RA'EGA
O ESPAÇO GEOGRÁFICO EM ANÁLISE
https://revistas.ufpr.br/raega

ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

Analysis and classification of the cabruca agroforestry system with machine learning in the Atlantic Forest Biome – Brazil

Análise e classificação do sistema agroflorestal cabruca com aprendizagem de máquina no Bioma Mata Atlântica – Brasil

Vinícius de Amorim Silva [*], **Hercules da Silva Carvalho** [**], **Paulo Sérgio Vila Nova Souza** [***]

[*] Center for Training in Technosciences and Innovation, Federal University of Southern Bahia - UFSB, vinicius@ufsb.edu.br
[**] Center for Training in Technosciences and Innovation, Federal University of Southern Bahia – UFSB, hercules.carvalho@cja.ufsb.edu.br
[***] Center for Training in Agroforestry Sciences, Federal University of Southern Bahia - UFSB, paulosvn@gfe.ufsb.edu.br

http://dx.doi.org/10.5380/raega.v64i1.101334

Abstract

The absence of labeling and spectral similarity of cabruca systems in the Atlantic Forest biome for the Almada, Cachoeira, and Una River basins in the databases hinders the proper classification of these areas. Thus, this study aims to analyze the classification of cabruca areas in the mentioned basins through the use of images of different resolutions and machine learning. The methodology includes: determining land cover and land use classes, acquiring satellite images, processing the images with algorithms in Python, and evaluating supervised classification. The results show that cropping algorithms are effective in processing satellite images, preserving essential geographic information. Furthermore, the Neural Network model achieved an accuracy of 92%, regardless of the diversity of spatial resolutions. However, it is recommended to use the classification results of cabruca areas as estimates, due to the difficulty of the algorithms in distinguishing these areas from Dense Ombrophilous Forest.

Keywords:
Artificial Intelligence, Geoprocessing, Shaded Cocoa, Biodiversity Conservation.

Resumo

A ausência de rotulagem e similaridade espectrais dos sistemas cabruca no bioma Mata Atlântica para as bacias hidrográficas dos rios Almada, Cachoeira e Una nos bancos de dados dificulta a classificação adequada dessas áreas. Assim, este trabalho objetiva analisar a classificação das áreas de cabruca nas referidas bacias por meio do uso de imagens de diferentes resoluções e aprendizado de máquina. A metodologia inclui: a determinação das classes de cobertura e uso da terra, a obtenção de imagens de satélite, o processamento das imagens com algoritmos em Python

ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

e a avaliação da classificação supervisionada. Os resultados mostram que os algoritmos de recorte são eficazes no processamento das imagens de satélite, preservando informações geográficas essenciais. Além disso, o modelo de Redes Neurais alcançou uma acurácia de 92%, independe da diversidade de resoluções espaciais. No entanto, recomenda-se usar os resultados da classificação das áreas de cabruca como estimativas, devido à dificuldade dos algoritmos em distinguir essas áreas da Floresta Ombrófila Densa.

Palavras-chave:
Inteligência Artificial, Geoprocessamento, Cacau sombreado, Conservação da biodiversidade.

I. INTRODUCTION

The classification of cabruca areas is essential for assessing the impact of agricultural practices on biodiversity conservation and the maintenance of ecosystem services. Cabruca, an agroforestry system in the southern region of Bahia, represents a multifunctional strategy for cacao production and conservation of forest remnants (Valadares, 2016; Embrapa, 2021; Xavier; Nascimento Jr; Chiapetti, 2021a). The practice of cabruca, which involves the cultivation of cacao under the shade of dense ombrophilous vegetation, stands out, in areas where forests are scarce and fragmented, as a sustainable approach to agriculture and environmental conservation of native species (Xavier; Nascimento Jr; Chiapetti, 2021b).

