UFSM

Ciência Florestal

ISSN 1980-5098
OPEN ACCESS

Ci. Fl., Santa Maria, v. 34, n. 2, e68716, p. 1-21, Apr./June 2024 • https://doi.org/10.5902/1980509868716
Submitted: 3rd/12/2021 • Approved: 19th/10/2023 • Published: 7th/06/2024

Articles

Classification of Semideciduous Seasonal Forest successional stages using Sentinel-1-2 and SRTM data on Google Earth Engine

Classificação de estágios sucessionais da Floresta Estacional Semidecídua utilizando dados Sentinel-1-2 e SRTM no Google Earth Engine

Vinícius Lorini da Costa[I], **Marcos Wellausen Dias de Freitas**[I]

[I]Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

ABSTRACT

Remote sensing data used in this study included MSI (Multispectral Instrument) Sentinel-2, SAR (Synthetic Aperture Radar) Sentinel-1, GLCM (Grey Level Co-Occurrence Matrix) texture data derived from Sentinel-1, and geomorphometric data derived from SRTM (Shuttle Radar Topography Mission) images. The input data was divided into separate groups for machine learning algorithms, including Support Vector Machine (SVM), Classification and Regression Tree (CART), and Random Forest (RF), which were implemented on the Google Earth Engine platform. RF showed the highest overall accuracies (93 to 97%), regardless of the dataset used as input, with the Kappa index ranging from 0.89 (optical and SAR data) to 0.95 (optical, SAR, and geomorphometric data). CART showed identical overall accuracy values (92.5%) except for the dataset supplemented with SAR texture data, which showed slightly lower accuracy (91.7%), with the Kappa index ranging from 0.89 to 0.91. The worst performance was classifying optical data by SVM, resulting in 59% accuracy and a Kappa index of 0.37. However, the synergy of optical, SAR, and geomorphometric data classified by SVM achieved 75% accuracy.

Keywords: Supervised classification; Cloud computing; Secondary Forest

RESUMO

Foram utilizados dados de sensoriamento remoto adquiridos pelos sensores MSI (*Multispetral Instrument*) do satélite Sentinel-2 e SAR (*Synthetic Aperture Radar*) Sentinel-1, dados de textura GLCM (*Grey Level Co-Ocurrence Matrix*) derivados das imagens Sentinel-1 e dados geomorfométricos derivados de imagens SRTM (*Shuttle Radar Topography Mission*). Os dados compuseram diferentes grupos de entrada para os classificadores de aprendizagem de máquina *Support Vector Machine* (SVM), *Classification and Regression Tree* (CART) e *Random Forest* (RF), implementados na plataforma *Google Earth Engine*. O RF apresentou as maiores exatidões globais (93 a 97%), independente do conjunto de dados utilizados como entrada, com o índice Kappa variando de 0,89 (dados ópticos e SAR) a 0,95 (dados ópticos, SAR e geomorfométricos). O CART apresentou valores idênticos de exatidão global (92,5%) exceto para o conjunto de dados acrescido dos dados de textura SAR, que apresentou exatidão ligeiramente mais baixa (91,7%), com índice Kappa variando de 0,89 a 0,91. O pior desempenho foi o da classificação de dados ópticos por SVM, resultando em 59% de exatidão e 0,37 de índice Kappa. Todavia, a sinergia de dados ópticos, SAR e geomorfométricos classificados por SVM atingiu 75% de exatidão.

Palavras-chave: Classificação supervisionada; Computação em nuvem; Floresta Secundária

1 INTRODUCTION

Secondary forests are altered environments, common in the landscape, constantly increasing in extent as primary forests are degraded and as productive areas are abandoned, becoming areas in regeneration. Depending on disturbance history and soil-climatic characteristics, these forests vary in age, floristic composition and growth rate (AGUILAR, 2005). Secondary forest vegetation is commonly grouped into three succession stages: initial, intermediate, and advanced, each with particular characteristics regarding species composition and vegetation structure (LU, 2003). Mapping these stages is a fundamental part of these studies, in addition to environmental monitoring and conservation, allowing for quantitative and qualitative assessment of forest remnants as well as their spatial distribution (SOTHE, ALMEIDA, LIESENBERG, SCHIMALSKI, 2017). Since 1993, primary and secondary forests in intermediate and advanced stages of regeneration in the Brazilian Atlantic Forest biome and its associated ecosystems have been subject to legal regulations regarding their use, conservation, and parameters for characterization and classification of these stages due to the biome's degradation state (BRAZIL, 1993; 1994; 2006).

Forest succession is described by changes in a community over time. It represents the gradient from rapid growth of pioneer species, which promotes the ideal environment for developing slower-growing species when they detect an opening in the canopy. Among the parameters for differentiation of forest successional stages are age (or regeneration time), average height, average basal area, physiognomy, and remote sensing data (LU, 2003). For forests in Rio Grande do Sul, CONAMA Resolution 33/94 (BRAZIL, 1994) is the legal instrument that defines the successional stages of secondary vegetation of the Atlantic Forest and associated ecosystems, which include, among other typologies, the Seasonal Semideciduous Forest (SSF).

To evaluate the potential and limitations of using radiometric data for characterizing secondary forest vegetation in Altamira/PA, PONZONI (2004) correlated biophysical data collected manually in the field with Landsat-5/TM sensor radiometric data and found no statistically significant differences between the stages. However, they concluded that the relative position of the spectral curves of each stage is related to increased shade levels, with lower reflectance values in the advanced stage. According to FOODY & CURRAN (1994), forest succession is characterized by increases in leaf biomass, woody biomass, and canopy roughness, with maximum roughness achieved with the formation of a mature tertiary canopy. Generally, it is difficult to differentiate between the intermediate and advanced stages using optical remote sensing data. They also state that C-band SAR images, when acquired in different polarizations, have a high potential for differentiating leaf biomass and canopy roughness. Mapping forest successional stages through optical remote sensing images presents challenges in classification because the reflectance spectra of the involved classes are very similar (VIEIRA, 2003).

Aspects related to terrain with excellent temporal stability, such as slope, aspect, and elevation, have been reported as critical biogeographic parameters influencing vegetation growth and development. For forest classification purposes, they can increase accuracy by 10% (LIU, 2019; CHIANG & VALDEZ, 2019).

