

**VITOR ALVES DA SILVEIRA**

**ANÁLISE GEOESPACIAL DE INSELBERGS NO DOMÍNIO DA MATA  
ATLÂNTICA E CAATINGA, BRASIL**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal, para obtenção do título de Magister Scientiae.

Orientador: Alexandre Rosa dos Santos

Coorientador: Carlos Ernesto G. R. Schaefer

**VIÇOSA - MINAS GERAIS  
2021**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

S587m  
2021  
Silveira, Vitor Alves da, 1989-  
Modeling and mapping of inselbergs in the domain of the a  
atlantic forest and caatinga, Brazil / Vitor Alves da Silveira. –  
Viçosa, MG, 2021.  
48 f. : il. (algumas color.) ; 29 cm.

Orientador: Alexandre Rosa dos Santos.  
Dissertação (mestrado) - Universidade Federal Viçosa.  
Referências bibliográficas: f. 40-48.

1. Aprendizado do computador. 2. Engenharia geotécnica.  
3. Sensoriamento remoto. 4. Previsão geofísica. I. Universidade  
Federal Viçosa. Departamento de Engenharia Florestal.  
Programa de Pós-Graduação em Ciência Florestal. II. Título.

CDO adapt. CDD 22. ed. 634.9585


VITOR ALVES DA SILVEIRA

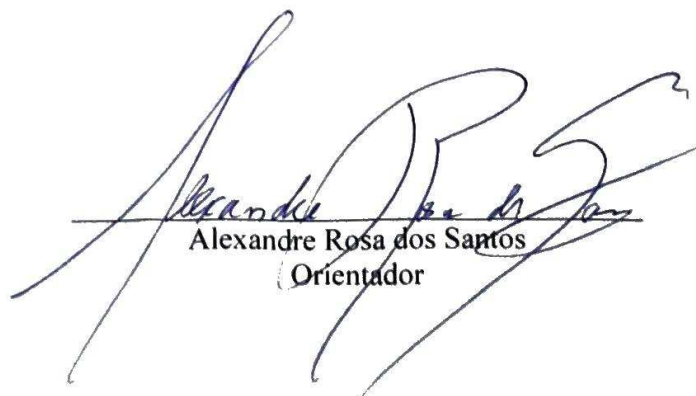
ANÁLISE GEOESPACIAL DE INSELBERGS NO DOMÍNIO DA MATA  
ATLÂNTICA E CAATINGA, BRASIL

Dissertação apresentada à Universidade Federal de  
Viçosa, como parte das exigências do Programa de  
Pós-Graduação em Ciência Florestal, para a  
obtenção do título de *Magister Scientiae*.

APROVADA: 22 de fevereiro de 2020.

Assentimento:

  
Vitor Alves da Silveira  
Autor

  
Alexandre Rosa dos Santos  
Orientador

Aos meus pais, Roberto Telles da Silveira e Leila Alves da Silveira,  
pelo amor, educação, paciência, apoio e oportunidades.

À minha irmã Júlia Alves da Silveira.

À minha filha Ana Luiza Nascimento da Silveira,

Aos meus avós e toda minha família.

À Universidade Federal Viçosa.

## **DEDICO**

Em meandros e chapadão formam uma linda trama,  
Aplainamento todo sedento por cascalho arredondado,  
Que não fica parado...  
As nuances, elevações, variações de clima úmido para clima seco.  
Ou vice-versa, de tudo isso e mais um pouco.  
Segue-se entre cânions o leito menor,  
Por dentro o talvegue,  
Acima, o leito maior se esbaranha,  
De forma doce, com altíssimo terraço.  
O detentor de hematita, caulinita e feldspato,  
Esse, o mais rico em Potássio.  
A colmatagem, tratado da coverdade,  
Ultrapassa a curva e continuam as tardes, os sons  
E as areias nas paisagens típicas de Inselbergs  
(Vitor Alves da Silveira)

## AGRADECIMENTOS

Aos meus familiares, sobretudo meus Pais, por me inspirarem todos os dias e pelo apoio de sempre.

A Deus e a vida por mais uma realização.

Ao meu Orientador Alexandre Rosa dos Santos e Coorientador Carlos Ernesto Schaefer pelos ensinamentos e conhecimentos transmitidos, excelentes pesquisadores internacionais, uma honra tê-los como meus orientadores e próximos de mim, ao longo dessa pesquisa.

Aos servidores e terceirizados da Universidade Federal de Viçosa - UFV pelo cuidado e dedicação à Universidade e conseqüentemente conosco, docentes e discentes.

Aos professores do Programa de Pós-Graduação em Ciência Florestal, especialmente o Alexandre Simões Lorenzon, pelas dicas preciosas e conversas ao longo do mestrado e também, aos professores Márcio Rocha Francelino e Elpídio Inácio, do Departamento de Solos da UFV pelas excelentes contribuições ao longo do curso.

Ao pesquisador e programador Ganso, pela paciência, parceria e boa vontade de me ajudar, principalmente na parte metodológica desse trabalho, sem a sua contribuição não seria possível obter dados valiosos nessa pesquisa.

A todos que me ajudaram nessa pesquisa, em especial ao estagiário Hígor Brandão, pela grande parceria de pesquisa e amizade, sem você este trabalho não seria o mesmo.

Ao Herval, Wendelo, Otávio, Matheus, Dani, Prímula e Dudu pelos ensinamentos, amizade e bela equipe que formamos no trabalho de campo pelo Nordeste de Minas Gerais e Sul da Bahia para validar os dados dessa pesquisa e gerar novos dados de vegetação e solos inéditos para essa região do Brasil. Deixo aqui meu reconhecimento e gratidão por vocês e pelo professor Carlos Ernesto G. R. Schaefer por nos acompanhar, orientar e possibilitar esse encontro único e científico em nossas vidas.

Ao Projeto Mata Seca pelo financiamento da expedição de campo na região Nordeste de Minas Gerais e Sul da Bahia.

“O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001

## RESUMO

SILVEIRA, Vitor Alves da, M.Sc., Universidade Federal de Viçosa, fevereiro de 2021. **Modelagem e mapeamento de inselbergs no domínio da Mata Atlântica e Caatinga, Brasil.** Orientador: Alexandre Rosa dos Santos. Coorientador: Carlos Ernesto Gonçalves Reynaud Schaefer.

Os Inselbergs são afloramentos de rochas de proporções monolíticas, abrigando níveis altos de diversidade de plantas e endemismo em diversas regiões geográficas do globo. No Brasil encontra-se uma das maiores concentrações de Inselbergs do planeta, e pesquisas são necessárias para evidenciar locais representativos para conservação. Nesse contexto, o presente estudo avaliou o desempenho de algoritmos de aprendizado de máquina no mapeamento de Inselbergs. Especificamente, buscou-se prever e espacializar os Inselbergs usando algoritmos, selecionando covariáveis importantes na espacialização desses afloramentos rochosos em gradiente climático do domínio atlântico do Brasil. Técnicas de classificação de imagens, algoritmos de aprendizado de máquina e modelagem geoespacial foram usadas para mapear a distribuição dos Inselbergs e selecionar áreas relevantes para conservação ecológica e ambiental desses ambientes ameaçados por pressões antrópicas e mudanças climáticas. De acordo com os resultados, os modelos estudados foram classificados com concordância substancial – Substantial (0,61-0,80) a quase perfeito – Almost Perfect (0,81-1,00) na classificação. Os algoritmos GBM, svmRadialSigma, C5.0 e RF apresentaram os melhores resultados na aproximação da realidade no mapeamento das paisagens típicas de Inselbergs estudadas. A metodologia utilizada neste estudo revela grande potencial de uso para subsidiar decisões de seleção de paisagens típicas de Inselbergs no contexto dos biomas Mata Atlântica e Caatinga para proteção e gestão ambiental eficiente da biodiversidade e geodiversidade. A metodologia proposta pode ser adaptada à diferentes áreas e biomas do mundo, com sucesso.

**Palavras-chave:** Aprendizado de máquina. Geotecnologia. Sensoriamento Remoto. Predição.

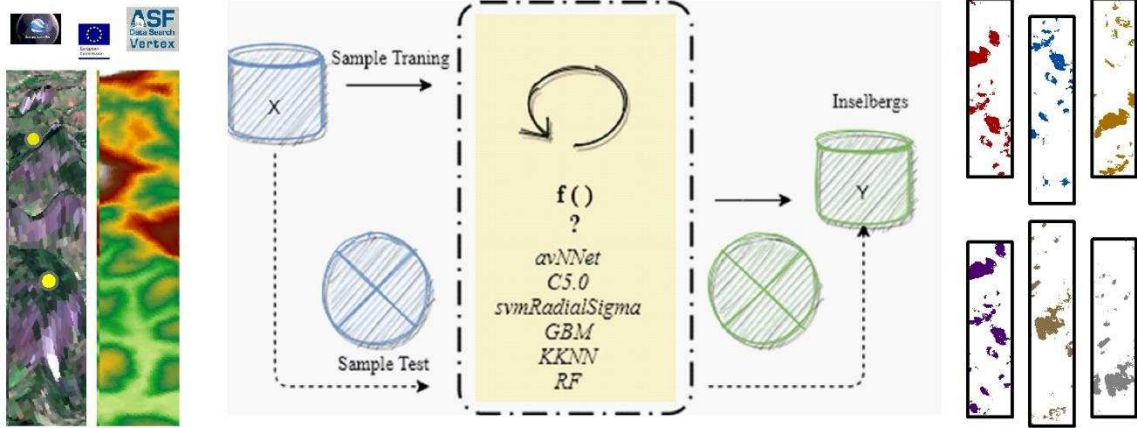
## ABSTRACT

SILVEIRA, Vitor Alves da, M.Sc., Universidade Federal de Viçosa, February, 2021. **Modeling and mapping of Inselbergs in the domain of the Atlantic Forest and Caatinga, Brazil.** Advisor: Alexandre Rosa dos Santos. Co-advisor: Carlos Ernesto Gonçalves Reynaud Schaefer.

Inselbergs are outcrops of rocks of monolithic proportions, harboring high levels of plant diversity and endemism in various geographical regions of the globe. Brazil has one of the largest concentrations of Inselbergs on the planet, and research is needed to highlight representative sites for conservation. In this context, the present study evaluated the performance of machine learning algorithms in Inselberg mapping. Specifically, we sought to predict and spatialize the Inselbergs using algorithms, selecting important covariates in the spatialization of these rocky outcrops in the climatic gradient of the Atlantic domain of Brazil. Image classification techniques, machine learning algorithms and geospatial modeling were used to map the distribution of Inselbergs and select areas relevant to ecological and environmental conservation of these environments threatened by anthropogenic pressures and climate change. According to the results, the studied models were classified with substantial agreement - Substantial (0.61-0.80) to almost perfect - Almost Perfect (0.81-1.00) in the classification. The GBM, svmRadialSigma, C5.0 and RF algorithms showed the best results in approaching reality in mapping the typical Inselbergs landscapes studied. The methodology used in this study reveals great potential for use to support decisions on the selection of typical Inselberg landscapes in the context of the Atlantic Forest and Caatinga biomes for the protection and efficient environmental management of biodiversity and geodiversity. The proposed methodology can be successfully adapted to different areas and biomes in the world.

**Keywords:** Machine Learning. Geotechnology. Remote Sensing. Prediction.

# RESUMO GRÁFICO





## SUMÁRIO

1. INTRODUCTION .....	10
2. MATERIAL AND METHODS .....	12
2.1. Physical and natural aspects of the study area .....	12
2.2. Geology .....	13
2.3. Inselbergs modeling and prediction methodology .....	14
2.3.1. General Settings .....	14
2.4. Spatial database .....	15
2.4.1. Images Sentinel-2/MSI .....	15
2.4.2. Images ALOS PALSAR .....	16
2.5. Separation of bands and generation of spectral indices .....	16
2.6. Generation of topographic covariates .....	16
2.7. Sample collections .....	18
2.8. Selection of covariates .....	18
2.8.1. Removing covariates with variance close to zero .....	18
2.8.2. Removal of covariates by correlation .....	18
2.8.3. Removing covariates by importance .....	19
2.9. Algorithms .....	19
2.10. Inselbergs training and prediction .....	20
2.11. Maps and final results .....	21
2.12. Assertiveness Index of Algorithms – IAA .....	21
3. RESULTS .....	22
3.1. Covariables selected by RFE for each algorithm .....	22
3.2. Evaluation of Inselberg mapping with machine learning algorithms .....	24
3.3. Final maps and prediction accuracy .....	26
3.4. Final maps of the GBM algorithm .....	32
3.5. Algorithm Assertiveness Index – IAA .....	33
4. DISCUSSION .....	34
4.1. Importance of selected covariates in the spatialization of Inselbergs .....	34
4.2. Inselbergs precision performance by machine learning algorithms .....	37
4.3. Precision accuracy maps and maps of Inselbergs prediction .....	37
4.4. Algorithm Assertiveness Index – IAA .....	39
5. CONCLUSION .....	39

Declaration of Competing .....	39
Acknowledgments .....	39
References .....	40

## 1. Introduction

Brazil has one of the highest concentrations of Inselbergs of the world and according to Ab'Sáber (2003) on the Atlantic façade, from humid to semi-arid, typical Inselbergs landscapes occur. These are preferably distributed to the north of the Planalto da Borborema, in Quixadá (Ceará), Patos (Paraíba), Itatim (Bahia), Seridó region (Rio Grande do Norte) and Pernambuco (Jatobá, 1994; Maia et al., 2016; Maia and Nascimento, 2018; Rodrigues et al., 2019).

The largest and most significant areas of Inselbergs are located along the Atlantic Mobile Belt, whose main distribution area in the Atlantic Forest Biome of Southeast of Brazil (Ab'Sáber, 1967; Schaefer, 2012). The “sugar loaf” model, also known as Bornhardt, is typical and found in large concentrations in this region, mainly in the state of Espírito Santo, Rio de Janeiro, Northeast Minas Gerais and South Bahia (Paula et al., 2016).

Plant communities in Inselbergs are functionally ignored, and more research is needed to enhance knowledge about species distribution and selection of representative sites for conservation measures (Paula et al., 2016). Geomorphological features in semi-arid Inselbergs, especially on the region of Milagres (Bahia), are still poorly studied (Rios, 2017). In Northeast of Minas Gerais, these rocky outcrops are home to the last remnants of forest fragments in their surroundings, due to the difficulty of access, forming important forest fragments (Paula et al., 2017). The entire region is representative of Inselbergs, and needs scientific attention due to the large gap in biological, geomorphological and geospatial investigations (Oliveira-Filho et al., 2005). In the study area, there are still no conservation units representative of these environments to ensure the protection of natural resources (Paula et al., 2017), especially the conservation of the last remnants of Atlantic Forest and Caatinga Forest in the vicinity of rock outcrops.