In this way, geoprocessing becomes a crucial tool for analyzing cabruca and Dense Ombrophilous Forest (DOF) areas, providing *insights* into the carbon sequestration capacity and conservation levels of each. However, the classification levels of each. However, the classification of these areas faces difficulties in finding characteristics that differentiate them, as a consequence of the spectral similarities between the vegetation and the limitations of remote sensing techniques, which hinder the accurate interpretation of spectral indices in satellite images (Valadares, 2016; Lisboa, 2022).

The difficulty in distinguishing between cabruca and DOF is identified in the works of Souza et al. (2011) and Teixeira et al. (2013), in which both classes were categorized as one, and in Fonseca et al. (2023), where its use as an estimate of "shaded cacao" state is recommended. The complexity of this distinction forces researchers to aggregate the aforementioned classes. However, this approach compromises the precision of analyses and interventions aimed at conserving and managing these environments. There are machine learning and neural network techniques, which present advantages in terms of accuracy in the supervised classification of satellite images (Ouchra; Belangour; Erraissi, 2023; Bhosale; Patankar, 2023).

73
ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

Investments in remote sensing technologies and machine learning methods are essential for the development of structured approaches with a theoretical bias, such as the encoding of different spectral resolutions into homogeneous mathematical structures, ensuring that all elements of the multidimensional array are of the same data type. In this sense, it is hypothesized that the diversity of resolution may provide distinct spectral characteristics, contributing to a more detailed analysis. One of the main challenges is the low availability of dedicated databases that include cabruca and FOD classes, configured to contain images of different resolutions.

In this context, when specifically assessing areas of the Atlantic Forest biome, particularly in the Almada River Basin (BHRA), the Cachoeira River Basin (BHRC), and the Una River Basin (BHRU), it is identified that the natural systems of the cocoa-growing region, such as secondary forests, are still unlabeled, which hinders the classification of these elements. Thus, the aim of this study is to analyze the classification of cabruca and FOD areas in the aforementioned basins, associating them with the use of machine learning and a database with different spectral resolutions.

II. Materials and Methods

For this work, a structured approach in logical steps is adopted. Starting with the determination of land cover and use classes, as well as collection areas. Following this, satellite images with different resolutions are obtained. Then, image processing is carried out. Finally, the database is evaluated, with the classification of the study areas using a supervised classification model (Figure 1).

For the execution of the steps, the following research platforms are used: a) Google Earth Engine, Google Colaboratory; and, b) software QGIS (version 3.22.16), under the General Public License (GNU).

O ESPAÇO GEOGRÁFICO EM ANÁLISE
https://revistas.ufpr.br/raega
ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