The Sentinel-2 mission features the Multispectral Instrument (MSI), a passive sensor that collects solar energy the Earth reflects. The MSI has a spectral resolution of 13 bands, spatial resolution ranging from 10 to 60 meters depending on the band, and a radiometric resolution of 12 bits, resulting in 4096 brightness levels. Product 2A includes atmospheric correction over product 1C and provides land surface reflectance. Additionally, band QA60 contains information about cloud presence (ESA, 2019). Sentinel-2 data, despite having a shorter time series when compared to Landsat, provides information with superior spatial resolution and data from the red edge (bands 5, 6, and 7). These are particularly promising for their ability to detect fine differences in chlorophyll pigments, as higher chlorophyll content can indicate higher canopy density, structural complexity of the community, or nitrogen content in plant tissue (LAURIN, 2016).

The Sentinel-1 mission is equipped with the Synthetic-Aperture Radar (SAR). This side-looking active sensor operates in the C band with a frequency of 5.405 GHz and supports single (VV and HH) or dual (VV+VH and HH+HV) polarization operation. This active sensor transmits and receives microwave radiation in two polarizations, acquiring information about surface roughness and geometric arrangement of terrain components. Microwave penetration into the forest depends mainly on the wavelength of the radiation. Frequency, incidence angle, and polarization are system parameters, and dielectric constant, temperature, surface orientation, and roughness are target parameters. The degree of depolarization and the amount of backscattered energy in the sensor are typically controlled by local topography, the presence of vegetation, structures, and the occurrence of water bodies (PERIASAMY, 2018).

In forest environments, double-bounce backscattering outcomes from the interaction between radiation pulses, trunks, and ground, returning to the sensor, whereas in volumetric backscattering the radiation interacts with the canopy volume, being scattered multiple times as it propagates, is mainly observed, with the latter being the primary source of cross-polarization radiation (HV or VH) (SAATCHI, 2019). In

C band, radiation penetrates about one meter, interacts with the canopy volume, and provides information about leaf biomass, saturating at around 60 to 70 tons per hectare (FOODY & CURRAN, 1994). This band is less suitable than other microwave bands, such as L and P bands, as the backscattering of this relatively small wavelength saturates in locations with biomass density of 30-50 t/ha (HANSEN, 2020). The L band saturates at approximately 100 t/ha (JOSHI, 2016), and the P band saturates in the range of 150-200 t/ha (HANSEN, 2020) due to the longer wavelength, greater penetration into the canopy, and interaction with larger diameter trunks and branches.

Research on remote sensing of forest vegetation has been developed to map species distribution (CHIANG, 2019; XIE, 2021), above-ground biomass (PERIASAMY, 2018; LAURIN, 2018; CHEN, 2019; LIU, 2019; FORKUOR, 2020), vegetation type (WOOD, 2012; LAURIN, 2013; 2016; ERINJERY, 2018; PRATICÒ, 2021), succession of secondary forests (LU, 2003, 2014; SOTHE, ALMEIDA, LIESENBERG, SCHIMALSKI, 2017), among others. Many of these forest image classification studies have used machine learning techniques, which require data for algorithm training and predetermination of classes, known in the literature as supervised classification (TASSI & VIZZARI, 2020). In supervised classification, the quality of classification depends on labeled training data, as these data must be statistical representatives of the surface for the classifier to perform satisfactorily (WHITE, 2018).

Acquiring training data is a costly component in a mapping program, and there is a desire to reduce the volume of these data, particularly in actions with limited resources, such as those related to biodiversity conservation, as the financial resources required for *in situ* data collection increase as the volume of data increases (FOODY, 2007). Additionally, appropriate variable selection and refined algorithms are the two main points that can improve vegetation classification performance (SOTHE, ALMEIDA, LIESENBERG, SCHIMALSKI, 2017).

In this study, we aimed to map the successional stages of the Seasonal Semideciduous Forest (SSF) in the Serra do Sudeste region of Rio Grande do Sul using

machine learning techniques on the Google Earth Engine (GEE) platform. Specifically, we aimed to: a) evaluate the potential of different Sentinel-1, Sentinel-2, and SRTM datasets for classifying successional stages of SSF in SSRS, and b) analyze the performance of Support Vector Machine, Classification and Regression Tree, and Random Forest algorithms for classifying successional stages of SSF in SSRS using different SR data sets. The results of this research are important for impact assessment, monitoring, and conservation of forest resources, allowing for quantitative and qualitative assessment of forest remnants, temporally and spatially.