Environmental mapping and land cover monitoring for conservation purposes are one of the main applications of Earth observation satellite sensor data (Galiano et al., 2012). For this, a variety of classification methods have been used (Galiano et al., 2012). Satellite image classification techniques can be performed with unsupervised algorithms, such as K-means or ISODATA, supervised parametric algorithms, such as Maximum Likelihood – ML (Jensen, 2005) and various alternative machine learning algorithms, such as Support Vector Machines (SVM) (Cortes and Vapnik, 1995; Mountrakis et al., 2011), Random Forest – RF (Breiman, 2001), Gradient Boosting Machine – GBM (Friedman, 2001a), C5.0 (Quinlan, 2004), KNN (Hechenbichle and Schliep, 2004), Artificial neural networks (Mas and Flores, 2008) and

avNNet (B.D. Ripley, 1996) which are used in different areas of scientific knowledge, from spatial modeling research to tropical or temperate soils (Adhikari et al., 2013; Bonfatti et al., 2016; Gomes et al., 2019; Ließ et al., 2016), texture modeling and apparent density of soils (Adhikari et al., 2013), soil classification (Brungard et al., 2015), up to slope stability assessment studies (Zhou et al., 2019) and landslide susceptibility mapping (Huang et al., 2020).

Machine learning approaches have become widely accepted, as evidenced by their use in mapping land cover and large areas (Huang et al., 2020; Maxwell et al., 2018). The use of these algorithms allows us to consider a wide variety of predictive covariates (Souza et al., 2018). Studies show that these algorithms proved to be superior to conventional parametric classifiers, for example, ML, by presenting greater precision and better performance in supervised classification, especially due to the efficiency in processing large databases (Gašparović and Jogun, 2018; Ghimire et al., 2012; Shi and Yang, 2016; Yu et al., 2014; Zeraatpisheh et al., 2017). Due to the excellent ability to deal with nonlinear relationships between dependent and independent variables (Voyant et al., 2017), the machine learning algorithms employed in the supervised classification of satellite images proposed in this research are efficient and capable of modeling spectral signatures of different classes of land use and coverage (Emadi et al., 2020; Maxwell et al., 2018).

Given the environmental, cultural and social relevance of the typical landscapes of Inselbergs in Minas Gerais and Bahia, associated with improvements in satellite imaging technology, it is now possible to carry out mappings of large areas, in a more accurate and systematic way (Friedl et al., 1999; Shendryk et al., 2019) for the purpose of protecting and conserving strategic natural resource for global biodiversity.

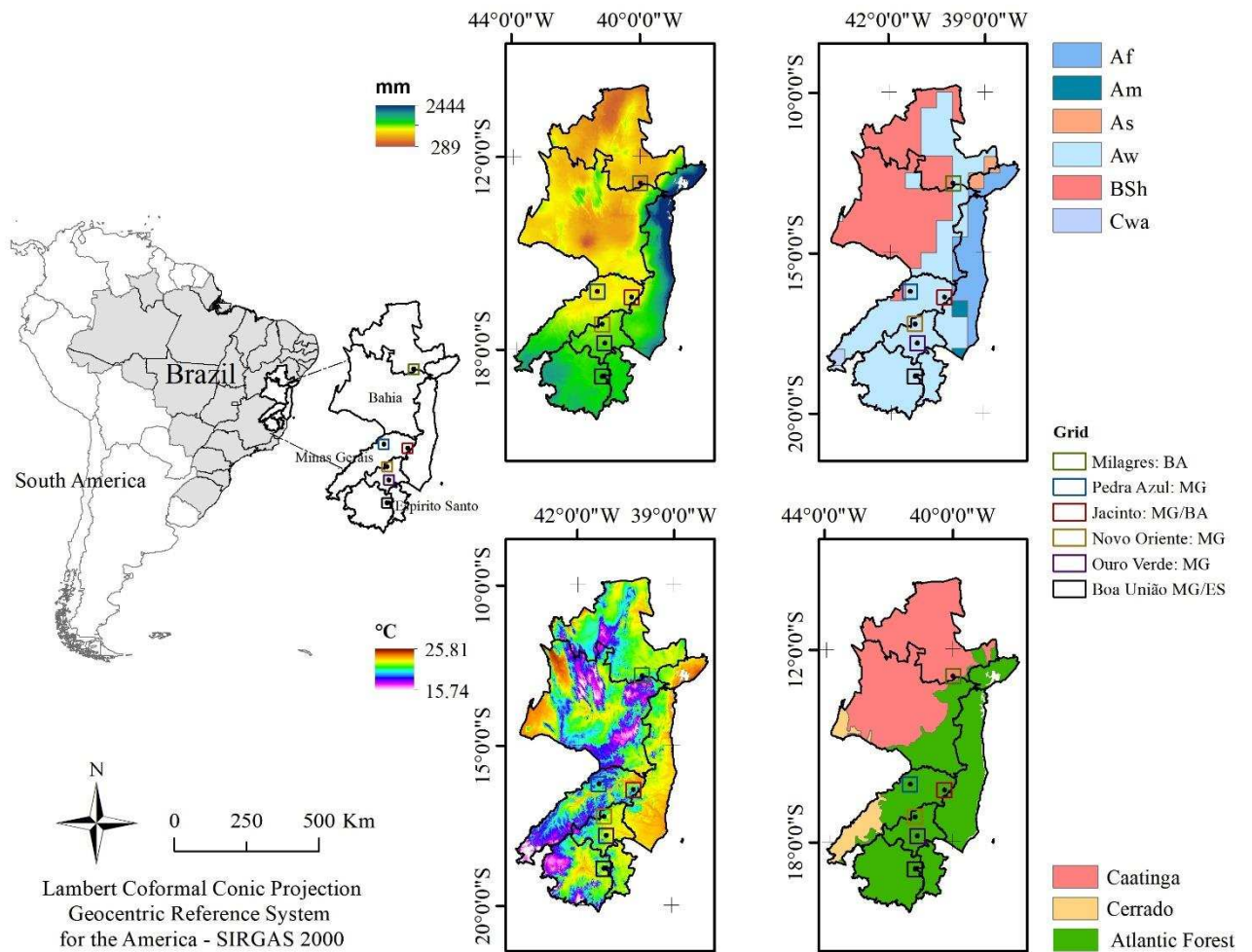
Given the above, the general objective of this study was to map the Inselbergs on the Brazilian Atlantic façade, along a climatic gradient. Specifically, we aim to: i) predict and spatialize Inselbergs using machine learning algorithms with remote sensor data; ii) select the most assertive model in the prediction of the Inselbergs; and iii) select the environmental covariates by importance for each algorithm. Final maps were also generated, from the results found in the classification for the best final agreement algorithm with the Sentinel-2 / MSI image for the entire study area.

## 2. Material and Methods

### 2.1. Physical and natural aspects of the study area

The climatic gradient of the study area is located between latitudes 9°35'0''S and 20°18'20''S and longitudes 43°17'30''W and 37°55'50''W, at Minas Gerais state (Jequitinhonha Valley, Mucuri Valley and Rio Doce Valley), Espírito Santo (West) and Bahia states (Center North and Center South), Brazil (Fig. 1). Six square grids with 50km x 50km each were selected, arranged in a south-north gradient (wet-dry), totaling 15,000 km<sup>2</sup> of research. The average elevation between the grid ranges from 130 meters in the Boa União grid to 1235 m in the Jacinto grid.

According to the Köppen-Geiger climate classification, the predominant climates in the grid vary from Aw: hot and humid tropical, with rainy season in summer and dry season in winter, to BSh: humid semi-arid climate. The average annual precipitation is 912 mm, with a maximum of 1255 mm (Jacinto - Minas Gerais) and a minimum of 504 mm (Milagres - Bahia), distributed between the months of November to April and followed by a long period of drought (Jémisson Mattos dos Santos, 2010). The average annual temperature is 22.66 °C, with a maximum of 24.67 °C and a minimum of 19 °C. The studied area is under the domain of the Atlantic Forest and Caatinga biome (Fig. 1).



**Fig. 1.** Map of the study area with bioclimatic variables annual precipitation, average annual temperature, climatic and phytogeographic domains.

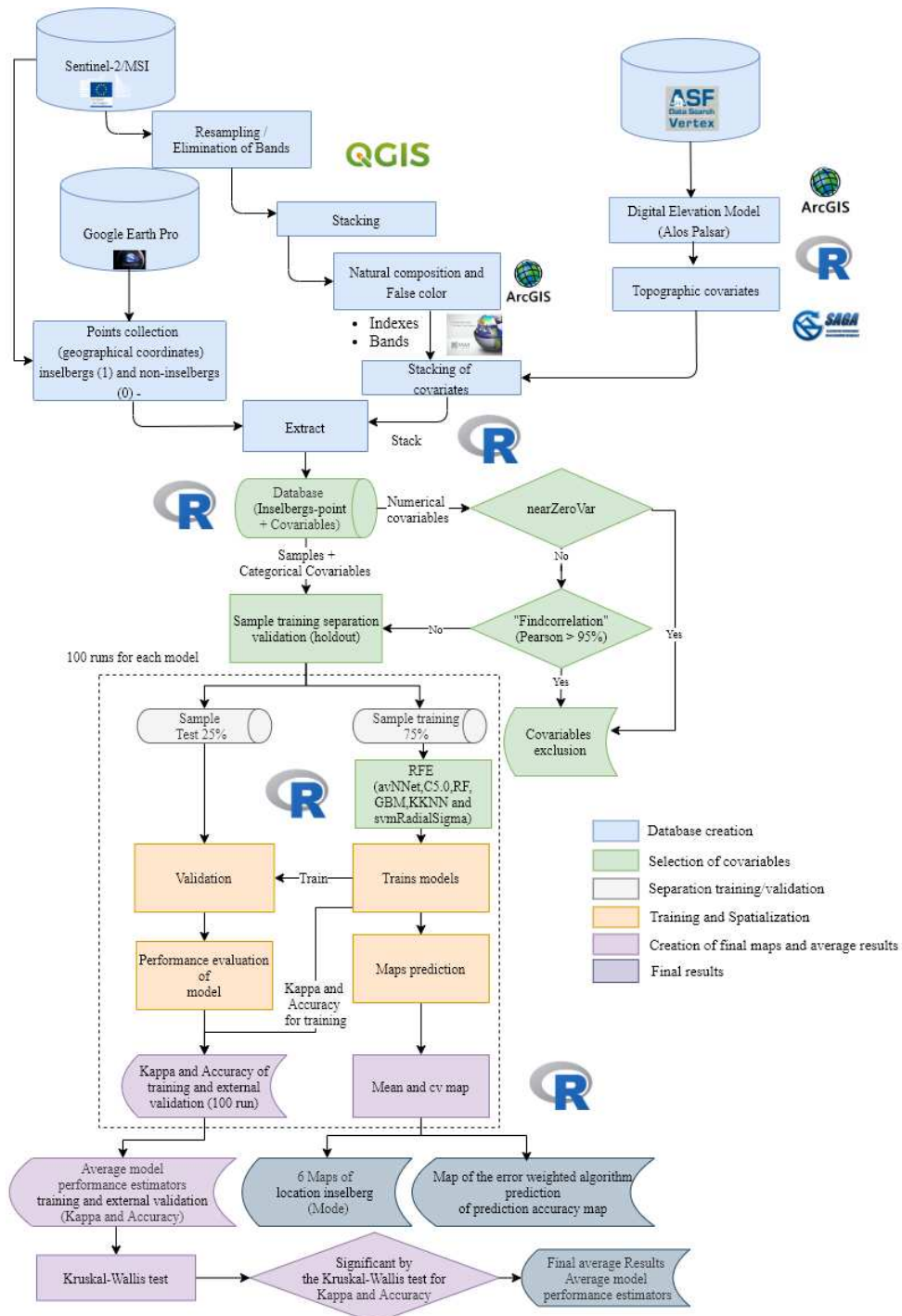
## 2.2. Geology

The Atlantic Forest Coastal zone, dissected by rivers that drain directly into the Atlantic Ocean, has several residual granitic nuclei dotting the landscape, represented by sugar-loaf forms and Inselbergs. They are true witnesses of drier paleoclimates, when erosive mechanisms exhumed the resistant granitic, exposing them as residual mountains (Schaefer, 2012). On the other hand, in the northeastern landscape - semi-arid climates, these resistant granite mountains-islands show a heritage of wetter paleoclimates (Bigarella and Andrade, 1964), revealed by the presence of deeps Oxisols on some flat tops (Schaefer, 2012).

### 2.3. Inselbergs modeling and prediction methodology

#### 2.3.1. General Settings

The modeling of Inselbergs was carried out using machine learning (ML) resources. The methodological framework includes techniques for selecting covariates by correlation and importance (Fig. 2). The methodology also includes statistical methods to optimize the result of determining uncertainty of the prediction of the Inselbergs.



**Fig. 2.** Methodological flowchart of data processing.

## 2.4. Spatial database

### 2.4.1. Images Sentinel-2/MSI

Twelve 100x100 km<sup>2</sup> orthoimages were used for the first semester of 2019 of the multispectral sensor (Multispectral Instrument - MSI), installed on board the Sentinel-2 Satellite and produced by the European Space Agency (ESA) (Table 1). For the Milagres grid, 4 images were mosaicked and for the remaining grid 2 images. Sentinel-2/MSI has 13 spectral bands, of high and medium spatial resolution (four 10 m bands, six 20 m bands and three 60 m bands) and radiometric resolution of 12 bits per pixel (ESA, 2019). In the resampling of pixels, bands 2, 3, 4, 8 were used, with 10m resolution and bands 5, 6, 7, 8A, 11 and 12 of 20m of spatial resolution for a 12.5m grid. The objective was to make all bands compatible in the same spatial resolution and make them compatible with the MDE. It is worth noting that resampling does not interfere with image quality (Reis et al., 2019).

Band number	Central wavelength	Bandwidth	Previous order	New order
2 - Blue	492.4 nm	66 nm	2	1
3 - Green	559.8 nm	36 nm	3	2
4 - Red	664.6 nm	31 nm	4	3
5 - Red Edge 1	704.1 nm	15 nm	5	4
6 - Red Edge 2	740.5 nm	15 nm	6	5
7 - Red Edge 3	782.8 nm	20 nm	7	6
8 - Near Infra-Red	832,8 nm	106 nm	8	7
8A - Red Edge 4	864.7 nm	21 nm	8A	8
11 - Short Wave Infra-Red (SWIR-1)	1613.7 nm	91 nm	11	9
12 - Short Wave Infra-Red (SWIR-2)	2202.4 nm	175 nm	12	10

**Table 1**

Spectral bands for the SENTINEL-2A/MSI sensor.

Using the QGIS Software, after the resampling process to the 12.5m resolution, the bands were stacked - Layer Stacking (Accumulated Layer) to a single GEOTIFF file. Then the images were cut into 50 km x 50 km squares referring to study areas, using the ArcGIS 10.3.1 Software. The same clipping procedure was performed for the scenes of the MDE.



