indices that had the best performance to distinguish the species evaluated were the Normalized Difference Water Index and Soil Adjusted Vegetation Index. The machine learning method for species classification, the best performance was Random Forest (85%).

It should also be noted that one of the contributions of this work is to highlight the use of the R Studio software, in which the license is made available free of charge, thus allowing users the freedom to study the dynamics of the operation of their respective packages, as well as, adapt them to your needs.

This work is a precursor to research involving other species, not only in the biome addressed, but also in other Brazilian biomes, as well as manifesting the possibility of using the algorithms tested in a greater number of samples in future research.

5 References

Alvares, C.A, Stape, L.J, Sentelhas, C.P., Gonçalves, M.L.J. & Sparovek, G. 2013, 'Köppen's climate classification map for Brazil', *Meteorologisch e Zeitschrift*, vol. 22, no. 6, pp. 711-28, DOI:10.1127/0941-2948/2013/0507.

Breiman, L. 2001, 'Random forests', *Machine Learning*, vol. 45, no. 1, pp. 5–32, DOI: 10.1023/A:1010933404324.

Deering, D. W. & Rouse J. 1975, 'Measuring "Forage Production" of Grazing Units From Landsat MSS Data', *Proceedings of the 10th International Symposium on Remote Sensing of Environment*, Ann Arbor, pp. 1169-78.

Demarez, V. & Gastellu-etchegorry, J.P. 2000, 'A modelling approach for studdying forest chlorophyll content', *Remote Sensing of Environment*, vol. 71, no. 2, pp. 226-38, DOI:10.1016/S0034-4257(99)00089-9.

Gaiad, N.P., Martins P.A., Debastiani, A., Corte, P.A. & Sanquetta, R.C. 2017, 'Uso e cobertura da terra apoiados em algoritmos baseados em aprendizado de máquina: o caso de Mariana - MG', *Enciclopédia Biosfera*, vol. 14, no. 25.

Gao, B. 1996, 'NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space', *Remote Sensing of Environment*, vol. 58, no. 3, pp. 257-66, DOI:10.1016/S0034-4257(96)00067-3.

Gitelson, A.A., Kaufman, Y. & Merzlyak, M.N. 1996, 'Use of a green channel in remote sensing of global vegetation from EOS-MODIS', *Remote Sensing Environment*, vol. 58, no. 3, pp. 289-98, DOI:10.1016/S0034-4257(96)00072-7.

Gonçalves, W.G, De Sá, J.A.S. & Ribeiro, H.M.C. 2017, 'Aplicação de máquinas de vetores de suporte na classificação automática de tipologias florestais', *Revista Seminário Estadual de Água e Floresta*.

Huete, A.R. 1988, 'A soil-adjusted vegetation index (SAVI)', *Remote sensing of environment*, vol. 25, no. 3, pp. 295-309, DOI:10.1016/0034-4257(88)90106-X.

Inmet - Instituto Nacional de Meteorologia 2018, *Análise do Tempo e do Clima*, viewed 21 November 2018, <http://www.inmet.gov.br/sonabra/pg_dspDadosCodigo_sim.php?QTgwMw>.

Inpe - Instituto Nacional de Pesquisas Espaciais 2013, viewed 4 December 2018, <https://www.gov.br/inpe/pt-br>.

Inza, I., Calvo, B., Armañanzas, R., Bengoetxea, E., Larrañaga, P. & Lozano, A.J. 2010, 'Machine learning: an indispensable tool in bioinformatics', *Bioinformatics methods in clinical research*, vol. 593, pp. 25-48.

Jordan, C.F. 1969, 'Derivation of leaf-area index from quality of light on the forest floor', *Ecology*, vol. 50, no. 4, pp. 663-6, DOI:10.2307/1936256.

Kovacs, J.M., Wang, J. & Flores Verdugo, F. 2005, 'Mapping mangrove leaf area index at the species level using IKONOS and LAI 2000 sensors for the Agua Brava Lagoon, Mexican Pacific', *Estuarine, Coastal and Shelf Science*, vol. 62, no. 1-2, pp. 377-84, DOI:10.1016/j.ecss.2004.09.027.

Liaw, A. & Wiener, M. 2002, 'Classification and Regression by RandomForest', *R News*, vol. 2-3, pp. 18-22.

Longhi, S.J., Araújo, M.M., Kelling, B.M., Hoppe, J., Muller, I. & Borsoi, A.G. 2000, 'Aspectos fitossociológicos de fragmento de floresta estacional decidual, Santa Maria, RS', *Ciência Florestal*, vol. 10, no. 2, pp. 59-74, DOI:10.5902/19805098471.

Mcfeeters, S.K. 1966, 'The use of normalized difference water index (NDWI) in the delineation of open water features', *International Journal of Remote Sensing*, vol. 17, no. 7, pp. 1425-32, DOI:10.1080/01431169608948714.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C. & Lin, C.C. 2018, *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*, version 1.7-0, viewed 10 September 2018, <https://cran.r-project.org/package=e1071>.