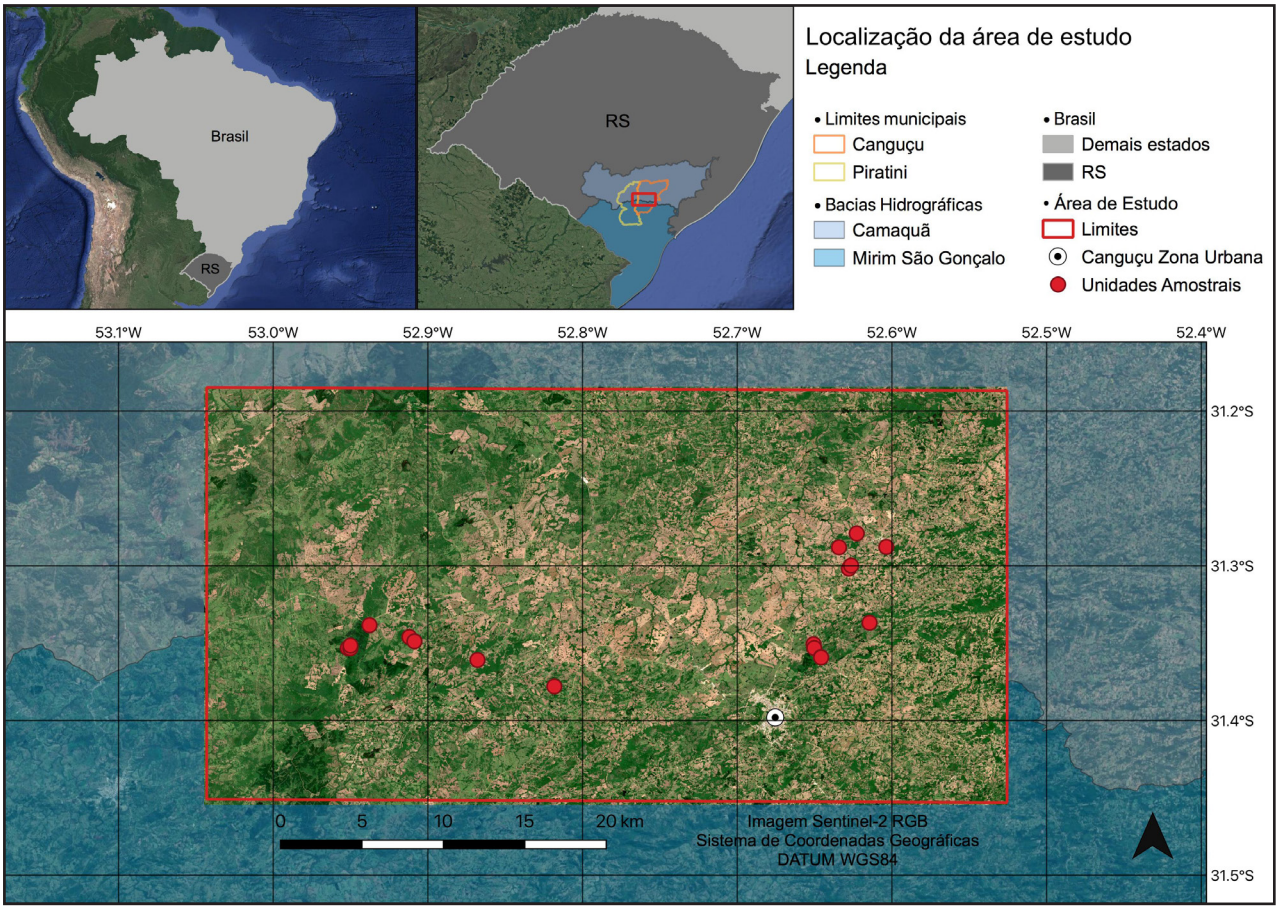
2 MATERIALS AND METHODS

2.1 Study area

The study area comprises portions of the municipalities of Canguçu and Piratini, covering approximately 1450 km² and is divided between the Camaquã and Piratini-São Gonçalo-Mangueira hydrographic basins (Figure 1). Monthly historical data on temperature and precipitation for the period from 2009 to 2019 were analyzed, collected by the Canguçu meteorological station (latitude -31.40333, longitude -52.700833, altitude: 446.81m), processed, and made available by the Brazilian National Institute of Meteorology. The lowest average temperature was 11.25°C in July, and the highest was 21.76°C in January. ALVARES (2013) mapped the Köppen-Geiger climatic classification for Brazil and indicated the predominant occurrence of a humid subtropical climate with temperate summers (Cfb) in Rio Grande do Sul on the Plateau of Araucarias. Meanwhile, the region of the study area presents a humid subtropical climate with hot summers (Cfa) and patches of Cfb climate. WREGE (2016) simulated the future occurrence of Araucaria angustifolia using two Representative Concentration Pathways (RCP) of greenhouse gas concentrations until the year 2100 proposed by the Intergovernmental Panel on Climate Change (IPCC): RCP 4.5 (less pessimistic) and RCP 8.5 (more pessimistic). In both simulations, the fundamental niche areas for the

species in the Serra do Sudeste region of Rio Grande do Sul disappear entirely in 2041 due to climate change. FRITZSONS (2018) considers this species to be a bioindicator of Cfb climate.

Figure 1 – Study area location



Source: Authors (2023)

The field-forest mosaic altered by rural anthropic use, is the landscape physiognomy as highlighted by CORDEIRO & HASENACK (2009). As CARLUCCI (2011) discussed, *Araucaria angustifolia* is present in nuclei, which are small but phytogeographically important due to latitude. Despite the Araucaria forest showing a natural tendency to expand over the fields, this movement is obstructed by fire and grazing actions (BEHLING, 2009). The areas of the Serra do Sudeste with Araucaria presence are prioritized for the creation of conservation units aiming at species conservation (CNCFlora, 2021).

With the precise objective of generating vector layers of training and validation samples for native forest classes, a meticulous vegetation survey was conducted in 18 sample units (SU) of 1000m² (50x20m) randomly installed in forest fragments. Here, all individuals with diameter at breast height (DBH) ≥ 5 cm were identified for species, had DBH measured with a measuring tape, and total height estimated with a four-meter graduated pole. The vertices of the SU were marked with GPS for the generation of vector data representing the successional stages of the secondary forest, which were classified according to federal parameters. The intermediate regeneration stage presented mean DBH and height of 11 cm and 7.1 m, respectively; while the advanced regeneration stage presented mean DBH and height of 15.8 cm and 10.05 m, respectively. The survey resulted in 195 point geometries, with 108 referring to the intermediate stage and 87 to the advanced stage. Formations in the initial regeneration stage were not included.