### 2.4.2. Images ALOS PALSAR

Eighteen scenes of the Digital Elevation Model - MDE ALOS PALSAR, four scenes for the Milagres, Hyacinth, Blue Stone and two scenes for the Novo Oriente and Ouro Verde squares were used, all with 12.5 m of spatial resolution. PALSAR is a synthetic aperture radar that operates in the L Band, obtains day or night images of the Earth's surface regardless of weather conditions and generates radiometrically corrected terrain products (RTC), that is, it corrects the geometry and radiometry of the opening radar synthetic to produce a superior product for scientific applications (Jaxa, 2020; Sena et al., 2020).

### 2.5. Separation of bands and generation of spectral indices

From the Sentinel-2/MSI images, 14 spectral covariables were generated for each study region. The spectral covariates were obtained using Sentinel-2 / MSI images in the transition from the rainy to the dry period, between March and July. The separate bands of interest were: BLUE, GREEN, RED, NIR, SWIR 1, SWIR 2. In the ENVI 5.0 Software, spectral indices were calculated NDVI - Normalized Difference Vegetation Index, SAVI - Soil Adjusted Vegetation Index, OSAVI - Optimized Soil Adjusted Vegetation Index, RENDVI - Red Edge Normalized Difference Vegetation Index, CMR - Clay Minerals Ratio, FMR - Ferrous Minerals Ratio, GNDVI - Green Normalized Difference Vegetation Index e MNDWI - Modified Normalized Difference Water Index.

### 2.6. Generation of topographic covariates

The generation of topographic covariates was performed in Software R (v 3.5.3), using the interface with the RSAGA package (Brenning, 2008). In the R environment, the MDE was used to generate **48** topographic covariates (morphometric maps), including the MDE itself for stacking. Maps include aspect, curvatures, hills, valleys and other information on the topography of the terrain (Gomes et al., 2019; Olaya and Conrad, 2009). After stacking the covariates, that is, the concatenation of several raster units in a single raster, data extraction was performed to separate and select the Inselbergs (Table 2)

<b>Spectral and topographic variables</b>	<b>Brief Description</b>
Band 1	Blue
Band 2	Green
Band 3	Red
Band 7	Near Infra-Red
Band 9	Short Wave Infra-Red (SWIR-1)
Band 10	Short Wave Infra-Red (SWIR-2)
Savi	Soil Adjusted Vegetation Index
Ndvi	Normalized Difference Vegetation Index
Osavi	Optimized Soil Adjusted Vegetation Index
Rendvi	Red Edge Normalized Difference Vegetation Index
Cm	Clay Minerals

Fmr	Ferrous Minerals Ratio
Gndvi	Green Normalized Difference Vegetation Index
Mndwi	Modified Normalized Difference Water Index
Aspect	Slope orientation
Convergence Index	Convergence Index
Cross Sectional Curvature	Transverse Curvature
Curvature classification	Curvature classification
Difference	Difference in hydrological gradient
Diffuse Solar Radiation (January)	Diffuse January radiation
Diffuse Solar Radiation (June)	June diffuse radiation
Digital Elevation Model	Digital elevation model
Direct Solar Radiation (January)	January direct radiation
Direct Solar Radiation (June)	June direct radiation
Diurnal Anisotropic Heating	Anisotropic daytime heating
Easternness	Sine of the slope orientation
Flow Line Curvature	Flow line curvature
General Curvature	General curvature
Gradient	Hydrological gradient
Hill	Hills
Hill Index	Hill index
Landforms	Relief Shapes
Longitudinal Curvature	Longitudinal curvature
Mass Balance Index	Balance index between erosion and deposition
Maximal Curvature	Maximum curvature
Mid-Slope Position	Medium tilt position
Minimal Curvature	Minimum curvature
Morphometric Protection Index	Protection index of a point in relation to the surrounding relief
Multiresolution Index of Ridge Top Flatness	Index that identifies flat areas with high attitudes
Multiresolution Index of Valley bottom Flatness	Multiresolution index of lower valley leveling
Normalized Height	Vertical distance between the base and the summit of the normalized slope
Northernness	Slope orientation cosine
Planar Curvature	Planar curvature
Profile Curvature	Profile curvature
Real Surface Area	Actual cell area calculation
Slope	Declivity
Slope Height	Vertical distance between the base and the summit of the slope
Slope Index	Slope index
Standardized Height	Slope between the base and the summit of the standard slope
Surface Specific Points	Quantitative land surface points
Tangential Curvature	Tangential curvature
Terrain Ruggedness Index	Quantitative topography roughness index
Terrain Surface Convexity	Surface terrain convexity
Topographic Position Index	Vertical difference between the base and the summit of the standardized slope
Topographic Wetness Index	Topographic control index in hydrological processes
Total Curvature	Total curvature
Total Solar Radiation (January)	January total solar radiation
Total Solar Radiation (June)	June total solar radiation
Valley	Valley
Valley Depth	Valley depth calculated by vertical distance at hydrographic level
Valley Index	Hills elevation index based on Valley Depth
Vector Ruggedness Measure	Measurement of the variation in the roughness of the slope terrain

**Table 2**

Group of topographic and spectral covariates selected to predict Inselbergs.

## 2.7. Sample collections

One thousand two hundred and seventy-eight sampling points were collected with their respective geographical coordinates for the different classes of land use and land cover in the Google Earth Pro computational application. At the end of this step, 424 points were separated for the Inselbergs usage class and categorized as 1. The remaining 854 points, referring to the other classes of land use and coverage, were designated as agriculture, forestry, mining, urban area, pasture, native forest, exposed soil and water bodies. These were categorized as 0 in the classification process by machine learning algorithms.

## 2.8. Selection of covariates

Processes for selecting covariates are necessary to reduce the computational cost, remove noisy/redundant predictors and increase the model's parsimony (Muñoz-Romero et al., 2020; Seasholtz and Kowalski, 1993; Zhang et al., 2020). Using this precept, the covariate selection process was carried out in three steps: 1) removal of covariables with variance close to zero; 2) removal by correlation; and 3) removal by importance.

### 2.8.1. Removing covariates with variance close to zero

Nesta etapa foi feita a retirada das covariáveis que apresentaram variância próxima a zero. Covariáveis que apresentam variância nula ou próxima a zero, não trazem ganhos de performance na modelagem, podendo ser eliminadas. Todas covariáveis numéricas foram avaliadas pela função “nearZeroVar” do pacote Caret (Kuhn, 2019). As covariáveis que passaram por esta fase foram para segunda fase de remoção por correlação. Nesta fase nenhuma covariável foi eliminada.

### 2.8.2. Removal of covariates by correlation

In the second step, correlation removal was performed, and Pearson's correlation was calculated for all numerical covariates. The covariables that presented a correlation greater than or equal to 95% were evaluated in pairs, with the removal of the covariate that had the highest value of the sum of the absolute correlations of the other covariates that entered the entry in the removal phase by correlation. Five topographic covariates were removed: `terrain_ruggedness_index`, `solrad_total1`, `curv_cross_secational`, `solrad_total2` and `solrad_ration2` and five spectral variables: GREEN, SWIR2, NDVI, OSAVI E GNDVI. The “findCorrelation” function of the Caret package (Kuhn, 2019) was applied at this stage and the remaining covariates from this process were stacked, together with the categorical covariates. The samples were separated into 75% for training and 25% for validation (hold-out). The training group was applied in the third selection phase.

### 2.8.3. Removing covariates by importance

In this phase, the Recursive Feature Elimination - RFE technique contained in the Caret package (Kuhn, 2019) was applied. The RFE is a backward selection method that automatically reduces the number of covariates, based on the importance of the covariates to predict the studied natural phenomenon (Kuhn and Johnson, 2013a). The results generated in the RFE are specific to each algorithm, requiring the application of this technique for each algorithm you want to test. The RFE was performed using as a base the set of covariables that remained from the selection process by correlation and tested at least fourteen subsets 5,6,7 ..., 13,14,15, 20 and 35 covariables and the total number of covariables. The optimization of subsets of ideal covariates was based on repeated cross-validation (repeatedcv) with 10 folds, 10 repetitions and 5 values of each of the internal hyperparameters of each tested algorithm (tuneLength). The hyperparameters for each algorithm are described in the Caret package manual in chapter 6 (“Models described”), available at <https://topepo.github.io/caret/train-models-by-tag.html>. The metric of to choose the best subset for each model was Kappa.

### 2.9. Algorithms

Six algorithms were tested: Model Averaged Neural Network – avNNNet (B. D. Ripley, 1996), C5.0 (Quinlan, 2004), Gradient Boosting Machine – GBM (Friedman, 2001b), Random Forest – RF (Breiman, 2001), k-Nearest Neighbors – KNN (Schliep et al., 2016) e Support Vector Machines com Radial Basis Function Kernel – SVMRadialSigma (Cortes and Vapnik, 1995). These algorithms cover a large part of the families of algorithms used in prediction in surveys using a large bank of remote sensing covariates.

### 2.10. Inselberg training and prediction

Model training was done using the ideal subset in the importance selection phase (RFE). The optimization of the internal parameters of the models in the training was done using repeated cross validation (repeatedcv) with 10 folds, 10 repetitions and 5 values of each of the internal hyperparameters of each tested algorithm (tuneLength).

To evaluate the performance of the algorithms, the confusion matrix was used to derive the Kappa (k) Eq. (1) and Accuracy Eq. (2) indices (Congalton, 1991) and the precision of the presence-absence predictions was measured using the statistics: Eq. sensitivity (3) and Eq. specificity (4). Jacob Cohen, in 1960, proposed kappa statistics as a measure of agreement between evaluators on categorical variables (Hazra and Gogtay, 2017a). The kappa coefficient provides the numerical classification and depicts the degree of agreement of the data between the detection result and the basic referential truth (Morales et al., 2018). The k value ranges

from poor ( $< 0$ ) to almost perfect ( $0.8 < K \leq 1.0$ ) (Landis and Koch, 1977) (Table 3). The calculated accuracy or global accuracy index indicates the probability that the studied and classified classes correspond to the true data, also presenting values ranging from 0 to 1, according to the aforementioned k values.

The sensitivity assessment metric is equivalent to producer precision, commonly used in the remote sensing literature and represents the percentage of positive observations correctly predicted, whereas specificity is the percentage of negative observations correctly predicted (Tatem et al., 2003) in the prediction of the Inselbergs.

<b>Kappa statistic</b>	<b>Strength of agreement</b>
$< 0$	Poor
$0 - 0.2$	Slight
$0.2 - 0.4$	Fair
$0.4 - 0.6$	Moderate
$0.6 - 0.8$	Substantial
$0.8 - 1$	Almost Perfect

**Table 3**

Guidelines of Landis and Koch, 1977.

$$k = \frac{n \sum_{i=1}^c n_{ii} - \sum_{i=1}^c n_{i+} + n_{+i}}{n^2 - \sum_{i=1}^c n_{i+} + n_{+i}} \quad (1)$$

Where:  $k$  = Kappa estimate;  $n_{ii}$  = line value  $i$  and column  $i$  (observed agreement);  $n_{i+}$  = the sum of the line  $i$ ;  $n_{+i}$  = column sum  $i$  the confusion matrix (product of the marginals, the expected agreement being);  $n$  = total number of samples; and  $C$  = total number of classes.

$$Accuracy = \frac{\sum x_i}{n} \times 100 \quad (2)$$

Accuracy is the global accuracy, in which:  $x_i$  = the sum of all elements on the diagonal of the confusion matrix; and  $n$  = total number of samples.

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

Where:  $TP$  = number of true positives;  $FN$  = false negatives; and  $TP + FN$  = number of all positives in the supervised classification.

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

In which:  $TN$  = number of true negatives;  $FP$  = false positives; and  $TN + FP$  = number of all negatives in the supervised classification.

The final maps for the best models were evaluated on their prediction and estimation accuracy for the Inselbergs in the study area. The maps have more weight when choosing the model with better performance. At the end of the training, the predicted maps for each model were generated in the respective work squares.

### **2.11. Maps and final results**

The process of selecting variables in the third phase, training and map prediction was repeated 100 times with different training and validation samples. This process is important to assess the variability of the prediction, since these subsets must generate different results and, consequently, performance for each model (Kuhn and Johnson, 2013b). The final results of the performance were calculated by means of the 100 rounds. The final maps were made by calculating the mode of the 100 rounds for each pixel of the six study areas. It is worth noting that mode is a measure of position that indicates the region of maximum frequencies in a given set of values (Neto, 2002).

For the study of the uncertainty of the prediction of the Inselbergs by the models, the prediction accuracy map was also calculated. The prediction accuracy map shows the pixels where the model has always chosen the same class in the pixel every 100 times the model has been run. This map demonstrates the ability to visualize the uncertainties of the classification, showing the capacity for accuracy and precision in the prediction of the algorithms.

The Kruskal-Wallis non-parametric test was also applied to the evaluation metrics: kappa index and accuracy. Models and repetitions of 100 times that the models were run with different samples of training and testing were considered as parameters. With the best methods defined statistically, the results of the maps will be evaluated, verifying with the spatialization more consistent with the reality of the phenomenon of spatialization of the Inselbergs for the study area.

### **2.12. Assertiveness Index of Algorithms - IAA**

The IAA Eq. (5) was calculated to assess the assertiveness of the best performing algorithms in the prediction metrics, along a wet to dry climate gradient.

$$IAA = \frac{A_{modelo}}{A_{total}} \times 100 \quad (5)$$

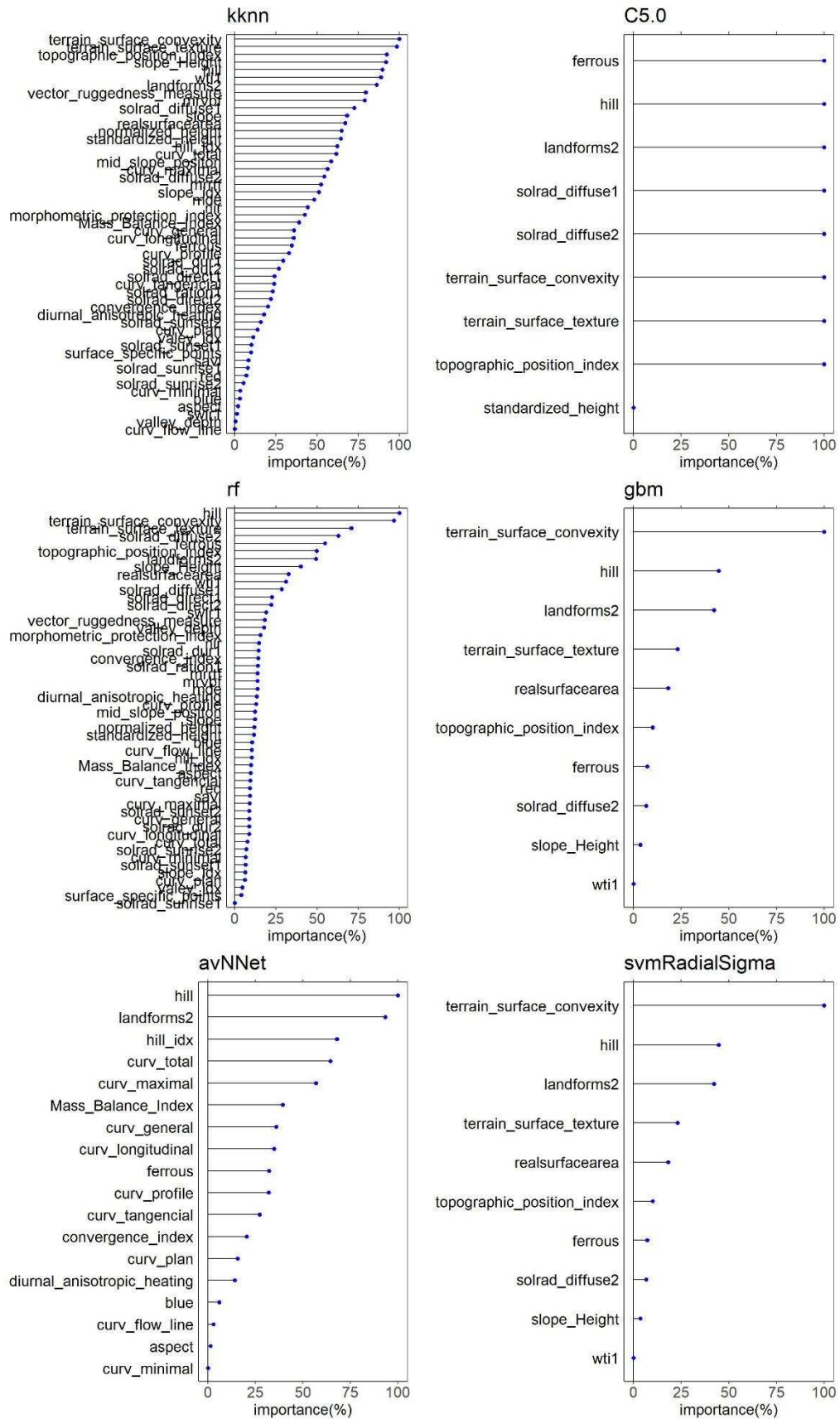
Where:  $A_{modelo}$  = Area found by the machine learning algorithm; and  $A_{total}$  = Total area of the reference Inselbergs in the analyzed grid.