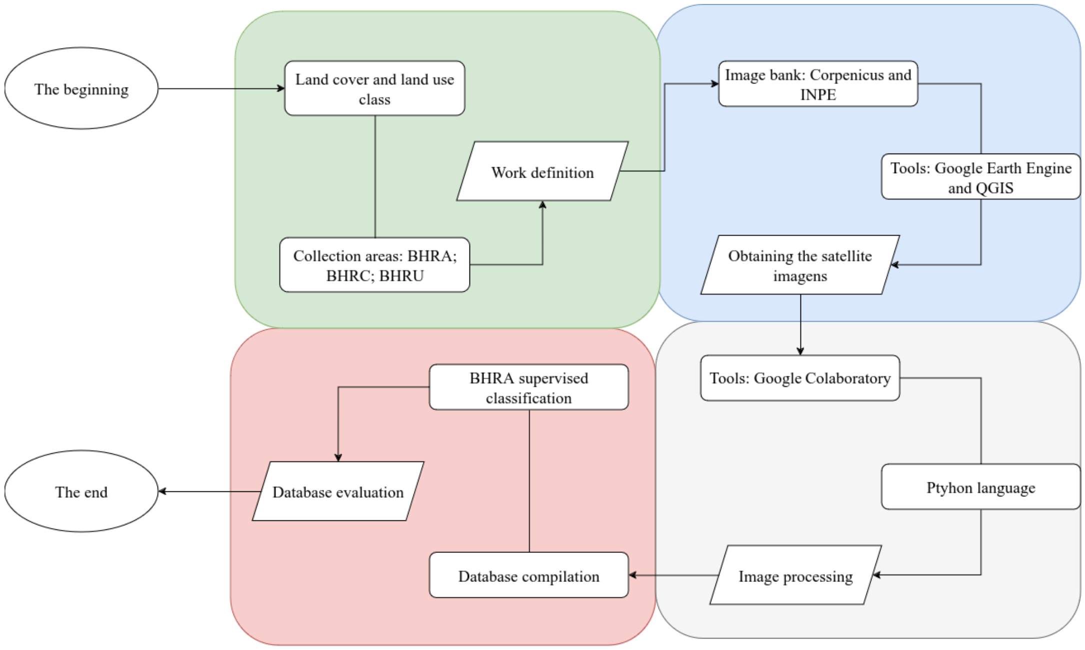

Figure 1 – Methodology flowchart. Source: The authors 2024.

III. Work Definition

The dynamics of cabruca areas present different configurations, which adapt both to the landowner and to local environmental conditions. Among these configurations, the following stand out: cabrucas with Atlantic Forest remnants, characterized by a tree canopy that provides shade for cocoa cultivation; thinned cabrucas, with lower density due to the reduced presence of large trees; and cabrucas integrated into consortium systems, where cocoa is cultivated alongside other species, such as banana (Sambuichi, 2006; Xavier; Nascimento Jr; Chiapetti, 2021a; Xavier; Nascimento Jr; Chiapetti, 2021b).

In the analysis of the classification of cabruca and FOD areas, it is essential to understand the relationship between these configurations and their reflectance. It is expected that dense cabruca areas will present greater spectral similarity with FOD areas, while thinned or consortium cabrucas will exhibit spectral characteristics with lower similarity. For this purpose, a dedicated geographic database was used, containing land cover and land use images of the study area, as well as field recognition of the different configurations present in each image. The information is extracted from the BHRA, BHRC, and BHRU, being labeled according to their respective classes, following the standard established by the Brazilian Institute of Geography and Statistics (IBGE).

https://revistas.ufpr.br/raega
ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

IV. STUDY AREA

The study area is located in the southern state of Bahia, encompassing units within the Atlantic Forest biome (Figure 2). BHRA includes the municipalities of Almadina, Coaraci, Itajuípe, Uruçuca, Barro Preto, and Ilhéus, with over 247,000 inhabitants. BHRC covers the municipalities of Barro Preto, Buerarema, Firmino Alves, Floresta Azul, Ibicaraí, Ilhéus, Itabuna, Itaju da Colônia, Itapé, Itororó, Jussari, Santa Cruz da Vitória, and São José da Vitória, and a population of over 470,000 inhabitants. Meanwhile, BHRU comprises the municipalities of Arataca, Santa Luzia, São José da Vitória, and Una, with approximately 47,000 inhabitants (IBGE, 2022).