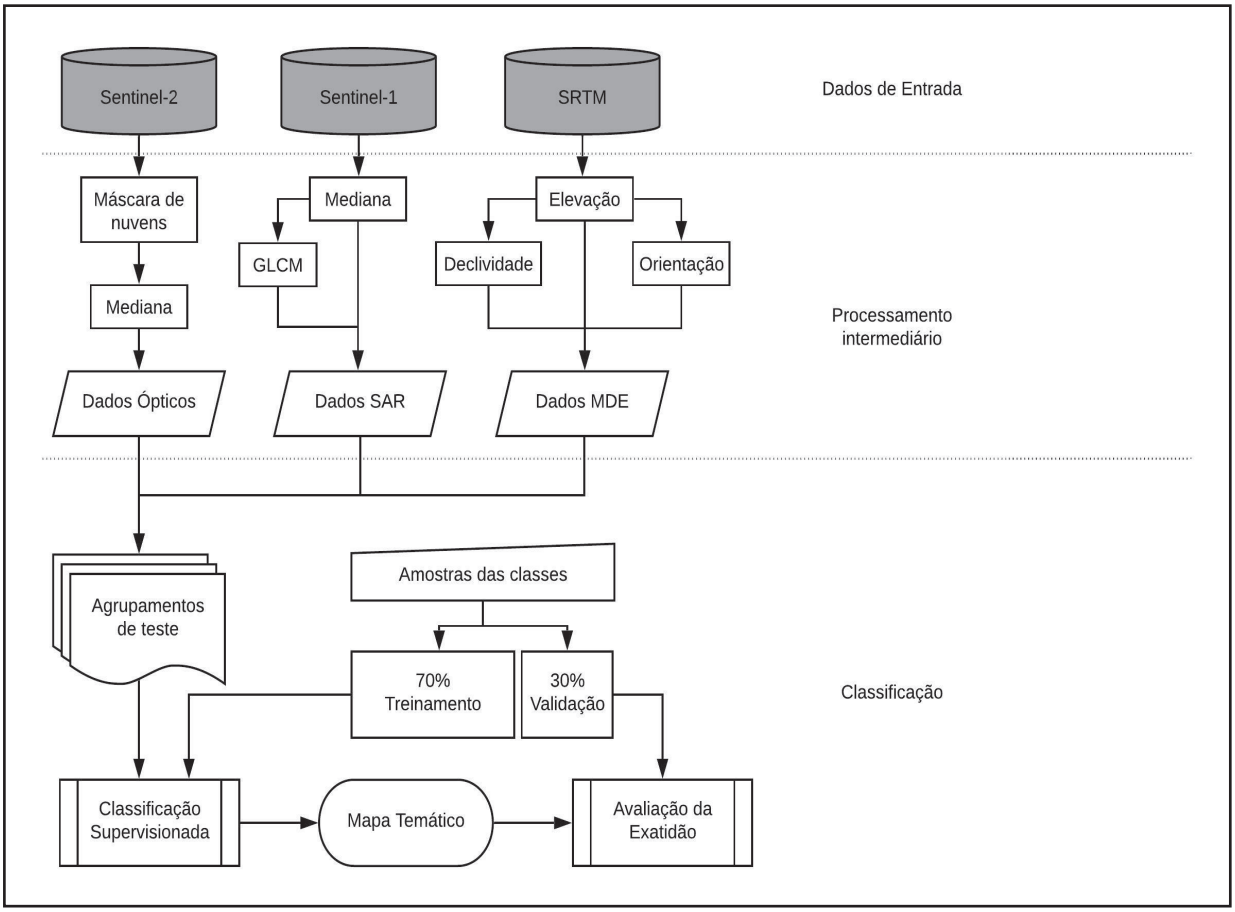
From the perspective of horizontal ecological importance, the tree species that compose the secondary forest fragments, in descending order of Importance Value Index (MUELLER-DOMBOIS & ELLENBERG, 1974), were *Ocotea pulchella* (canela-lageana), *Blepharocalyx salicifolius* (murta), *Sebastiania commersoniana* (branquilho), *Casearia decandra* (guaçatunga), *Myrcianthes gigantea* (araçá-do-mato), *Cupania vernalis* (camboatá-vermelho), *Zanthoxylum rhoifolium* (mamica-de-cadela), *Vitex megapotamica* (tarumã), *Nectandra megapotamica* (canela-preta), and *Allophylus edulis* (chal-chal).

2.2 Remote Sensing Data

The SRTM product version 3.0 (SRTM Plus) with a resolution of 1 arcsecond (approximately 30m), the Sentinel-2 image collection, level 2A, which provides orthoimages with atmospherically corrected surface reflectance values, and the Sentinel-1 image collection at GRD level (Ground Range Detected) were used. The GRD images provide pre-processed images for thermal noise removal, radiometric calibration, and terrain correction. The GRD images are then processed for backscattering coefficient (σ°) in decibels (dB) on a logarithmic scale, all available on GEE.

All processing was conducted on the GEE platform and is presented in the flowchart in Figure 2. From the SRTM DEM, geomorphometric variables of slope and aspect were derived using specific functions embedded in GEE (XIE, 2021).

Figure 2 – Data processing flowchart



Source: Authors (2023)

For Sentinel-2 data, images from the study area obtained between September 9th and December 22nd, 2019, with cloud cover of less than 10%, were selected using an automatic cloud masking procedure using the QA60 band. These selection criteria returned seven images aggregated using median temporal aggregation (CARRASCO, 2019). For analysis, bands B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12 were selected, with spatial resolutions of 10 and 20m, depending on the band.

For Sentinel-1 data, images from the study area acquired between September 9th and December 22nd, 2019 (spring), in IW mode (Interferometric Wide Swath, GRD –

Ground Range Detected, High Resolution), with relative orbit 24 and spatial resolution of 10m (High resolution) were filtered. The available images are dual-polarization (VV+VH) in decibels. A median reducer was applied to the SAR image stack based on STROMANN (2020), which used a central tendency measure reducer on SAR image stacks with the same objective and claimed that speckle noise is reduced while the characteristics of different land use and land cover classes are captured.

From the median of the SAR bands, three metrics based on the Gray Level Co-occurrence Matrix (GLCM), proposed by HARALICK (1973), were calculated using a 3x3 pixel moving window. The calculated metrics were Contrast, which measures the local Contrast of the image; Inverse Difference Moment (IDM), which measures the gray level homogeneity; and Entropy, which measures the randomness of the gray level distribution (TASSI & VIZZARI, 2020). According to HALL-BAYER (2017), GLCM textures have been successful in increasing classification accuracies and are extensively used in forest studies where textures are formed by different shading patterns due to the structural differences of trees. GLCM textures Contrast and Entropy are associated with the degree of variation in gray shade, which correlates with forest biomass since backscattering increases in rugged canopies and decreases in younger canopies, where the canopy is typically arranged in only one layer (KUPLICH, 2005).

For the supervised classification of images, the Random Forest (RF), Classification and Regression Tree (CART), and Support Vector Machine (SVM) classifiers, all implemented in GEE, were used. RF is the only classifier that requires defining the parameter of the number of decision trees to be created. ERINJERY (2018) used 10 and 100 trees, TASSI & VIZZARI (2020) used 50 trees, and XIE (2021) optimized the parameter by implementing a learning curve ranging from 1 to 100 trees and obtained an optimal value of 71 trees. In this study, 50 decision trees were used in the RF classifier (TASSI & VIZZARI, 2020). CART and SVM classifiers were used in the default configuration and did not require parameters settings. For classifier training, a dataset comprising 70% of the samples was created and randomly selected by GEE using the *randomColumn()* function, which divides the dataset using the desired proportion.

For training and validation, 391 sample points of the target classes were classified *in situ*, each corresponding to different pixels (without overlap or repetition) within the study area. The samples were distributed in land use and land cover classes (water, agriculture, native grassland, forestry, intermediate regeneration secondary forest, and advanced regeneration secondary forest). The samples were divided into a 70/30 ratio for training and validation, totaling 270 training elements and 121 validation elements.

Different test groups were formed (Table 1) to verify the performance of remote sensing datasets and their synergies for increasing classification accuracy. Classification accuracy assessment informs the ability of the classifier used to identify the desired targets in a given image (XIE, 2021). Accuracy assessment was applied to the validation dataset, which evaluated accuracy based on the confusion matrix, from which overall accuracy and Kappa index were derived (SULOVA & ARSANJANI, 2021).