### 3. Results

#### 3.1. Covariables selected by RFE for each algorithm

The contribution of the selected topographic and spectral covariates to the mapping and spatialization of the Inselbergs varied according to each machine learning algorithm, as shown in Fig. 3.

The svmRadialSigma and GBM algorithms selected the least number of predictive covariates in the selection of importance among the analyzed algorithms. Using as reference the number of eight most important covariates selected by the RFE, the topographic covariates terrain\_surface\_texture, terrain\_surface\_convexity, hill, slope\_height and the spectral covariate ferrous minerals ratio were selected by at least 3 different algorithms among the most important variables (Fig. 3). The topographic covariate Hill was important in all the algorithms analyzed in the classification of the Inselbergs: GBM - 64.91%; avNNNet - 100%; C5.0 - 100%; RF - 100%; KKNN - 91.29% and svmRadialSigma - 44.62%. Another multiselect covariate was the Ferrous Minerals Ratio, a spectral covariate, selected as important by 5 algorithms among the 6 analyzed: C5.0 - 100%, RF - 43.30%, avNNNet - 21.13% and svmRadialSigma - 7.18% GBM - 0.61% (Fig. 3).



**Fig. 3.** Metrics of importance of covariates for the main random predictors of the studied algorithms. The x-axis denotes the importance of the variable in percentage. The y-axis denotes the unique identifier for each selected covariate.



### 3.2. Evaluation of Inselberg mapping with machine learning algorithms

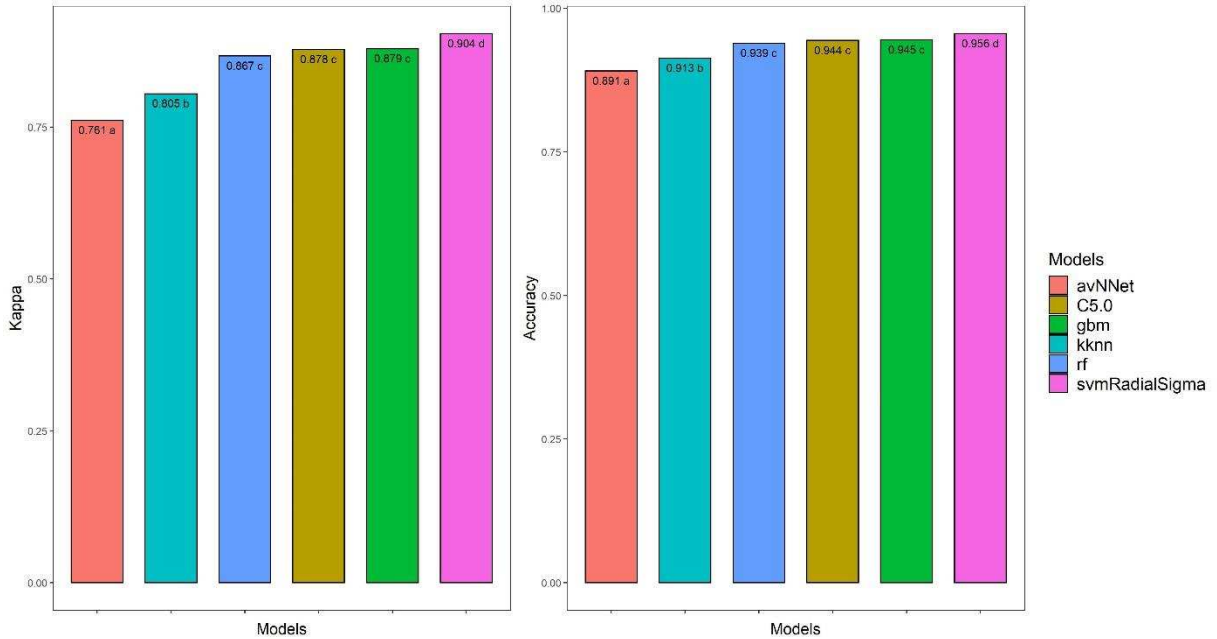
The analyzed algorithms for the classification of Inselbergs showed accuracy values ranging from 0.89 to 0.96 and kappa index, 0.76 to 0.90 (Table 4; Fig. 4). In general, svmRadialSigma and GBM showed the best performances, with kappa of 0.90 and 0.88, respectively. The kknns and avNNet algorithms showed the lowest performance for the data set used, the results are shown in Table 4 and Fig. 4.

According to the Kruskal-Wallis test performed for the evaluation metrics, the tested algorithms showed statistically significant differences in the final results (Fig. 4). The prediction statistics did not show overfitting, since the difference between the training and validation sets for accuracy and kappa index were close to zero (Table 4). This shows that the training and prediction samples showed an adequate number for the mapping. Our samples were well sampled and the MDE and the bands determined the Inselbergs well.

<b>Model</b>	<b>Kappa training</b>	<b>Kappa validation</b>	<b>Overfitting</b>
KKNN	0,8101	0,8047	-0,0054
Random Forest	0,8801	0,8668	-0,0133
AvNNet	0,7850	0,7615	-0,0235
C5.0	0,8939	0,8781	-0,0158
Gradiente Boosting Machine	0,8931	0,8791	-0,0140
svmRadialSigma	0,9064	0,9036	0,0028
	<b>Accuracy training</b>	<b>Accuracy validation</b>	<b>Overfitting</b>
KKNN	0,9149	0,9125	-0,0024
Random Forest	0,9447	0,9388	-0,0059
AvNNet	0,9017	0,8910	-0,0107
C5.0	0,9515	0,9443	-0,0072
Gradiente Boosting Machine	0,9512	0,9449	-0,0063
svmRadialSigma	0,9572	0,9560	-0,0012
	<b>Sensitivity mean</b>	<b>Specificity mean</b>	
KKNN	0,9549	0,8351	
Random Forest	0,9460	0,9255	
AvNNet	0,9159	0,8455	
C5.0	0,9578	0,9147	
Gradiente Boosting Machine	0,8628	0,8443	
svmRadialSigma	0,9860	0,9339	

**Table 4**

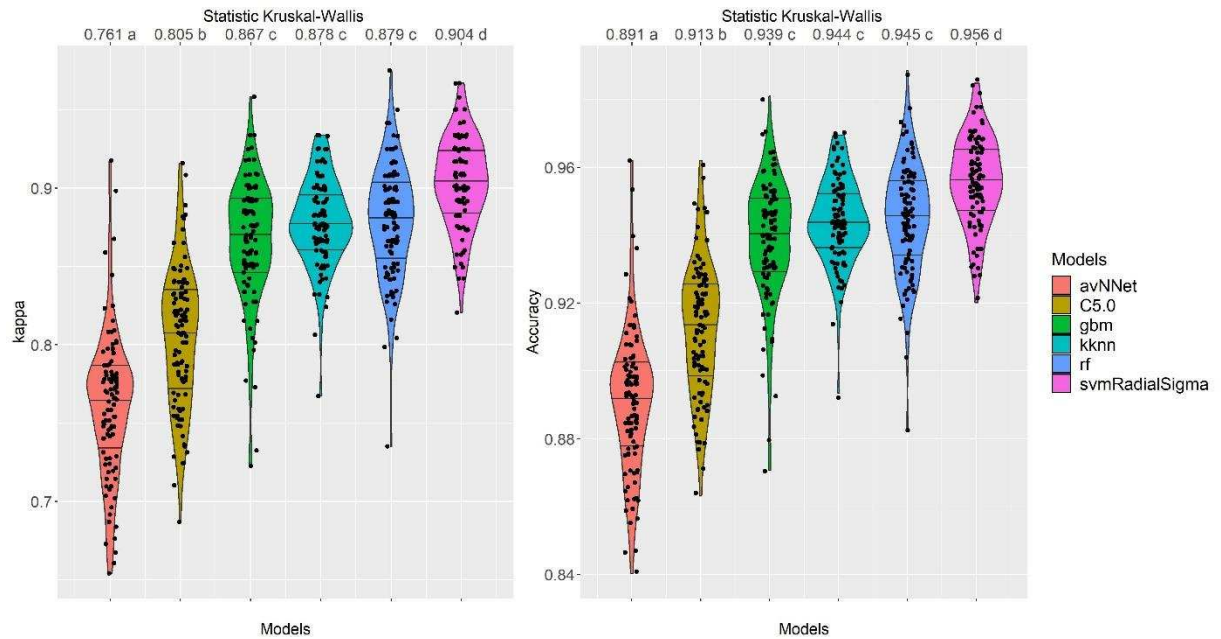
Kappa index values, accuracy, overfitting, average sensitivity and average specificity for the machine learning models analyzed.



**Fig. 4.** Validation metrics of prediction, accuracy and kappa index plotted for the respective analyzed algorithms.

All models analyzed produced specificities and sensitivities greater than 80%, that is, they showed a high capacity to avoid false negatives and positives, respectively. The svmRadialSigma algorithm showed higher mean sensitivity and specificity. GBM had the lowest sensitivity and the kknn model the lowest specificity value (Table 4).

It is possible to notice in Fig. 5 that the KKNN and avNNet algorithms presented larger box size and line length, indicating that the classification process varied much in these algorithms. In the C5.0 algorithm, a more uniform classification is noted, that is, less variation occurred during the classification of the Sentinel-2/MSI image in this algorithm. A small and less accentuated shift is observed between the median lines in the four best classifiers shown in Fig. 5.



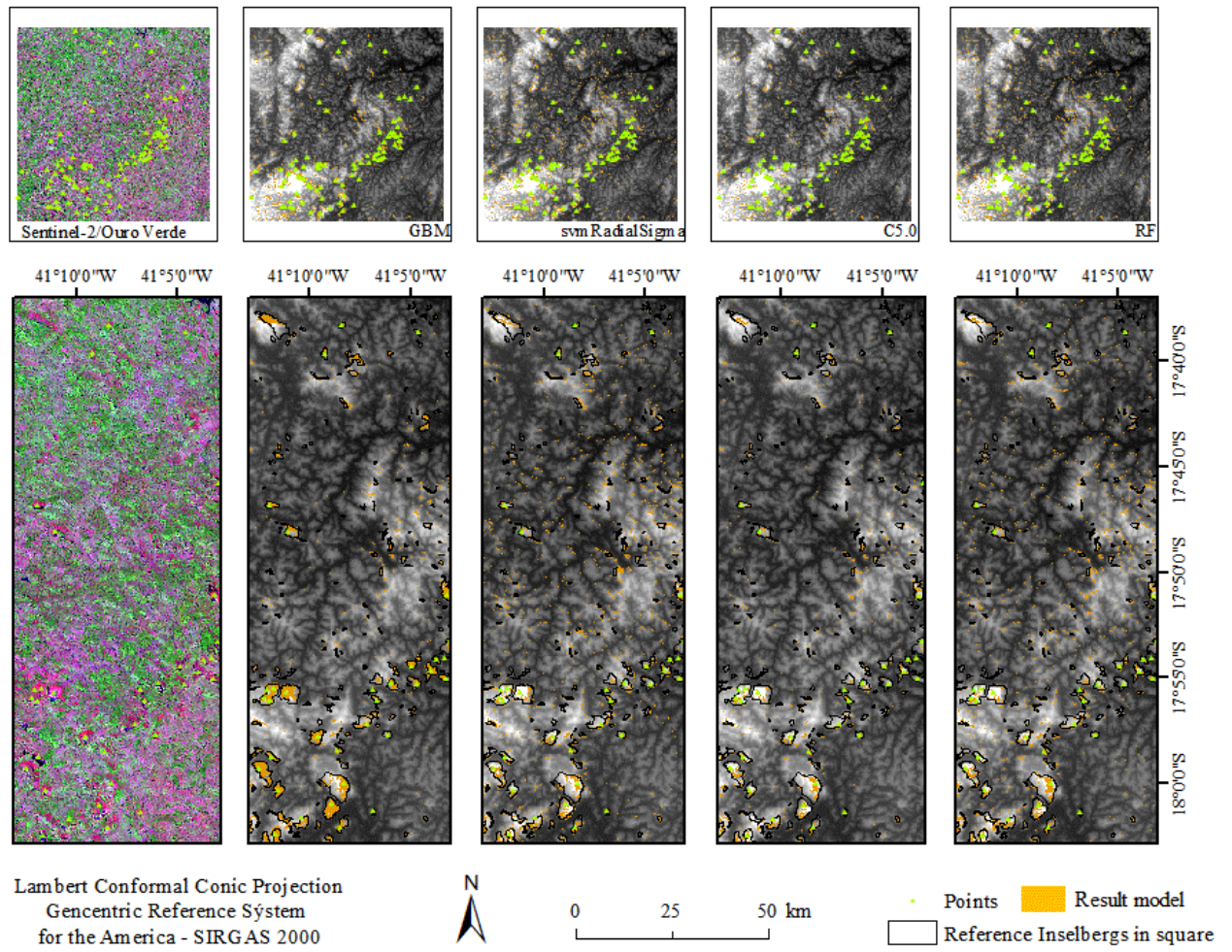
**Fig 5.** Violinsplot Chart for Accuracy and Kappa Index.