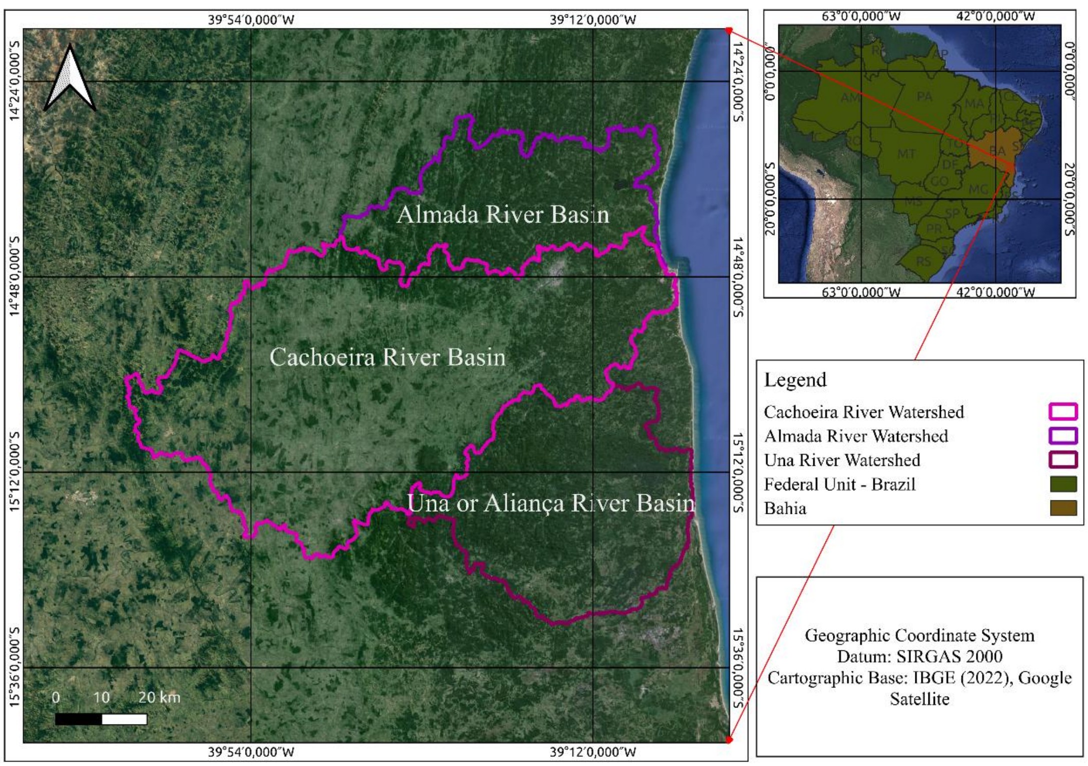

Figure 2 – Map of the location the limits of the basins. Source: IBGE, 2022. Google Satellite, 2023. Elaboration: The authors, 2024.

V. OBTAINING SATELLITE IMAGES

For the construction of the *dataset*, images from CBERS04A and Sentinel-2 satellites with resolutions of 2, 8, 10, 20, 55, and 60 meters, obtained from INPE and Copernicus, are used. The selection is based on the hypothesis that the diversity of resolution can offer different spectral characteristics.

RA'EGA
O ESPAÇO GEOGRÁFICO EM ANÁLISE
https://revistas.ufpr.br/raega
ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

For image acquisition the percentage of cloud cover and concentration points are evaluated, ensuring no overlap with urban infrastructures, DOF points, areas validated in the yield, and water bodies. The 2-meter resolution images were generated through the processing of the panchromatic (PAN) band using the pansharpening algorithm, combined with color information from the 8-meter images obtained by merging the red, green, and blue bands. Similarly, the 20- and 60-meter images were obtained together with the 10-meter Sentinel-2 image (Table 1).

Table 1 – Information about the images used in the work.

Resolution	Optical sensor	Wavelength					Satellite	Date of acquisition
		B Band	G Band	R Band	NIR Band	PAN Band		
8 meters	WPM	0,45-0,52µm	0,52-0,59µm	0,63-0,69µm	0,77-0,89µm	0,45-0,90 µm	CBERS04A	06-22-2023
8 meters	WPM	0,45-0,52µm	0,52-0,59µm	0,63-0,69µm	0,77-0,89µm	0,45-0,90 µm	CBERS04A	05-13-2021
8 meters	WPM	0,45-0,52µm	0,52-0,59µm	0,63-0,69µm	0,77-0,89µm	0,45-0,90 µm	CBERS04A	07-02-2020
10 meters	MSI	0,49µm	0,56µm	0,66µm	0,83µm	-	Sentinel-2	14-05-2024
20 meters	MSI	0,49µm	0,56µm	0,66µm	0,83µm	-	Sentinel-2	14-05-2024
60 meters	MSI	0,49µm	0,56µm	0,66µm	0,83µm	-	Sentinel-2	14-05-2024
55 meters	WFI	0,45-0,52µm	0,52-0,59µm	0,63-0,69µm	0,77-0,89µm	-	CBERS04A	13-07-2023

Source: INPE, 2024. ESA, 2024. Elaboration: The authors, 2024.