Table 1 – Input data groups used

Datasets	Description
S2	Sentinel-2 (B2, B3, B4, B5, B6, B7, B8, B8A, B11 e B12) Median
S2-MDE	**S2** + Elevation, Slope and Terrain Aspect
S2-S1	**S2** + SAR Sentinel-1 (VV, VH) median
S2-S1-MDE	All above
S2-S1-D-GLCM	**S2-S1-MDE** + GLCM textute metrics *Contrast, IDM* and *Entropy* of VV and VH polarizations

Source: Authors (2023)

3 RESULTS AND DISCUSSIONS

Fifteen land use and land cover classifications and forest succession stages were conducted in the study area using three machine learning classifiers (RF, CART, and SVM). Five datasets, including optical, geomorphometric, SAR, and SAR-derived texture data, were used as input to evaluate the synergistic potential for classifying stages of native forest vegetation succession.

The Random Forest algorithm returned the highest overall accuracies (93 to 97%) regardless of the input dataset, with the Kappa index ranging from 0.89 (optical and SAR data) to 0.95 (optical, SAR, and geomorphometric data). CART achieved values very close to those of RF, presenting identical overall accuracy (92.5%), except for the dataset augmented with SAR texture data, which showed slightly lower accuracy (91.7%), with the Kappa index ranging from 0.89 to 0.91. The poorest performance was from optical data classification by SVM, resulting in 59% accuracy and a Kappa index of 0.37. However, the synergy of optical, SAR and geomorphometric data classified by SVM achieved 75% accuracy, representing a 16.6% increase in accuracy and a 0.25 increase in the Kappa index (0.62), with this classifier showing the most significant variations related to the inclusion of other types of data (Table 2).

Table 2 - Global Accuracy and Kappa Index Values Achieved by Each Classification Algorithm and Dataset Used

	SVM		RF		CART	
Data input	**Exat.**	**Kappa**	**Exat.**	**Kappa**	**Exat.**	**Kappa**
S2	0,586	0,37	0,950	0,92	0,925	0,91
S2-MDE	0,719	0,55	0,942	0,91	0,925	0,89
S2-S1	0,694	0,56	0,933	0,89	0,925	0,91
S2-S1-MDE	0,752	0,62	**0,966**	**0,95**	0,925	0,91
S2-S1-MDE-GLCM	0,719	0,52	0,950	0,92	0,917	0,89

Source: Authors (2023)

When analyzing the accuracy data of producer and user achieved by each algorithm for the forest classes per dataset (Table 3), Random Forest returned more homogeneous results, with less variation among the obtained values. However, when examining the producer and user accuracies achieved by the CART classifier, it is observed that these results vary significantly between 0.77 and 0.97. This considerable variation indicates that commission and omission errors occur more frequently in different experiments.

Table 3 – User and Producer Accuracies Achieved by Each Algorithm for the Main Classes of Interest for Each Input Dataset

		S2		S2-MDE		S2-S1		S2-S1-MDE		S2-S1-MDE-GLCM	
		FM	**FA**	**FM**	**FA**	**FM**	**FA**	**FM**	**FA**	**FM**	**FA**
SVM	User	0,00	1,00	0,31	0,84	0,19	0,84	0,50	0,78	0,29	0,82
	Producer	0,00	0,39	0,80	0,53	1,00	0,51	0,87	0,61	0,73	0,53
CART	User	0,77	0,97	0,81	0,91	0,77	0,97	0,81	0,94	0,81	0,91
	Producer	0,95	0,84	0,88	0,85	0,95	0,84	0,91	0,86	0,88	0,85
RF	User	0,88	0,91	0,85	0,91	0,85	0,88	**0,92**	**0,94**	0,88	0,91
	Producer	0,88	0,91	0,88	0,88	0,85	0,88	**0,92**	**0,94**	0,88	0,91

Source: Authors (2023)

In where: FM: Intermediate Stage Forest; FA: Advanced Stage Forest.

Forest differentiation through remote sensing is considered by LAURIN (2013) to be challenging due to the smooth transition between different successional stages and the spectral overlap of these classes. Spectral overlap refers to the situation where the spectral signatures of different land cover types are similar, making it difficult to distinguish between them using remote sensing data. Authors report that the use of different types of remote sensing data (optical, SAR, indices, terrain models) tends to improve the performance of algorithms, whether for classification or regression (CHEN, 2019; FORKUOR, 2020).

Analyzing the estimates by algorithm in Table 4, CART returned consistently the same amount of area per landcover class, regardless of the dataset. The overestimation of the silviculture class by this algorithm is demonstrated in Figure 3, which shows two cutouts of the study area and the classification results for the S2-S1-MDE dataset. In both presented cutouts, CART showed the highest overestimations of the silviculture class. The quantities returned by SVM indicate overestimations for the native forest classes compared to the other algorithms. Furthermore, observing the distribution pattern in the cutouts of Figure 3, confusion between silviculture and native forest fragments is noted, as well as difficulty in separating forest classes, possibly related to spectral overlap between such classes.

Table 4 – Heatmap (red for lower values, green for higher values) indicating the estimated absolute area by classification algorithm and dataset for each class

		AC	CA	SC	AG	FM	FA
SVM	S2	388,4	166,3	8,2	2,8	0	912,5
	S2-D	419,2	265,1	54,9	3,2	176,1	559,3
	S2-S1	303,4	307,9	237,1	2,6	51,5	575,7
	S2-S1-MDE	493,9	268,7	164,2	2,4	144	404,7
	S2-S1-MDE-GLCM	419,3	265,1	54,2	3,2	176,1	559,2
CART	S2	241,6	363,5	419	0,8	345,5	107,8
	S2-D	241,6	363,5	418,9	0,8	348,4	104,9
	S2-S1	241,6	363,5	419	0,8	345,5	107,8
	S2-S1-MDE	241,6	363,5	418,9	0,8	339,8	113,4
	S2-S1-MDE-GLCM	241,6	363,5	418,9	0,8	348,4	104,2
RF	S2	412,8	444,8	237,2	2,4	223,5	157,6
	S2-D	407,5	425,1	246,5	2,6	201,1	195
	S2-S1	402,2	418,4	251,8	2,2	274,2	129,5
	S2-S1-MDE	360,4	445,6	251,8	2,5	288,4	129,2
	S2-S1-MDE-GLCM	417,6	467,1	218,3	2,7	170,3	201,9

Source: Authors (2023)

In where: AC: Agriculture; CA: Grassland; SC: Silviculture; AG: Water; FM: Intermediate Stage Forest; FA: Advanced Stage Forest. Areas in km².