The AvNNNet presented four very good classifications (Fig. 5), but the performance was not superior to the svmRadialSigma and GBM algorithms, which showed a very good classification peak compared to the other algorithms. Random Forest had four bad ratings, but its worst rating was not lower than that found for the avNNNet algorithm, which had the lowest performance among the algorithms analyzed by the kappa index and accuracy and statistical analysis. Analyzing the outliers of each algorithm (Fig. 5), all classifiers present at least one bad classification as outlier.

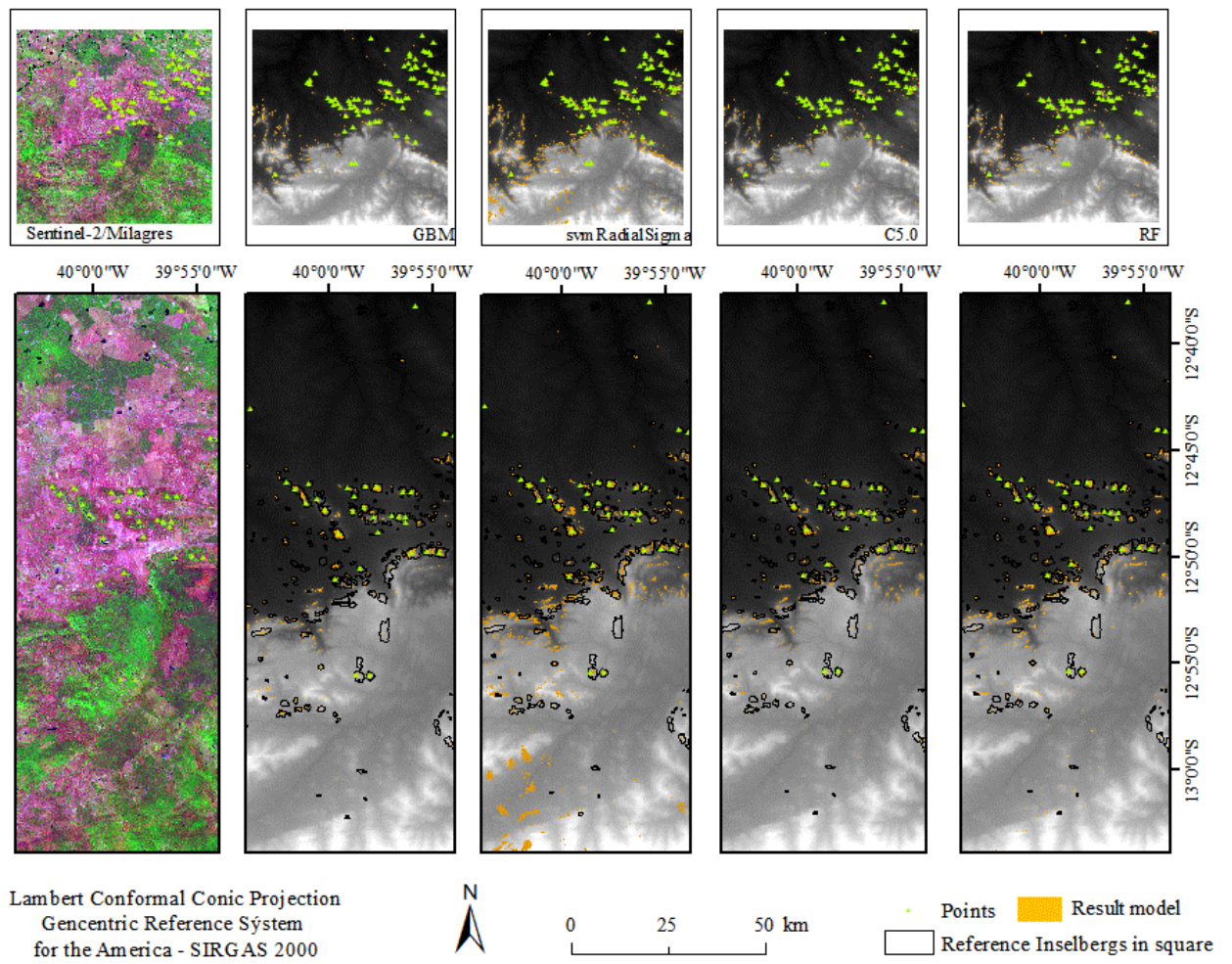
A normal distribution trend was observed for the data shown in Fig. 5. SvmRadialSigma has a more stable prediction, closer to the mean and mode, while avNNNet showed more instability in the prediction, followed by kkn, both had the lowest precision. SvmRadialSigma showed the highest accuracy and GBM, C5.0 and RF showed no statistical difference, being the second best accuracy and kappa index found.

### 3.3. Final maps and prediction accuracy

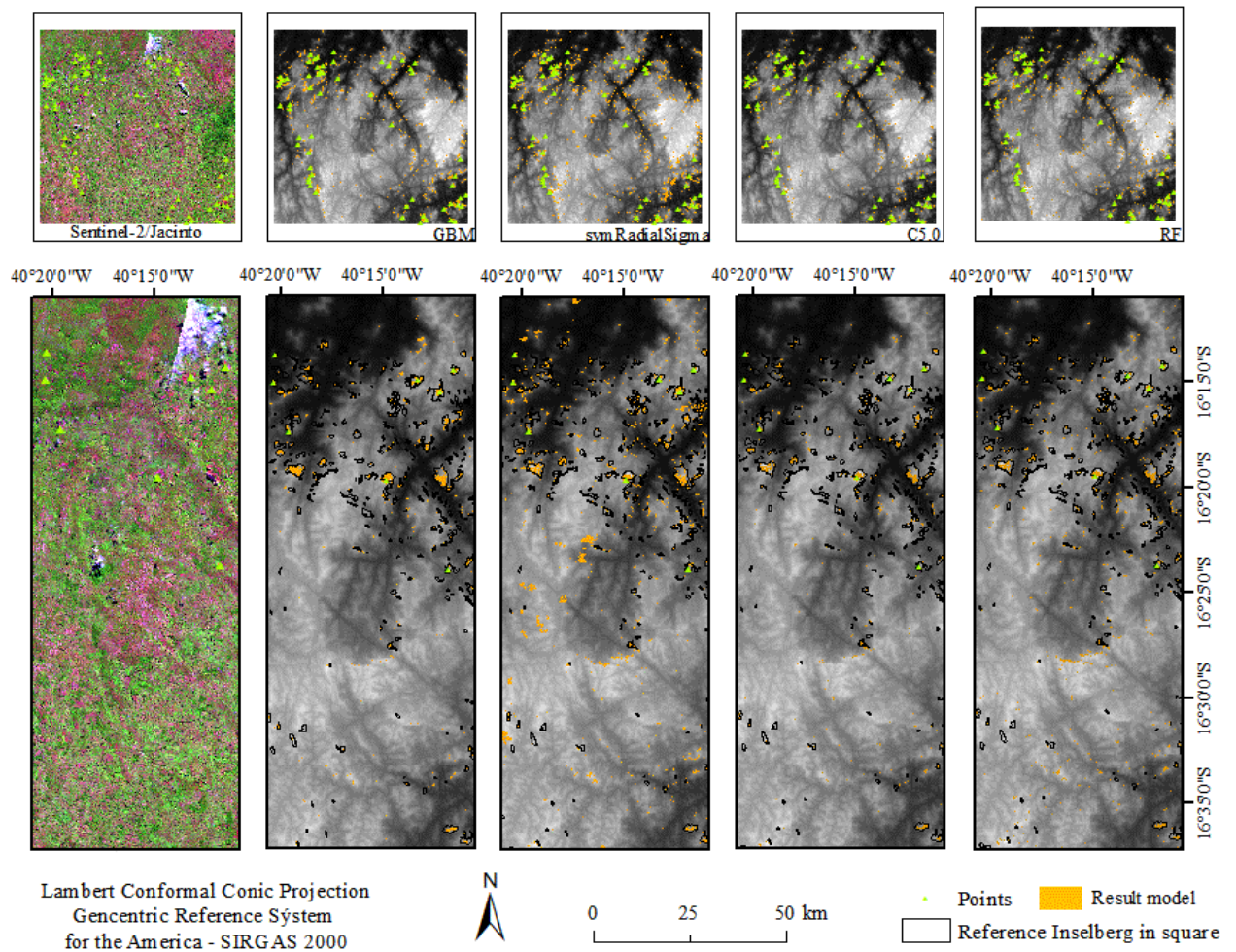
In general, machine learning algorithms showed good identification and spatialization of inselbergs. To visualize the prediction results, from the final maps, the best performance algorithms in the mapping were selected: GBM, svmRadialSigma, C5.0 and Random Forest for the Ouro Verde (Fig. 6), Milagres (Fig. 7) and Jacinto (Fig. 8) boxes.



**Fig. 6.** Maps generated by the best performance GBM, SVM, C5.0 and RF classifiers for Ouro Verde grid in Minas Gerais, Brazil.



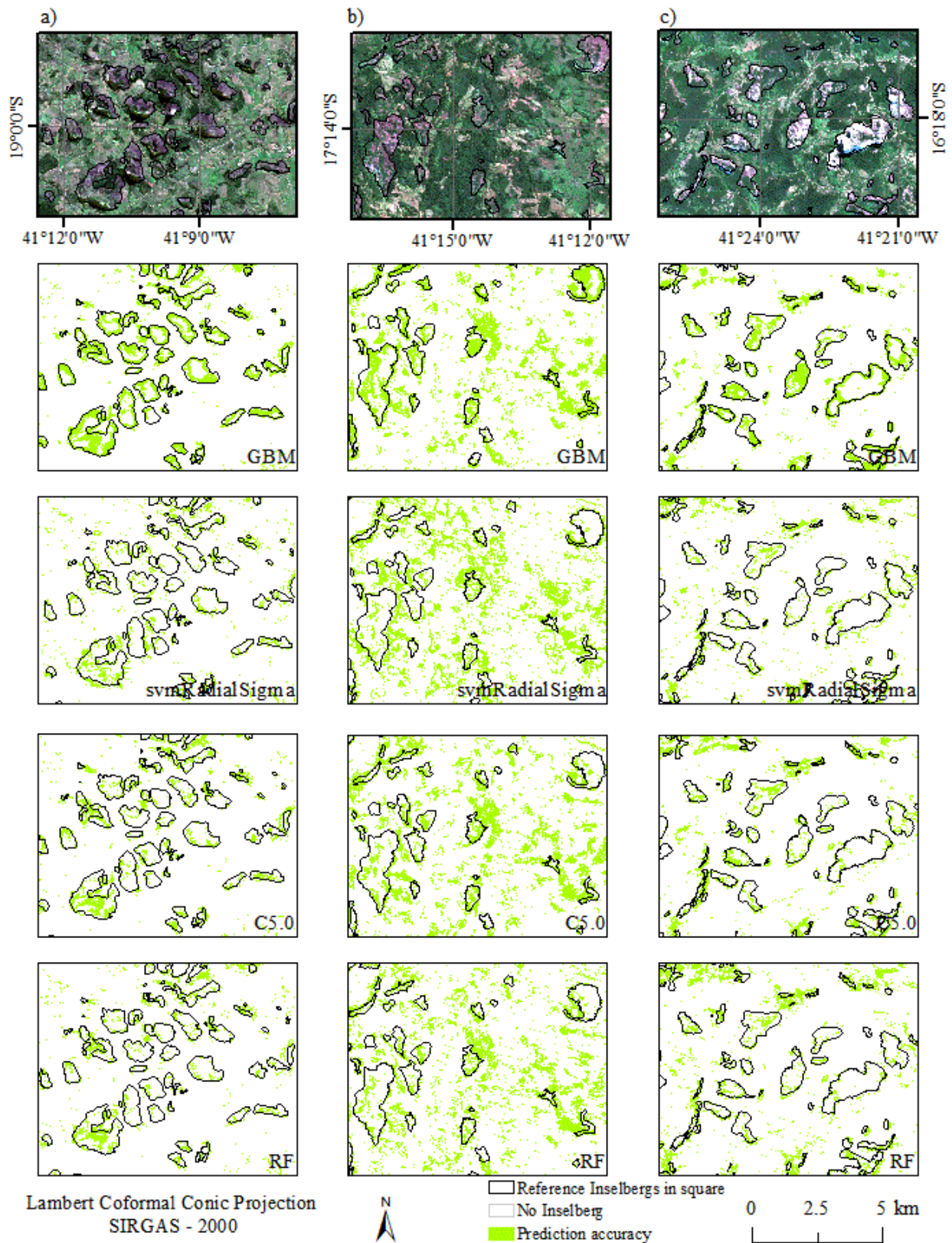
**Fig. 7.** Maps generated by the GBM, svmRadialSigma, C5.0 and RF classifiers with the best performance for Milagres grid in Bahia, Brazil.



**Fig. 8.** Maps generated by the best performance GBM, SVM, C5.0 and RF classifiers for Jacinto grid in Minas Gerais and Bahia, Brazil.