Mitchell, T. 1997, *Machine Learning*, McGraw Hill, New York.

Monteiro, F.P. 2015, 'ClasSIS: uma metodologia para classificação supervisionada de imagens de satélite em áreas de assentamento localizados na Amazônia', PhD thesis, Universidade Federal Pará, Belém.

Oshiro, T.M. 2013, 'Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica', PhD thesis, Universidade de São Paulo, Ribeirão Preto.

Paula Filho, P.L. 2013, 'Reconhecimento de espécies florestais através de imagens macroscópicas', PhD thesis, Universidade Federal do Paraná, Curitiba.

Pinty, B. & Verstraete, M.M. 1992, 'GEMI: A non-linear index to monitor global vegetation from satellites', *Vegetation*, vol. 101, no. 1, pp. 15-20, DOI:10.1007/BF00031911.

Qi, J., Chehbouni, A., Huete, A., Kerr, Y.H. & Soroosshian, S. 1994, 'A Modified Soil Adjusted Vegetation Index', *Remote Sensing and the Environment*, vol. 48, no. 2, pp. 119-26, DOI:10.1016/0034-4257(94)90134-1.

R Core Team 2014, R: *A language and environment for statistical computing*, R Foundation for Statistical Computing, Austria, viewed 21 November 2018, <http://www.R-project.org/>.

Rezende, S.O. 2003, *Sistemas inteligentes: fundamentos e aplicações*, Manole, Barueri.

Richardson, A.J. & Wiegand, C.L. 1977, 'Distinguishing vegetation from soil background information'. *Photogrammetric Engineering and Remote Sensing*, vol. 43, no. 12, pp. 1541-52.

Rouse, J.W., Haas, R.H., Schell, J.A. & Deering, D.W. 1974, *Monitoring the vernal advancement retrogradation of natural vegetation*, Final Report Type III, NASA/GSFC, Greenbelt.

Santacruz, A. 2015, *Image Classification with RandomForests in R (and QGIS)*, viewed 2 December 2018, <http://amsantac.co/blog/en/2015/11/28/classificationr.html>.

Santos, K.N. 2016, *Utilização de técnicas de aprendizado de máquina para predição de crises epiléticas*, PhD thesis, Universidade Federal do Rio Grande do Norte, Natal.

Souza, B.F.S., Teixeira, S.A., Silva, F.T.A.F., Andrade, M.E. & Braga, S.P.A. 2010, 'Avaliação de classificadores baseados em aprendizado de máquina para a classificação do uso e cobertura da terra no bioma Caatinga', *Revista Brasileira de Cartografia*, vol. 62, pp. 385-99, DOI:10.14393/rbcv62n0-43717.

Souza, C.G., Carvalho, L., Aguiar, P. & Arantes, B.T. 2016, 'Algoritmo de aprendizagem de máquina e variáveis de Sensoriamento Remoto para o mapeamento da cafeicultura', *Boletim de Ciências Geodésicas*, vol. 22, no. 4, pp. 751-73, DOI:10.1590/S1982-21702016000400043.

Souza, J.F. & Kai, P.M. 2014, 'Classificação de folhas usando medidas invariantes', PhD thesis, Universidade Estadual de Mato Grosso do Sul, Campo Grande.

Tan, P., Steinbach, M. & Kumar, V. 2009, *Introdução ao datamining: mineração de dados*, Editora Ciência Moderna, Rio de Janeiro.

Usda - United States Department of Agriculture 2003, *Keys to Soil Taxonomy*, 9th edn, USDA, Washington.

Vapnik, V. 1995, *The nature of statistical learning theory*, Springer-Verlag.

Venables, W.N. & Ripley, B.D. 1992, *Modern Applied Statistics with S*, 4th edn, Springer, New York.

Author contributions

Táscilla Magalhães Loiola: conceptualization; formal analysis; methodology; validation; writing - original draft; writing – review and editing; funding acquisition; visualization. **Roberta Aparecida Fantinel:** conceptualization; formal analysis; methodology; validation; writing - original draft; writing – review and editing; funding acquisition; visualization. **Fernanda Dias dos Santos:** conceptualization; methodology. **Franciele de Bastos:** conceptualization; methodology; writing - original draft; writing – review and editing. **Mateus Sabadi Schuh:** conceptualization; methodology. **Pablo Fernandes:** conceptualization; formal analysis; methodology; writing - original draft; writing – review and editing. **Bruna Andriele Simões:** conceptualization; methodology. **Rudiney Soares Pereira:** conceptualization; formal analysis; methodology; supervision.

Conflict of interest

The authors declare no potential conflict of interest.

Data availability statement

All data included in this study are publicly available in the literature.

Funding information

Not applicable.

How to cite:

Loiola, T.M., Fantinel, R. A., Santos, F.D., Bastos, F., Schuh, M.S, Fernandes, P., Oliveira, B.A.S., & Pereira, R.S. 2023, 'Use of Machine Learning Algorithms in the Classification of Forest Species', *Anuário do Instituto de Geociências,* 46:50490. https://doi.org/10.11137/1982-3908_2023_46_50490