The dataset that yielded the best accuracy and Kappa index results was the combination of optical, SAR, and geomorphometric data, indicating the potential of this synergy for classifying land use and land cover with high spectral overlap. GLCM texture data did not incorporate the expected effects of increased accuracies. When mapping above-ground biomass distribution, FORKUOR (2020) achieved 90% accuracy using Random Forest regression, correlating vegetation biophysical variables with optical time series and fused radar-derived indices, totaling 138 predictor variables. They reported that optical data outperformed radar data when used separately; however, the complementarity of both data types increased the accuracy of their experiments compared to using only one dataset. CHEN (2019) also considers optical data fundamental, highlighting the greater importance of data derived from the SRTM DEM compared to Sentinel-1 SAR data. This partly corroborates the results obtained in this study, such as the high accuracies achieved by RF using only optical data and the improved SVM performance with the addition of DEM and SAR data.

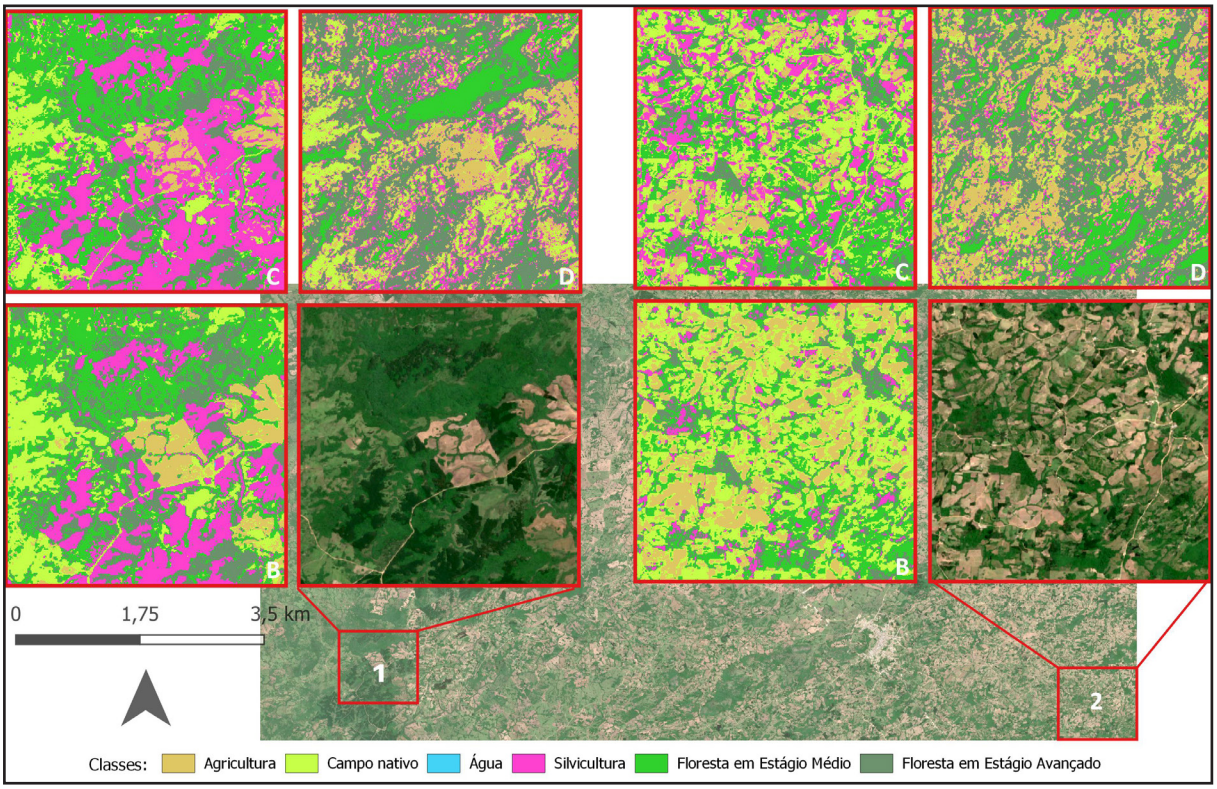
Regarding the SVM algorithm, it is more sensitive to the quality of training data, and the low classification performance achieved by this method in this study may be related to sample heterogeneity or the presence of outliers, as indicated by PRATICÒ (2021). SOTHE (2017) achieved a global accuracy of over 98% with SVM for classifying successional stages in Santa Catarina using input data composed of all Sentinel-2 bands, indicating that classifier parameter settings such as kernel type, as well as kernel parameters settings themselves, promote classification performance improvement. Another factor for the lower SVM performance in this research is related to the number of training samples. Increasing the number of training samples is a fundamental optimization factor for this algorithm in land use and land cover classification, as per STROMANN (2020), who used the default GEE settings and generated a learning curve, revealing that classifications with less than 100 samples resulted in accuracies between 60 and 70%, with a gradual accuracy increase to approximately 90% with 2000 samples.

In comparative terms with other studies using the same RF, CART, and SVM classifiers in GEE, this work showed similar performance to SHAHARUM (2020), who mapped oil palm cultivars in regions of the Malaysian peninsula, achieving global accuracies of 86.5%, 80.08%, and 93.16%, respectively. Despite the higher global accuracy of SVM, the authors highlighted that it classified oil palm samples as “other types of vegetation” for one of the analysed regions and considered decision tree-based algorithms (RF and CART) more efficient, with RF performing better overall, as in this study.

The same can be said regarding the performance of RF, SVM, and CART classification algorithms in forestry studies. This study performed similarly to PRATICÒ (2021), who used them to classify forest types in a Mediterranean region, obtaining global accuracies of 88%, 83%, and 80%, respectively, using spectral indices and seasonal optical compositions.

In this study, the same result obtained by HARALICK (1973) and HALL-BAYER (2017) was not observed, as the inclusion of texture data caused more confusion in the classifications of the three algorithms, resulting in lower global accuracies. This may be related to the window size used for calculating these metrics, as CHEN (2019) reported achieving better correlation values using an 11x11 pixel window.