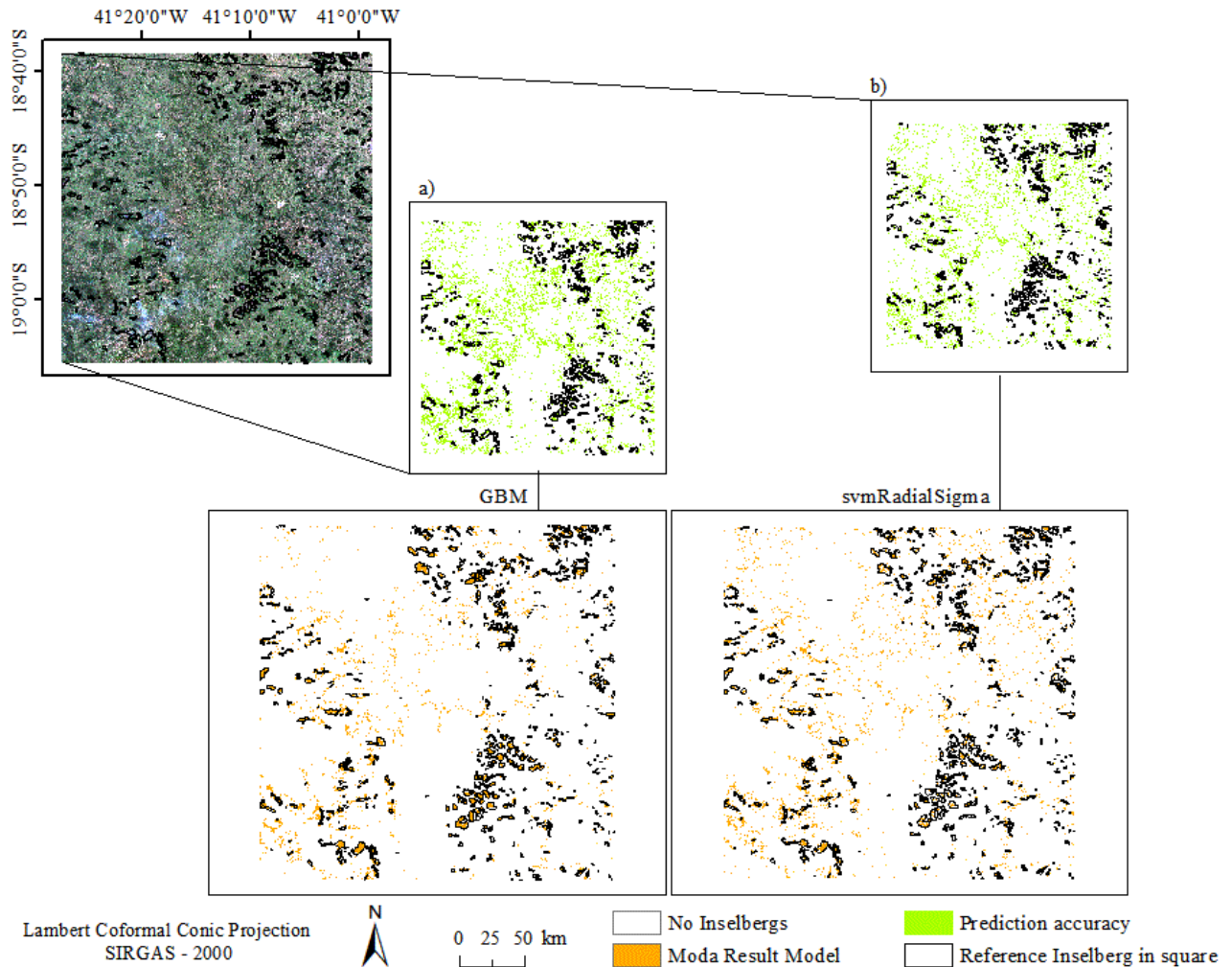
The final map of C5.0 had greater agreement in spatialization with the GBM algorithm, but it showed less differentiation between the polygons of Inselbergs (Figs. 6, 7 and 8). The RF generalized a little more and approached the Inselbergs' differentiating behavior found by svmRadialSigma.

Between the svmRadialSigma and GBM algorithms, the GBM was worse in identification. The intervals of changing accuracy of the prediction are shown in green (Fig. 9). In svmRadialSigma the spatialization of Inselbergs (green tone) indicates less dispersion in spatialization. C5.0 showed higher final agreement of accuracy with the GBM model. The svmRadialSigma model made less errors in prediction. The C5.0 and RF showed intermediate performance between GBM and svmRadialSigma in the spatialization of Inselbergs.



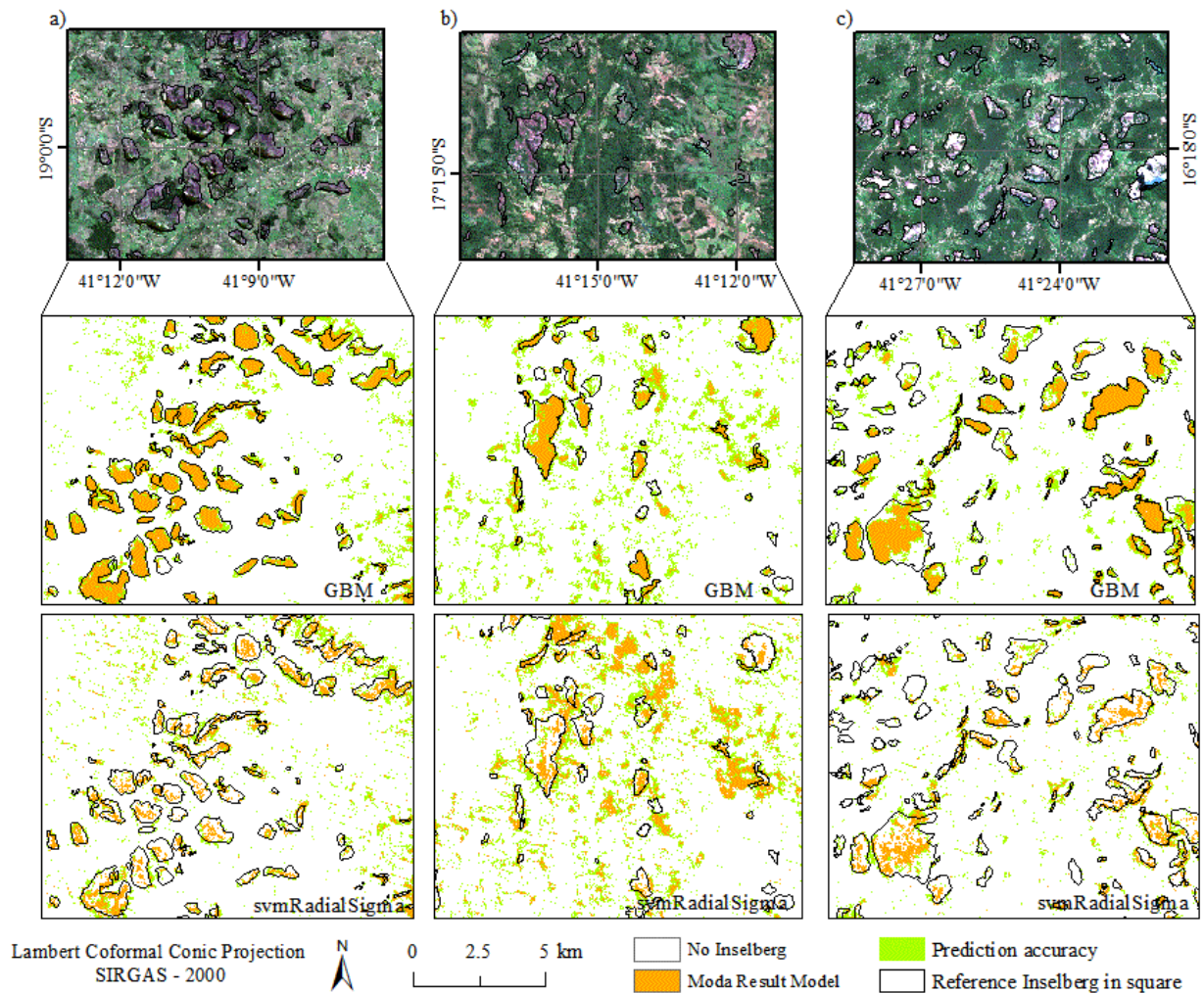
**Fig. 9.** Final map of the prediction accuracy of GBM, svmRadialSigma, C5.0 and RF algorithms for the grid: a) Boa União; b) Novo Oriente - Minas Gerais; and c) Pedra Azul, Brazil.

The final maps generated by the GBM algorithm were more accurate compared to svmRadialSigma. Also, the GBM, presented lower final precision in the mapping result, indicated by greater green spots in the final maps of the prediction accuracy (Figs. 9 and 10). Among the analyzed algorithms, svmRadialSigma showed less dispersion in the mapping of Inselbergs, indicating better precision. For better understanding of the reader, the GBM and svmRadialSigma algorithms were selected to show the final results and their overlap (Fig. 11).



**Fig. 10.** Final prediction accuracy map and final map for the Boa União grid: a) GBM b) svmRadialSigma.

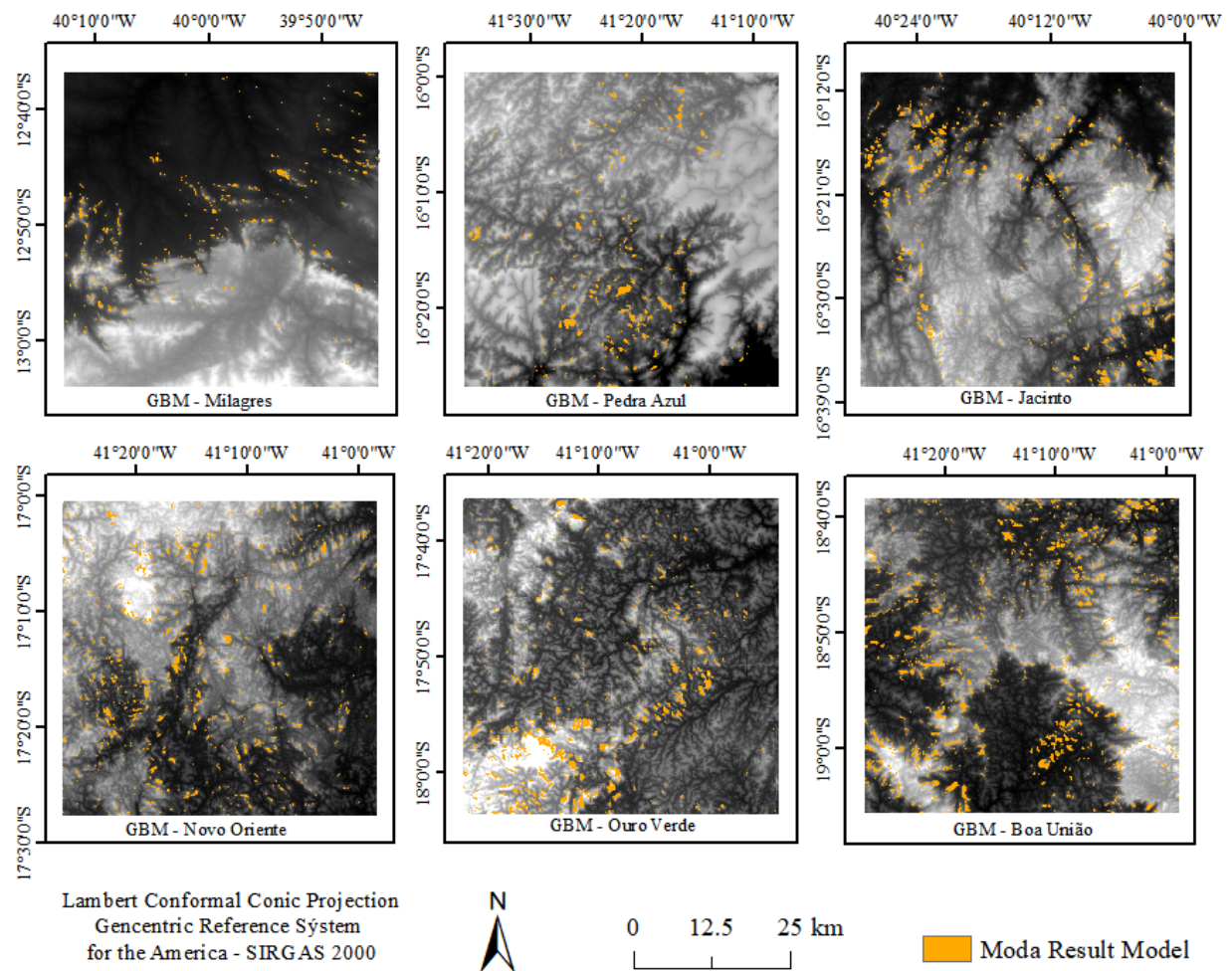




**Fig. 11.** Final map superimposed on the prediction accuracy map for the GBM and SVM algorithms: a) Boa União b) Novo Oriente and c) Pedra Azul, Brazil.

### 3.4. Final maps of the GBM algorithm

Based on the results obtained by statistical analysis and the final maps of modes and Inselbergs, the GBM algorithm was selected to demonstrate the final maps for the entire study area. In total 3612 Inselbergs were mapped by the GBM: Milagres (440), Pedra Azul (626), Jacinto (982), Novo Oriente (343), Ouro Verde (460) and Boa União (761).



**Fig. 12.** Final maps of the Inselbergs mapped by the GBM algorithm for all the studied squares.

### 3.5. Algorithm Assertiveness Index – IAA

The values found for the Algorithm Assertiveness Index are shown in Table 5.

Grid	IAA (%)	
	GBM	svmRadialSigma
Milagres	21,78	33,70
Pedra Azul	23,15	14,92
Jacinto	43,21	37,27
Novo Oriente	49,88	27,75
Ouro Verde	57,02	30,17
Boa União	58,36	38,23

**Table 5**

Assertiveness index for the algorithms analyzed in the prediction of the Inselbergs.

## 4. Discussion

### 4.1. Importance of selected covariates in the spatialization of Inselbergs

The methodological structure of the research allowed to evaluate the machine learning models and the most important covariates for the classification of Inselbergs, as well as to determine the contribution of each covariate in the prediction by algorithm. The covariables terrain\_surface\_texture, terrain\_surface\_convexity, slope\_height and hill selected by the RFE represented important geomorphometric parameters to distinguish the Inselbergs in the studied landscape (Fig. 3).

The covariable terrain\_surface\_terrain captures the ridges and valleys, being able to differentiate the peaks that outline the distribution of the valleys in the MDE - ALOSPALSAR. This covariate is calculated as the nested-means terrain classification, which is defined by both relief (Z) and spacing (X, Y), can be represented by such measures of spatial intricacy as drainage density and changes in sign of slope, aspect or curvature per unit area (Iwahashi and Pike, 2007; Mahmoudzadeh et al., 2020). The slope gradient and surface texture together are fundamental in automatic classification of steep topography, for example, typical Inselbergs landscapes, however, are unsuitable for discriminating between low relief features such as flood plains or river terraces, according to (Iwahashi and Pike, 2007).

According to the results found in this research, terrain surface convexity was important to predict Inselbergs, as shown in Fig. 3. This covariate is calculated as the ratio of the number of cells having positive curvature to the number of all valid cells within the specified search radius (Iwahashi and Pike, 2007). Similar to the texture measure, the convexity is independent of the magnitude of the relief and identifies the shapes, the positive curvatures of the surface and / or the local convexity in the landscape, producing positive values in convex areas upwards, negative values in concave areas and zero in flat slopes (Iwahashi and Pike, 2007). In this sense, it can be inferred that the Inselbergs are classified as positive curvatures on the landscape surface, associated with the plains or can be interpreted as isolated pontoons associated with higher elevations.

The slope\_height covariate selected by the algorithms is defined as the vertical distance between the crest and the tip of a slope (Gökçeoglu and Aksoy, 1996). As an important covariate, the height of the slope limits the size and spatial extent of the Inselbergs in the landscape. This covariate characterizes the height of the slope and assists in parameterizing the unevenness associated with the relief (Qiu et al., 2016). It must be considered that the slope of the slope and / or hill is another influential factor in the modeling processes of the surface and

appears in many automated methods for land classification, although the regional typologies of the land surface are built combining relief, inclination, spacing of resources and other derivatives of height and geographic location (Iwahashi and Pike, 2007).

The covariables `terrain_surface_texture`, `terrain_surface_convexity` and `slope_height` are in line with those suggested by (Iwahashi et al., 2001) and are used as diagnostic variables to differentiate relief classes, topographic targets, as well as to contrast stable slopes and escarpments. According to the selection of the RFE algorithm, they proved to be important in the process of spatialization and mapping of Inselbergs along a climate gradient. (Iwahashi and Kamiya, 1995) suggest these variables to create maps of land units. The concavity and convexity of the terrain surface are widely used to express topographic environments (Hutchinson and Gallant, 1999), so that the texture and local convexity resemble the attributes of the terrain, such as topographic profile or convexity of the plane (Iwahashi and Pike, 2007).