In addition, to ensure the reliability of the geographic data, the coordinates of the points where cabruca areas are located were collected during field data validation using the Timestamp Camera Free and the eTrex 10 handheld GPS. Subsequently, they are verified with Google Earth Pro, such as the coordinate points: -14° 47' 43.536''S -39° 10' 21.24''W or -14° 37' 42.072''S -39° 32' 21.056''W.

VI. IMAGE PROCESSING

The cabruca and FOD areas identified during field visits are subsequently mapped using QGIS tools. Then, to reduce the likelihood of human error, an algorithm was developed in Python to perform the reading and automatic clipping of the corresponding images, adjusting them for use in machine learning model training. This algorithm allows the manual definition of image dimensions, providing flexibility and control over the processing. In addition, it efficiently organizes the dataset by storing the clipped images in separate folders according to their respective classes.

It is noteworthy that, in addition to the main classes, the database also includes classes of urban infrastructure, pasture, and water bodies present in the basins. The delimitation of these areas is based on

ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

field visits, prior local knowledge, and visual information from the MapBiomas Project – Collection 9. Furthermore, the "cloud" class was incorporated due to the high frequency of this phenomenon in satellite images, influenced by the humid tropical climate of southern Bahia and by the moisture coming from the Atlantic Ocean.

VII. DATABASE EVALUATION

After compiling the data, the pixel information from the images is accessed through the reading of spectral bands and stored in a NumPy variable, a multidimensional array widely used in Python for the efficient handling of numerical data. This structure is essential for image processing and machine learning tasks, as it enables optimized mathematical and logical operations, such as the manipulation of large volumes of spatial data. From this point, machine learning algorithms are applied to evaluate the effectiveness of the constructed database. Models such as K-nearest neighbors (KNN), which base their predictions on the similarity between neighboring data points, and Neural Networks, capable of identifying more complex and subtle patterns in the data, are used (Alba et al., 2022; Fagundes; Júnior, 2022).

To ensure the best performance and optimize classification accuracy, the models' hyperparameters were tuned following the guidelines of the scikit-learn library, which provides robust tools for machine learning model optimization. Among the methods used, cross-validation stands out, ensuring a robust evaluation of the model, as well as hyperparameter tuning through Grid Search, which allows the identification of the optimal parameter combination to maximize accuracy in satellite image pattern recognition while avoiding overfitting (Homer et al., 2019; Smith; Jones, 2021).

The dataset was split into three parts: 70% for training, 15% for validation, and 15% for testing, ensuring adequate class balance. Model performance was evaluated using metrics such as accuracy, precision, recall, and F1-score, enabling a detailed comparison of the effectiveness of the models in classifying satellite images.

These algorithms are employed to identify pixel patterns in images with 8- and 10-meter resolution. In image classification, two distinct conditions are assessed: Condition A, corresponding to the region near the municipality of Itabuna-BA, characterized mainly by the predominance of FOD, urban infrastructure, and pasture classes, with limited occurrence of cabruca areas; and Condition B, referring to the area near CEPLAC-BA, composed of FOD, cabruca, and urban infrastructure.

ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

Furthermore, for a second round of evaluations, considering the correction of identified structural variables and questioning the hypothesis of resolution diversity, a database composed of 2-meter resolution images obtained from a single satellite was used.

Performance is evaluated using assessment metrics (accuracy, precision, recall, and F1-score) and the confusion matrix, contributing to determine the role of the image database and its influence on machine learning classification models. Finally, the manual estimation of the area of each class was carried out by overlapping data obtained from field visits and MapBiomas Collection 9, due to the difficulty of achieving a reliable classification for the study area. In this approach, it was considered that 0.05% of the FOD class corresponds to cabruca areas, based on field data validation.