Figure 3 – Representation of two study area cutouts and the classification results of the RF (B), CART (C), and SVM (D) algorithms for the S2-S1-DEM dataset



Source: Authors (2023)

4 CONCLUSIONS

The Sentinel-1 and SRTM geomorphometric data did not increase the global accuracies of the CART and RF classifiers; however, their synergistic use with optical data demonstrated slightly higher potential for RF classification. Even with only 270 training elements for all classes, with the intermediate and advanced successional

stage classes having 75 and 60 samples, respectively, the Random Forest classifier achieved a global accuracy performance of over 90%. This algorithm has high classification power even when the target classes exhibit spectral overlap and when the number of training samples is limited. It also has straightforward usage as it does not require the configuration of advanced parameters, making it relevant for classifying stages of natural vegetation succession. Based on the results obtained in this study, it is concluded that the synergy between optical, radar, and geomorphometric data showed the most significant potential for classifying the intermediate and advanced succession stages of the Semideciduous Seasonal Forest in the Southeastern Serra of Rio Grande do Sul using the Random Forest algorithm. Based on this model, it is estimated that there are 288.4 km² and 129.2 km² of secondary forests in intermediate and advanced regeneration stages, representing 19.88% and 8.91% of the area of interest, respectively.

ACKNOWLEDGEMENTS

This work was financially supported by CNPq (National Council for Scientific and Technological Development). Process number 132457/2019-9.

REFERENCES

AGUILAR, Alexis. Remote Sensing of Forest Regeneration in Highland Tropical Forests. **Giscience & Remote Sensing**, volume 42, p. 66-79, 2005.

ALVARES, C.A.; José Luiz Stape, Paulo Cesar Sentelhas, José Leonardo de Moraes Gonçalves e Gerd Sparovek. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, volume 22, p. 711 – 728, 2013.

BRASIL. Decreto nº 5.092 de 21 de maio de 2004. Define regras para identificação de áreas prioritárias para a conservação, utilização sustentável e repartição dos benefícios da biodiversidade, no âmbito das atribuições do ministério do meio ambiente. **Diário Oficial da União**, Página 2, 24/05/2004.

BRASIL. Decreto nº 750, de 10 de fevereiro de 1993. Dispõe sobre o corte, a exploração e a supressão de vegetação primária ou nos estágios avançado e médio de regeneração da Mata Atlântica, e dá outras providências. **Diário Oficial da União** - s 1, p 1801, 11/2/1993.

BRASIL. Lei Nº 11.428, de 22 de dezembro de 2006. Dispõe sobre a utilização e proteção da vegetação nativa do Bioma Mata Atlântica. **Diário Oficial da União** - s1, p1, 26/12/2006.

BRASIL. Resolução CONAMA nº 33, de 7 de dezembro de 1994. Define estágios sucessionais das formações vegetais que ocorrem na região da Mata Atlântica do Estado do Rio Grande do Sul, visando viabilizar critérios, normas e procedimentos para o manejo, utilização racional e conservação da vegetação natural. **Diário Oficial da União** 248, s1, pg. 21352-21353, de 30/12/1994.

CARLUCCI, M. B. *et al.* Conservação da Floresta com Araucária no Extremo Sul do Brasil. **Natureza & Conservação**, volume 9, p. 111 – 114, 2011.

CARRASCO, Luis; Aneurin W. O'Neil, R. Daniel Morton e Clare S. Rowland. Evaluating Combinations of Temporally Aggregated Sentinel-1, Sentinel-2 and Landsat 8 for Land Cover Mapping with Google Earth Engine. **Remote Sensing**, volume 11, 288, 2019.

CHEN, Lin *et al.* Optimal Combination of Predictors and Algorithms for Forest Above-Ground Biomass Mapping from Sentinel and SRTM Data. **Remote Sensing**, volume 11, 414, 2019.

CHIANG, Shou-Hao & VALDEZ, Miguel. Tree Species Classification by Integrating Satellite Imagery and Topographic Variables Using Maximum Entropy Method in a Mongolian Forest. **Forests**, volume 10, 2019.

CNCFlora. *Araucaria angustifolia* in Lista Vermelha da flora brasileira versão 2012.2 Centro Nacional de Conservação da Flora. Available in: http://cncflora.jbrj.gov.br/portal/ptbr/profile/araucaria%20angustifolia. Accessed in: 26 June 2021.

CORDEIRO, J. L. P.; HASENACK, H. Cobertura vegetal atual do Rio Grande do Sul. Em: Pillar, V. De P.; Müller, S. C.; Castilhos, Z. M. De S.; Jacques, A. V. Á. (Org.). **Campos sulinos**: **conservação e uso sustentável da biodiversidade.** Brasília: MMA, p. 285-299, 2009.

ERINJERY, Joseph J & Singh, Mewa & Kent, Rafi. Mapping and assessment of vegetation types in the tropical rainforests of the Western Ghats using multispectral Sentinel-2 and SAR Sentinel-1 satellite imagery. **Remote Sensing of Environment**, volume 216, p. 345-354, 2018.

ESA (EUROPEAN SPACE AGENCY), 2019. Sentinel Online. Available in: https://sentinels.copernicus.eu/web/sentinel/missions

FOODY, G., & CURRAN, P. Estimation of Tropical Forest Extent and Regenerative Stage Using Remotely Sensed Data. **Journal of Biogeography**, volume 21(3), p. 223-244, 1994.

FORKUOR, Gerald *et al.* Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets - A case study. **Remote Sensing of Environment, v**olume 236, 111496, 2020.

FRITZONS, E.; Wrege, M.S.; Mantovani, L.E. A Distribuição Natural Do Pinheiro-Do-Paraná No Estado Do Rio Grande Do Sul, Brasil: A Influência De Fatores Climáticos. **Revista Brasileira de Climatologia,** volume 22, p. 117 – 132, 2018.

HARALICK, R.M., Shanmugam, K. and Dinstein, I. Textural features for image classification**. IEEE Transactions on Systems, Man, and Cybernetics**, volume SMC-3, n.6, p.610-621, 1973.

HALL-BAYER, Mryka. Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. **International Journal of Remote Sensing**, 38:5, p. 1312 – 1338, 2017.

IBGE – Instituto Brasileiro De Geografia E Estatística (2006). Mapa de Regiões Fitoecológicas. https://www.ibge.gov.br/geociencias/downloads-geociencias.html

KUPLICH, T.M.; P. J. Curran, P.M. Atkinson. Relating SAR image texture to the biomass of regenerating tropical forests. **International Journal of Remote Sensing**, volume 26, p. 4829 – 4854, 2005.