Among the covariables selected a priori by the models using RFE (Gomes et al., 2019), the topographic covariate `hill` proved to be important in every study area to predict and spatialize the Inselbergs. All algorithms selected this covariate to detect the spatial variability of the Inselbergs in the landscape (Fig. 3). In this sense, the morphometric parameters of the terrestrial surface analyze and parameterize the shape of the surface, as well as, identify specific points on the surface and classify the terrain into ridges, slopes and channels (Olaya and Conrad, 2009).

The topographimetric covariates generated in Software R (v 3.5.3), from the interface with the RSAGA package (Brenning, 2008) represents a set of consistent algorithms for spatial analysis of Digital Elevation Models and spatialization of Inselbergs in different landscapes. The morphometric covariables such as slope, plain, curvature, solar radiation and topographic humidity index, also contributed in a different way to detect the Inselbergs in the studied landscape (Fig. 3). We highlight the need to extract primary and secondary morphometric attributes embedded in the Digital Elevation Models - MDE and use them to estimate the presence of Inselbergs.

Of all the spectral covariates selected by the algorithms, the Ferrous Minerals Ratio - FMR index showed the greatest significant importance among the different spectral covariates that predict the Inselbergs (Fig. 3). The FMR is generated from the ratio between the Short Wave Infra-Red and Near Infra-Red bands. Spectral indices are often used to classify land use and land cover to help distinguish land use and land cover classes, especially between vegetation types (Abdi, 2020). According to (Rowan and Mars, 2003), band ratio images

illustrate specific characteristics of spectral absorption of minerals in rocks. It is worth noting that the number of Fe is preferable in almost all granitoids where ferrous and ferric iron analyzes are available, one exception is the classification of suites that show a wide range of  $\text{Fe}^{3+} / \text{Fe}^{2+}$  due to late subsolid oxidation (Frost et al., 2001). It has long been recognized that there are fundamental differences between rock suites that undergo iron enrichment during differentiation, while silica abundance remains low, and those that undergo silica enrichment with only minimal FeO enrichment compared to MgO (Nockolds and Allen, 1956).

Generally, many minerals in the iron oxide group (hematite-  $\text{Fe}_2\text{O}_3$ , goethite -  $\text{FeO}(\text{OH})$ ), have a set of broad spectral absorption characteristics centered mainly in the regions of 500 and 850 at 910 nm (VNIR and SWIR), while magnetite -  $\text{Fe}_3\text{O}_4$  and ilmenite -  $\text{FeTiO}_3$  do not exhibit spectral features in the visible ranges (Formaggio et al., 1996; Mars, 2018). These considerations are more noticeable for hyperspectral images, in this sense, our results indicate that, what really acts in the differentiation of rocks and makes the FMR important in the prediction is water, mainly for multispectral images used in this research. Inselbergs, due to their large proportions of bare rock, reflect more in the shortwave infrared range, where there is less water, more reflectance and more water, less reflectance in the SWIR band. The biotite present in the composition of porphyritic granites, typical of the studied region, may also be acting. According to (Shabani and Lalonde, 2003), the composition of the main element of biotite can serve as a tool, among others, for the classification and characterization of granites, as well as, to understand its petrogenesis, but that none of these methods is infallible or must be used in isolation.

The Ferrous Minerals Ratio index can also be used to predict various chemical and physical characteristics of soils (Shepherd and Walsh, 2002). According to (Mathews et al., 1973), knowledge of the mineralogical composition is essential to evaluate the spectral behavior of soils, due to the various absorption features and the influence on albedo. The observed results indicate that the mineralogical / mineral indexes can be used to separate areas with a high predominance of exposed soil and rocky outcrops in the supervised classification of images by machine learning algorithms. The use of Sentinel-2/MSI bands and the technique of dividing bands were fundamental to separate the Inselbergs and differentiate them in the process of distinguishing classes of land use and land cover.

#### **4.2. Inselbergs precision performance by machine learning algorithms**

From the precision and performance metrics of the models, the results confirm the good performance / performance of the models effectively, above all, the Gradient Boosting Machine (GBM), proposed by (Friedman, 2001a). According to (Dixit et al., 2017), classification accuracy is the most important parameter for any classification algorithm. In this sense, the studied models were classified with substantial agreement - Substantial (0.61-0.80) to almost perfect - Almost Perfect (0.81-1.00) in the classification (Landis and Koch, 1977).

By not overfitting, the results found of accuracy and kappa index of validation corroborate to confirm the good results in the mapping, demonstrated by the small differences found in the training samples and data validation (Table 4 and Fig. 4). The kappa coefficient is more demanding and provides a measure of the difference between the actual data, the reference data and the classifier used to perform the classification versus the likelihood of agreement between the reference data and a random classifier (Adam et al., 2014).

Regarding the sensitivity and specificity measures used to assess the performance of supervised classification algorithms, the results indicate positive observations correctly predicted for Inselbergs, due to the high values of sensitivity and specificity found for the analyzed algorithms. In statistical practice, sensitivity is gained at the expense of specificity and vice versa (Hazra and Gogtay, 2017b). According to Table 4, it is possible to notice higher sensitivity values found for all analyzed algorithms, compared to specificity values. In short, sensitivity represented the percentage of positive observations correctly predicted and specificity, the percentage of negative observations correctly predicted for Inselbergs (Tatem et al., 2003).

#### **4.3. Precision accuracy maps and maps of Inselberg prediction**

The effectiveness of an automated analysis of the topography of DEMs can be assessed in several ways, including map overlay and statistical analysis (Iwahashi and Pike, 2007). The similarity in the values of accuracy, sensitivity and specificity between the algorithms used to map the Inselbergs showed the viability of the methodology for mapping the Inselbergs in tropical landscapes. Considering the focus on mapping and spatializing Inselbergs, GBM presented more accurate final maps in the real geolocation of Inselbergs in the field and Sentinel-2/MSI images. According to the results, the GBM erred more in the prediction attempts, however, in the final result it got more correct. The GBM better delimited the Inselbergs and their respective peaks (highest points of the rocky outcrops), compared to the other analyzed algorithms, mainly in relation to the SVMRadialSigma, which presented the

highest accuracy value and kappa index. It is possible to notice in Fig. 10 that more details about the shape of the Inselbergs are seen in the results of the GBM algorithm, even presenting a lower value in the kappa index and final accuracy compared to the SVM, as shown in Fig. 4.

The results demonstrate the importance of rotating the models several times to obtain results closer to reality and we emphasize that the best model obtained and selected by the result of the validation metric does not always present the most accurate result in the final agreement of the maps, therefore, the importance of the use and final analysis of prediction accuracy maps, satellite images and, when possible, experts in the study topic to determine the best performance algorithm.

The reasons for the mapping and conservation of Inselbergs in the Caatinga and Atlantic Forest Biomes include their high number of geographically restricted and threatened species that function as islands of terrestrial habitat (Bussell and James, 1997). In this sense, geospatial analysis, conservation measures and connection of the forest remnants of the Inselbergs and the surrounding areas are of paramount importance for maintaining these environments threatened by anthropic pressure.

It is worth mentioning that many Inselbergs in the world are threatened by alarming rates of mining, invasion of exotic grasses, water collection, tourism and urbanization, which results in the loss of biodiversity and degradation of their ecosystem services (Paula et al., 2015; Porembski et al., 2016). The mapping and spatialization associated with the high precision and accuracy of the classification obtained in this research provide reliable information on the number and extent of the main Inselbergs in this region of Brazil, as well as, it can be used as a reference to plan ecosystem services, forest restoration and management environmental efficiency for granite Inselbergs rock fields over a climate gradient.

It is worth mentioning that many Inselbergs in the world are threatened by alarming rates of mining, invasion of exotic grasses, water collection, tourism and urbanization, which results in the loss of biodiversity and degradation of their ecosystem services (Paula et al., 2015; Porembski et al., 2016). The spatialization of Inselbergs, associated with high precision and accuracy in the results of the classification obtained in this research, provide reliable approximate information on the number and spatial extent of the main Inselbergs in this region of Brazil under the climatic domain of the Atlantic Forest and Caatinga, as well as, this modeling can be used as a reference for planning ecosystem services, forest / ecological restoration and more assertive environmental management for granite Inselbergs rock fields.

#### **4.4. Algorithm Assertiveness Index - IAA**

According to the results of the spatial assertiveness of the Inselbergs analyzed separately by grid (Table 5), there was a better performance of the GBM algorithm in the grid inserted in the climatic domain Aw, humid climate. The svmRadialSigma algorithm showed better performance in the Milagres/Bahia grid, inserted predominantly in the BSh domain, dry climate. The results found from the IAA corroborate to understand that the machine learning algorithms can present different performances along a climatic gradient, whether in the domain of the Atlantic Forest and/or the Caatinga.

#### **5. Conclusion**

The machine learning algorithms used had adequate performances in mapping inselbergs from Brazil, with a kappa index ranging from 0.80 to 0.87. The Sentinel-2/MSI images classified by the supervised machine learning algorithms GBM, SVM, C5.0 and RF showed consistent and reliable results in the mapping and modeling of Inselbergs and associates landscapes.

The Model Gradient Boosting Machine - GBM presented the second best performance in the classification of the Inselbergs and the best agreement in the final maps. Machine learning algorithms can perform differently according to the climatic domain of the study area, whether it is predominantly wet or dry.

The methodology used in this research has applicability to other global regions with similar of Inselbergs and landscapes.

#### **Declaration of Competing**

Interest on behalf of all authors, the corresponding author states that there is no conflict of interest.

#### **Acknowledgments**

The authors would like to thank the Coordination for the Improvement of Higher Education Personnel (CAPES) for providing the Master's scholarship and financing this research project. We also thank the Graduate Program in Forest Sciences at the Federal University of Viçosa, to the Geoprocessing Laboratory (LABGEO) and to the Research Group



and Geotechnology applied to the Global Environment – GAGEN at the Federal University of Espírito Santos, Brazil.

## References

- Ab'Sáber, A.N., 2003. Os domínios de natureza no Brasil: potencialidades paisagísticas, 2<sup>a</sup>. ed. Ateliê Editorial, São Paulo.
- Ab'Sáber, A.N., 1967. Domínios morfoclimáticos e províncias fitogeográficas do Brasil. *Orientação*, Inst. Geogr. 1, 45–48.
- Abdi, A.M., 2020. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience Remote Sens.* 57, 1–20. <https://doi.org/10.1080/15481603.2019.1650447>
- Adam, E., Mutanga, O., Odindi, J., Abdel-Rahman, E.M., 2014. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *Int. J. Remote Sens.* 35, 3440–3458. <https://doi.org/10.1080/01431161.2014.903435>
- Adhikari, K., Kheir, R.B., Greve, M.B., Bøcher, P.K., Malone, B.P., Minasny, B., McBratney, A.B., Greve, M.H., 2013. High-Resolution 3-D Mapping of Soil Texture in Denmark. *Soil Sci. Soc. Am. J.* 77, 860–876. <https://doi.org/10.2136/sssaj2012.0275>
- Bigarella, J.J., Andrade, G.O. de, 1964. Considerações sobre a estratigrafia dos sedimentos cenozóicos em Pernambuco (Grupo Barreiras). Recife.
- Bonfatti, B.R., Hartemink, A.E., Giasson, E., Tornquist, C.G., Adhikari, K., 2016. Digital mapping of soil carbon in a viticultural region of Southern Brazil. *Geoderma* 261, 204–221. <https://doi.org/10.1016/j.geoderma.2015.07.016>
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brenning, A., 2008. Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models, in: J. Böhner, T.B.& L.M. (Ed.), *SAGA--Seconds Out (Hamburger Beiträge Zur Physischen Geographie Und Landschaftsökologie, Vol. 19)*. pp. 23–32.
- Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C., 2015. Machine

- learning for predicting soil classes in three semi-arid landscapes. *Geoderma* 239–240, 68–83. <https://doi.org/10.1016/j.geoderma.2014.09.019>
- Bussell, J.D., James, S.H., 1997. Rocks as museums of evolutionary processes. *J. R. Soc. West. Aust.* 80, 221–229.
- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* 37, 35–46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B)
- Cortes, C., Vapnik, V., 1995. Support-Vector Networks. *Mach. Learning* 20, 273–297. <https://doi.org/10.1007/BF00994018>
- Dixit, A., Hedge, N., Eswar Reddy, B., 2017. Texture feature based satellite image classification scheme using SVM. *Int. J. Appl. Eng. Res.* 12, 3996–4003.
- Emadi, M., Taghizadeh-Mehrjardi, R., Cherati, A., Danesh, M., Mosavi, A., Scholten, T., 2020. Predicting and mapping of soil organic carbon using machine learning algorithms in northern iran. *Remote Sens.* 12, 2234. <https://doi.org/10.3390/rs12142234>
- ESA, 2019. Sentinel–2 [WWW Document]. URL <https://sentinel.esa.int/web/sentinel/missions/sentinel-2> (accessed 12.1.19).
- Formaggio, A., Epiphano, J., Valeriano, M., Oliveira, J., 1996. Comportamento espectral (450–2.450 nm) de solos Tropicales de Sao Paulo. *Rev. Bras. Cienc. Do Solo* 20, 467–474.
- Friedl, M.A., Brodley, C.E., Strahler, A.H., 1999. Maximizing Land Cover Classification Accuracies Produced by Decision Trees at Continental to Global Scales. *IEEE Trans. Geosci. Remote Sens.* 37, 969–977.
- Friedman, J.H., 2001a. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232.
- Friedman, J.H., 2001b. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232.
- Frost, B.R., Barnes, C.G., Collins, W.J., Arculus, R.J., Ellis, D.J., Frost, C.D., 2001. A geochemical classification for granitic rocks. *J. Petrol.* 42, 2033–2048. <https://doi.org/10.1093/petrology/42.11.2033>
- Galiano, V.F.R., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification.