VIII. RESULTS AND DISCUSSION

In the analysis of the classification results obtained by the learning models, it is observed that the algorithms show difficulties in accurately distinguishing cabruca areas. For the KNN model, its performance on the test data reached an overall accuracy of 82%, indicating its ability to provide estimates close to the reference values.

However, there is a lower performance, below 50%, for the cabruca class when compared to the others, where the precision and the disparity reinforced in the F1-Score stand out, revealing an imbalance in the classification of the aforementioned class. On the other hand, the DOF, pasture and water body classes demonstrate the effectiveness of the database with accuracy above 80% (Table 2).

Table 2 – Results of the evaluation of the KNN model with the images from the dedicated database.

Class	Accuracy / %	Recall / %	F1-Score / %
Cabruca	48	36	41
Dense Ombrophilous Forest	86	84	85
Urban Infrastructure	60	78	68
Pasture	87	84	86
Walter body	89	90	90
Cloud	72	51	60

Authorship: the authors, 2024.

In Table 2, the model's performance with respect to the recall of the cabruca and cloud classes is presented, which identified 36% and 51% of the relevant instances, respectively. The model's low retrieval capacity compromises classification quality, especially for the cabruca class, possibly due to overlap with other classes. The difficulty in extracting adequate spectral features from the images further limits the classification process, particularly when spectral similarity occurs among different land cover classes, making it impossible

ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

to separate them based solely on reflectance values and requiring additional criteria, beyond spectral ones, such as photointerpretation (Moreira; Adami; Rudorff, 2004; Sena Souza et al., 2016).

According to Tan et al. (2022), one factor associated with this issue is the absence of sensors capable of reducing the effects of surface overlap. In the classification of images with 8- and 10-meter resolution, it was observed that, for Condition A—where cabruca occurrence is limited—some inconsistencies emerged, as validated by field data. In the 8-meter image, discrepancies were found in the classification of the water body class, which was mistakenly interpreted as urban infrastructure, while portions of pasture areas were misclassified as water bodies.

The results highlight the importance of the class separability index; a factor related to classification accuracy. According to Rocha et al. (2012) and Das and Pandeys (2019), low accuracy in the classification of closely associated mixed pixels is due to the smaller spectral difference compared to correctly classified categories. In heterogeneous environments such as urban areas, the presence of a complex mixture of spectral responses within a single pixel (mixels) result in a spectral signature that combines the basic land-use and land-cover materials. This spectral mixing hinders the identification of classes through pixel-by-pixel analytical techniques. Therefore, when addressing misclassifications, the use of morphological dilation algorithms is considered to mitigate mixed-pixel problems, along with techniques for shadow detection and removal (Rocha et al., 2012; Liang; Liang; Sun, 2023).

When the spectral response of these mixed pixels is very similar either among themselves or to other classes (pure or mixed), the spectral separability of the categories decreases, leading to classification confusion. For this reason, the model's performance in estimating urban infrastructure and pasture areas was evaluated. However, cabruca points close to the FOD class stand out, reinforcing the lack of distinct pattern identification and, consequently, the overlap between them.

For the 10-meter image, higher accuracy was observed in estimating urban infrastructure and pasture areas, with roads and pastures being identified with greater quality compared to the 8-meter images (Figure 3).

RA'EGA
O ESPAÇO GEOGRÁFICO EM ANÁLISE
https://revistas.ufpr.br/raega
ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