LAURIN, Gaia Vaglio *et al.* Above-ground biomass prediction by Sentinel-1 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data. **Journal of Applied Remote Sensing**, volume 12, p. 016008-1 - 016008-18, 2018.

LAURIN, Gaia Vaglio*,* Nicola Puletti, William Hawthorne, Veraldo Liesenberg, Piermaria Corona, Dario Papale, Qi Chen, Riccardo Valentini*.* Discrimination of tropical forest types, dominant species, and mapping of functional guids by hyperspectral and simulated multispectral Sentinel-2 data. **Remote Sensing of Environment**, Volume 176, p. 163-176, 2016.

LAURIN, Gaia Vaglio*,* Veraldo Liesenberg, Qi Chen, Leila Guerriero, Fabio Del Frate, Antonio Bartolini, David Coomes, Beccy Wilebore, Jeremy Lindsell, Riccardo Valentini*.* Optical and SAR sensor synergies for forest and land cover mapping in a tropical site in West Africa. **International Journal of Applied Earth Observation and Geoinformation,** volume 21, p. 7 – 16, 2013.

LIU, Yanan *et al.* Estimation of the forest stand mean height and aboveground biomass in Northeast China using SAR Sentinel-1B, multispectral Sentinel-2A, and DEM imagery**. ISPRS Journal of Photogrammetry and Remote Sensing,** volume 151, p. 277 – 289, 2019.

LU, Dengsheng *et al.* A comparative analysis of approaches for successional vegetation classification in the Brazilian Amazon. **GIScience & Remote Sensing**, volume 51, p. 695 – 709, 2014.

LU, Dengsheng. *et al.* Classification of successional forest stages in the Brazilian Amazon basin. **Forest Ecology and Management**, volume 181, p. 301–302, 2003.

MUELLER-DOMBOIS D, ELLENBERG H. **Aims and methods of vegetation ecology**. New York: John Wiley & Sons; 1974.

PAL, M. & MATHER, P.M. Support vector machines for classification in remote sensing. **International Journal of Remote Sensing**, volume 26, p. 1007 – 1011, 2005.

PERIASAMY, S. Significance of dual polarimetric synthetic aperture radar in biomass retrieval: An attempt on Sentinel-1. **Remote sensing of environment**, volume 217, p. 537-549, 2018.

PONZONI, F.J. e Ana Carolina Pinto Rezende. Caracterização espectral de estágios sucessionais de vegetação secundária arbórea em altamira (PA), através de dados orbitais. Árvore**,** volume 28, p. 535 – 545, 2004.

PRATICÒ, Salvatore *et al.* Machine Learning Classification of Mediterranean Forest Habitats in Google Earth Engine Based on Seasonal Sentinel-2 Time-Series and Input Image Composition Optimisation. **Remote Sensing,** volume 13, 586, 2021.

SAATCHI, Sassan em: Flores-Anderson, A. I., Herndon, K. E., Thapa, R. B., and Cherrington, E. (2019). **The SAR Handbook**: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation, 1st Edn. Huntsville, AL: NASA Marshall Space Flight Center.

SHAHARUM, Nur Shafira Nisa *et al.* Oil palm mapping over Peninsular Malaysia using Google Earth Engine and machine learning algorithms. **Remote Sensing Applications: Society and Environment**, volume 17, 100287, 2020.

SOTHE, Camile & Almeida, Cláudia & Liesenberg, Veraldo & Schimalski, Marcos. Evaluating Sentinel-2 and Landsat-8 Data to Map Sucessional Forest Stages in a Subtropical Forest in Southern Brazil. **Remote Sensing**, Volume 9, 838, 2017.

STROMANN, Oliver; Andrea Nascetti, Osama Yousif e Yifang Ban. Dimensionality Reduction and Feature Selection for Object-Based Land Cover Classification based on Sentinel-1 and Sentinel-2 Time Series Using Google Earth Engine. **Remote Sensing**, volume 12, 76, 2020.

SULOVA, Andrea & ARSANJANI, J.J. Exploratory Analysis of Driving Force of Wildfires in Australia: An Application of Machine Learning within Google Earth Engine. **Remote Sensing**, 13(10), 2021.

TASSI, Andrea & VIZZARI, Marco. Object-Oriented LULC Classification in Google Earth Engine Combining SNIC, GLCM, and Machine Learning Algorithms. **Remote Sensing**, volume 12, 3776, 2020.

VIEIRA, I. C. G. *et al.* Classifying successional forests using Landsat spectral properties and ecological characteristics in eastern Amazonia. **Remote Sensing of Environment**, volume 87, p. 470–481, 2003.

WHYTE, Andrew; K. P. Ferentinos, G. P. Petropoulos. A new synergistic approach for monitoring wetlands using Sentinels -1 and 2 data with object-based machine learning algorithms. **Environmental Modelling & Software**, volume 104, 2018, p. 40 – 54.

WOOD, E.M *et al.* Image texture as a remotely sensed measure of vegetation structure. **Remote Sensing of Environment,** volume 121, p. 516 – 526, 2012.

WREGE, M.S. *et al.* Predicting Current and Future Geographical Distribution of Araucaria in Brazil for Fundamental Niche Modeling. **Environment and Ecology Research**, volume 4, p. 269 – 279, 2016.

XIE, Bo *et al.* Analysis of Regional Distribution of Tree Species Using Multi-Seasonal Sentinel-1&2 Imagery within Google Earth Engine. **Forests**, volume 12, 2021.

Authorship Contribution

1 Vinícius Lorini da Costa

Biologist

https://orcid.org/0009-0003-2491-9927 • viniciuslorini@gmail.com

Contribution: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Visualization; Writing – original draft; Writing – review & editing

2 Marcos Wellausen Dias de Freitas

Geographer

https://orcid.org/0000-0001-9879-2584 • 00044703@ufrgs.br

Contribution: Formal analysis; Methodology; Resources; Supervision; Validation; Writing – review & editing

How to quote this article

COSTA, V. L.; FREITAS, M. W. D. Classification of Semideciduous Seasonal Forest successional stages using Sentinel-1-2 and SRTM data on Google Earth Engine. **Ciência Florestal**, Santa Maria, v. 34, n. 2, e68716, p. 1-21, 2024. DOI 10.5902/1980509868716. Available from: https://doi.org/10.5902/1980509868716. Accessed in: day month abbr. year.