- ISPRS J. Photogramm. Remote Sens. 67, 93–104.  
<https://doi.org/10.1016/j.isprsjprs.2011.11.002>
- Gašparović, M., Jogun, T., 2018. The effect of fusing Sentinel-2 bands on land-cover classification. *Int. J. Remote Sens.* 39, 822–841.  
<https://doi.org/10.1080/01431161.2017.1392640>
- Ghimire, B., Rogan, J., Galiano, V., Panday, P., Neeti, N., 2012. An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA. *GIScience Remote Sens.* 49, 623–643. <https://doi.org/10.2747/1548-1603.49.5.623>
- Gökçeoglu, C., Aksoy, H., 1996. Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region ( Turkey ) by deterministic stability analyses and image processing techniques. *Eng. Geol.* 44, 147–161. [https://doi.org/10.1016/S0013-7952\(97\)81260-4](https://doi.org/10.1016/S0013-7952(97)81260-4)
- Gomes, L.C., Faria, R.M., de Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Filho, E.I.F., 2019. Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma* 340, 337–350.  
<https://doi.org/10.1016/j.geoderma.2019.01.007>
- Hazra, A., Gogtay, N., 2017a. Biostatistics series module 7: the statistics of diagnostic tests. *Indian J. Dermatol.* 62, 18–24. <https://doi.org/10.4103/0019-5154.198047>
- Hazra, A., Gogtay, N., 2017b. Biostatistics series module 7: the statistics of diagnostic tests. *Indian J. Dermatol.* 62, 18–24. <https://doi.org/10.4103/0019-5154.198047>
- Hechenbichle, K., Schliep, K., 2004. *Weighted k-Nearest-Neighbor Techniques and Ordinal Classification*. Munich.
- Huang, F., Cao, Z., Guo, J., Jiang, S.H., Li, S., Guo, Z., 2020. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *Catena* 191, 104580. <https://doi.org/10.1016/j.catena.2020.104580>
- Hutchinson, M.F., Gallant, J.C., 1999. Representation of terrain, in: Longley, P.A., Maguire, D., Goodchild, M., Rhind, D. (Eds.), *Geographical Information Systems: Principles and Technical Issues*. Wiley, New York, pp. 105–124.
- Iwahashi, J., Kamiya, I., 1995. Landform Classification Using Digital Elevation Model by the Skills of Image Processing mainly using the Digital National Land Information. *Geoinformatics* 6, 97–108. [https://doi.org/10.6010/geoinformatics1990.6.2\\_97](https://doi.org/10.6010/geoinformatics1990.6.2_97)

- Iwahashi, J., Pike, R.J., 2007. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. *Geomorphology* 86, 409–440. <https://doi.org/10.1016/j.geomorph.2006.09.012>
- Iwahashi, J., Watanabe, S., Furuya, T., 2001. Landform analysis of slope movements using DEM in Higashikubiki area, Japan. *Comput. Geosci.* 27, 851–865. [https://doi.org/10.1016/S0098-3004\(00\)00144-8](https://doi.org/10.1016/S0098-3004(00)00144-8)
- Jatobá, L., 1994. *Geomorfologia do semi-árido*. Universidade Federal de Pernambuco, Recife.
- Jaxa, 2020. About ALOS - PALSAR [WWW Document]. URL <https://www.eorc.jaxa.jp/ALOS/en/about/palsar.htm> (accessed 11.10.20).
- Jémisson Mattos dos Santos, A.A.R.S., 2010. Genese Da Superfície Erosiva Em Ambiente Semi-Árido - Milagres/Ba: Considerações Preliminares. *Rev. Geogr. Especial V*, 236–247.
- Jensen, J.R., 2005. *Introductory Digital Image Processing: A Remote Sensing Perspective*, 3<sup>a</sup>. ed. Prentice Hall, Toronto.
- Kuhn, M., 2019. The caret Package [WWW Document]. URL <https://topepo.github.io/caret/index.html> (accessed 4.26.20).
- Kuhn, M., Johnson, K., 2013a. *Applied predictive modeling*. Springer, New York.
- Kuhn, M., Johnson, K., 2013b. *Applied predictive modeling*, 1<sup>a</sup>. ed. Springer, New York.
- Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. <https://doi.org/10.2307/2529310>
- Ließ, M., Schmidt, J., Glaser, B., 2016. Improving the spatial prediction of soil organic carbon stocks in a complex tropical mountain landscape by methodological specifications in machine learning approaches. *PLoS One* 11, e0153673. <https://doi.org/10.1371/journal.pone.0153673>
- Mahmoudzadeh, H., Reza, H., Taghizadeh-mehrjardi, R., Kerry, R., 2020. Geoderma Regional Spatial prediction of soil organic carbon using machine learning techniques in western Iran. *Geoderma Reg.* 21, e00260. <https://doi.org/10.1016/j.geodrs.2020.e00260>
- Maia, R.P., Bétard, F., Bezerra, F.H., 2016. Geomorfologia Dos Maciços De Portalegre E Martins – Ne Do Brasil: Inversão Do Relevo Em Análise. *Rev. Bras. Geomorfol.* 17, 273–

285. <https://doi.org/10.20502/rbg.v17i2.801>
- Maia, R.P., Nascimento, M., 2018. Relevos Graníticos do Nordeste Brasileiro. *Rev. Bras. Geomorfol.* 19, 373–389. <https://doi.org/http://dx.doi.org/10.20502/rbg.v19i2.1295>
- Mars, J.C., 2018. Mineral and lithologic mapping capability of worldview 3 data at Mountain Pass, California, using true- and false-color composite images, band ratios, and logical operator algorithms. *Econ. Geol.* 113, 1587–1601. <https://doi.org/10.5382/econgeo.2018.4604>
- Mas, J.F., Flores, J.J., 2008. The application of artificial neural networks to the analysis of remotely sensed data. *Int. J. Remote Sens.* 29, 617–663. <https://doi.org/10.1080/01431160701352154>
- Mathews, H.L., Cunningham, R.L., Petersen, G.W., 1973. Spectral Reflectance of Selected Pennsylvania Soils. *Soil Sci. Soc. Am. J.* 37, 421–424. <https://doi.org/10.2136/sssaj1973.03615995003700030031x>
- Maxwell, A.E., Warner, T.A., Fang, F., Maxwell, A.E., Warner, T.A., 2018. Implementation of machine-learning classification in remote sensing: an applied review. *Int. J. Remote Sens.* 39, 2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>
- Morales, G., Ramírez, A., Telles, J., 2018. Cloud Segmentation in High-Resolution Multispectral Satellite Imagery Using Deep Learning, in: *ArXiv*. pp. 280–288. <https://doi.org/10.1007/978-3-030-01424-7>
- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* 66, 247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- Muñoz-Romero, S., Gorostiaga, A., Soguero-Ruiz, C., Mora-Jiménez, I., Rojo-Álvarez, J.L., 2020. Informative variable identifier: Expanding interpretability in feature selection. *Pattern Recognit.* 98, 107077. <https://doi.org/10.1016/j.patcog.2019.107077>
- Neto, P.L. de O.C., 2002. *Estatística*, 2ª. ed. Edgard Blücher, São Paulo.
- Nockolds, S.R., Allen, R., 1956. The geochemistry of some igneous rock series—III. *Geochim. Cosmochim. Acta* 9, 34–77. [https://doi.org/10.1016/0016-7037\(56\)90056-4](https://doi.org/10.1016/0016-7037(56)90056-4)
- Olaya, V., Conrad, O., 2009. Geomorphometry in SAGA, in: Tomislav Hengl, H.I.R. (Ed.),

- Geomorphometry: Concepts, Software, Applications. Elsevier Science, Amsterdam, pp. 293–308. [https://doi.org/10.1016/S0166-2481\(08\)00012-3](https://doi.org/10.1016/S0166-2481(08)00012-3).
- Oliveira-Filho, A.T., Tameirão-Neto, E., Carvalho, W.A.C., Werneck, M., Brina, A.E., Vidal, C. V., Rezende, S.C., Pereira, J.A.A., 2005. Análise florística do compartimento arbóreo de áreas de floresta atlântica sensu lato na região das Bacias do Leste (Bahia, Minas Gerais, Espírito Santo e Rio de Janeiro). *Rodriguésia* 56, 185–235.
- Paula, L.F.A. de, Forzza, R.C., Neri, A. V., Bueno, M.L., Porembski, S., 2016. Sugar Loaf Land in south-eastern Brazil: a centre of diversity for mat-forming bromeliads on inselbergs. *Bot. J. Linn. Soc.* 181, 459–476. <https://doi.org/10.1111/boj.12383>
- Paula, L.F.A. de, Mota, N.F.O., Viana, P.L., Stehmann, J.R., 2017. Floristic and ecological characterization of habitat types on an inselberg in Minas Gerais, southeastern Brazil. *Acta Bot. Brasilica* 31, 199–211. <https://doi.org/10.1590/0102-33062016abb0409>
- Paula, L.F.A. de, Negreiros, D., Azevedo, L.O., Fernandes, R.L., Stehmann, J.R., Silveira, F.A.O., 2015. Functional ecology as a missing link for conservation of a resource-limited flora in the Atlantic forest. *Biodivers. Conserv.* 24, 2239–2253. <https://doi.org/10.1007/s10531-015-0904-x>
- Porembski, S., Silveira, F.A.O., Fiedler, P.L., Watve, A., Rabarimanarivo, M., Kouame, F., Hopper, S.D., 2016. Worldwide destruction of inselbergs and related rock outcrops threatens a unique ecosystem. *Biodivers. Conserv.* 25, 2827–2830. <https://doi.org/10.1007/s10531-016-1171-1>
- Qiu, H., Regmi, A.D., Cui, P., Cao, M., Lee, J., Zhu, X., 2016. Size distribution of loess slides in relation to local slope height within different slope morphologies. *Catena* 145, 155–163. <https://doi.org/10.1016/j.catena.2016.06.005>
- Quinlan, J.R., 2004. Data mining tools See 5 and C5. 0. [WWW Document]. URL <https://www.rulequest.com/see5-info.html> (accessed 4.26.20).
- Reis, B.P., Martins, S.V., Fernandes Filho, E.I., Sarcinelli, T.S., Gleriani, J.M., Leite, H.G., Halassy, M., 2019. Forest restoration monitoring through digital processing of high resolution images. *Ecol. Eng.* 127, 178–186. <https://doi.org/10.1016/j.ecoleng.2018.11.022>
- Rios, I.Q., 2017. Dinâmica Geomorfológica do Semiárido de Lagedo Alto - Bahia: Estudo de

três de Inselbergs representativos da região de Milagres. Universidade Estadual de Feira de Santana.

Ripley, B.D., 1996. Pattern recognition and neural networks, 1<sup>a</sup>. ed. Cambridge University Press, New York.

Ripley, B. D., 1996. Pattern recognition and neural networks, 1<sup>a</sup>. ed. Cambridge University Press, New York.

Rodrigues, W.F., Maia, R.P., Gomes, D.D.M., 2019. Condicionamento Morfoestrutural Do Inselberg Pedra Da Andorinha, Sertão Norte Do Ceará, Brasil. *Rev. Bras. Geomorfol.* 20, 861–876. <https://doi.org/10.20502/rbg.v20i4.1521>

Rowan, L.C., Mars, J.C., 2003. Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. *Remote Sens. Environ.* 84, 350–366. [https://doi.org/10.1016/S0034-4257\(02\)00127-X](https://doi.org/10.1016/S0034-4257(02)00127-X)

Schaefer, C.E.G.R., 2012. Bases físicas da paisagem brasileira: estrutura geológica, relevo e solos, in: Araújo, P. A., Alves, B.J.R. (Ed.), *Tópicos Em Ciência Do Solo*. Sociedade Brasileira de Ciência do Solo, Viçosa, pp. 1–69.

Schliep, K., Hechenbichle, K., Lizee, A., 2016. kkn: Weighted k-nearest neighbors.

Seasholtz, M.B., Kowalski, B., 1993. The parsimony principle applied to multivariate calibration. *Anal. Chim. Acta* 277, 165–177. [https://doi.org/10.1016/0003-2670\(93\)80430-S](https://doi.org/10.1016/0003-2670(93)80430-S)

Sena, N.C., Veloso, G.V., Fernandes-Filho, E.I., Francelino, M.R., Schaefer, C.E.G.R., 2020. Analysis of terrain attributes in different spatial resolutions for digital soil mapping application in southeastern Brazil. *Geoderma Reg.* 21, e00268. <https://doi.org/10.1016/j.geodrs.2020.e00268>

Shabani, A.T., Lalonde, E., 2003. Composition of Biotite from Granitic Rocks of the Canadian Appalachian Orogen: a potencial tectonomagmatic indicator? *Can. Mineral.* 41, 1381–1396. <https://doi.org/10.2113/gscanmin.41.6.1381>

Shendryk, Y., Rist, Y., Ticehurst, C., Thorburn, P., 2019. Deep learning for multi-modal classification of cloud, shadow and land cover scenes in PlanetScope and Sentinel-2 imagery. *ISPRS J. Photogramm. Remote Sens.* 157, 124–136. <https://doi.org/10.1016/j.isprsjprs.2019.08.018>

- Shepherd, K.D., Walsh, M.G., 2002. Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Sci. Soc. Am. J.* 66, 988–998. <https://doi.org/10.2136/sssaj2002.9880>
- Shi, D., Yang, X., 2016. An assessment of algorithmic parameters affecting image classification accuracy by random forests. *Photogramm. Eng. Remote Sensing* 82, 407–417. <https://doi.org/10.14358/PERS.82.6.407>
- Souza, C.M.P. de, Thomazini, A., Schaefer, C.E.G.R., Veloso, G.V., Moreira, G.M., Fernandes Filho, E.I., 2018. Multivariate Analysis and Machine Learning in Properties of Ultisols (Argissolos) of Brazilian Amazon. *Rev. Bras. Ciência do Solo* 42, e170419. <https://doi.org/10.1590/18069657rbc20170419>
- Tatem, A.J., Baylis, M., Mellor, P.S., Purse, B. V., Capela, R., Pena, I., Rogers, D.J., 2003. Prediction of bluetongue vector distribution in Europe and north Africa using satellite imagery. *Vet. Microbiol.* 97, 13–29. <https://doi.org/10.1016/j.vetmic.2003.08.009>
- Voyant, C., Notton, G., Kalogirou, S., Nivet, M.L., Paoli, C., Motte, F., Fouilloy, A., 2017. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* 105, 569–582. <https://doi.org/10.1016/j.renene.2016.12.095>
- Yu, Le, Liang, L., Wang, J., Zhao, Y., Cheng, Q., Hu, L., Liu, S., Yu, Liang, Wang, X., Zhu, P., Li, Xueyan, Xu, Y., Li, C., Fu, W., Li, Xuecao, Li, W., Liu, C., Cong, N., Zhang, H., Sun, F., Bi, X., Xin, Q., Li, D., Yan, D., Zhu, Z., Goodchild, M.F., Gong, P., 2014. Meta-discoveries from a synthesis of satellite-based land-cover mapping research. *Int. J. Remote Sens.* 35, 4573–4588. <https://doi.org/10.1080/01431161.2014.930206>
- Zeraatpisheh, M., Ayoubi, S., Jafari, A., Finke, P., 2017. Comparing the efficiency of digital and conventional soil mapping to predict soil types in a semi-arid region in Iran. *Geomorphology* 285, 186–204. <https://doi.org/10.1016/j.geomorph.2017.02.015>
- Zhang, L., Chen, H., Tao, X., Cai, H., Liu, J., Ouyang, Y., Peng, Q., Du, Y., 2020. Machine learning reveals the importance of the formation enthalpy and atom-size difference in forming phases of high entropy alloys. *Mater. Des.* 193, 108835. <https://doi.org/10.1016/j.matdes.2020.108835>



Zhou, J., Li, E., Yang, S., Wang, M., Shi, X., Yao, S., Mitri, H.S., 2019. Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories. *Saf. Sci.* 118, 505–518. <https://doi.org/10.1016/j.ssci.2019.05.046>