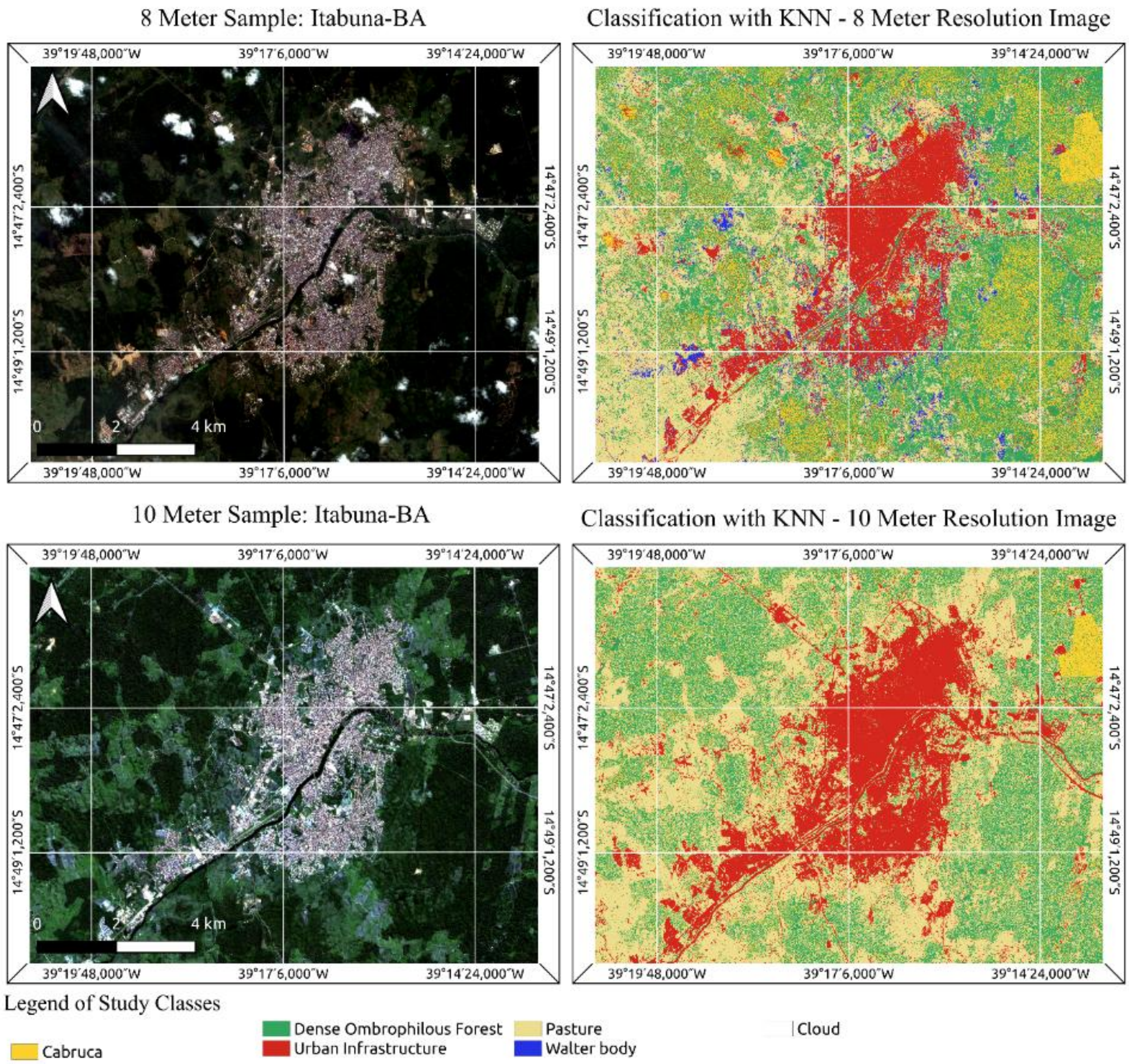

Figure 3 – Classification of land cover and use using the KNN model. Source: Elaborated by the authors, 2024.

For condition B, the ability to classify the cabruca areas present in the image is analyzed, reaching the polygon defined manually based on field data validation. This indicates that the model retrieved preserved information from the image database, particularly in the 10-meter resolution image, which shows a predominance of the FOD class around the polygon (Figure 4).

O ESPAÇO GEOGRÁFICO EM ANÁLISE
https://revistas.ufpr.br/raega
ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

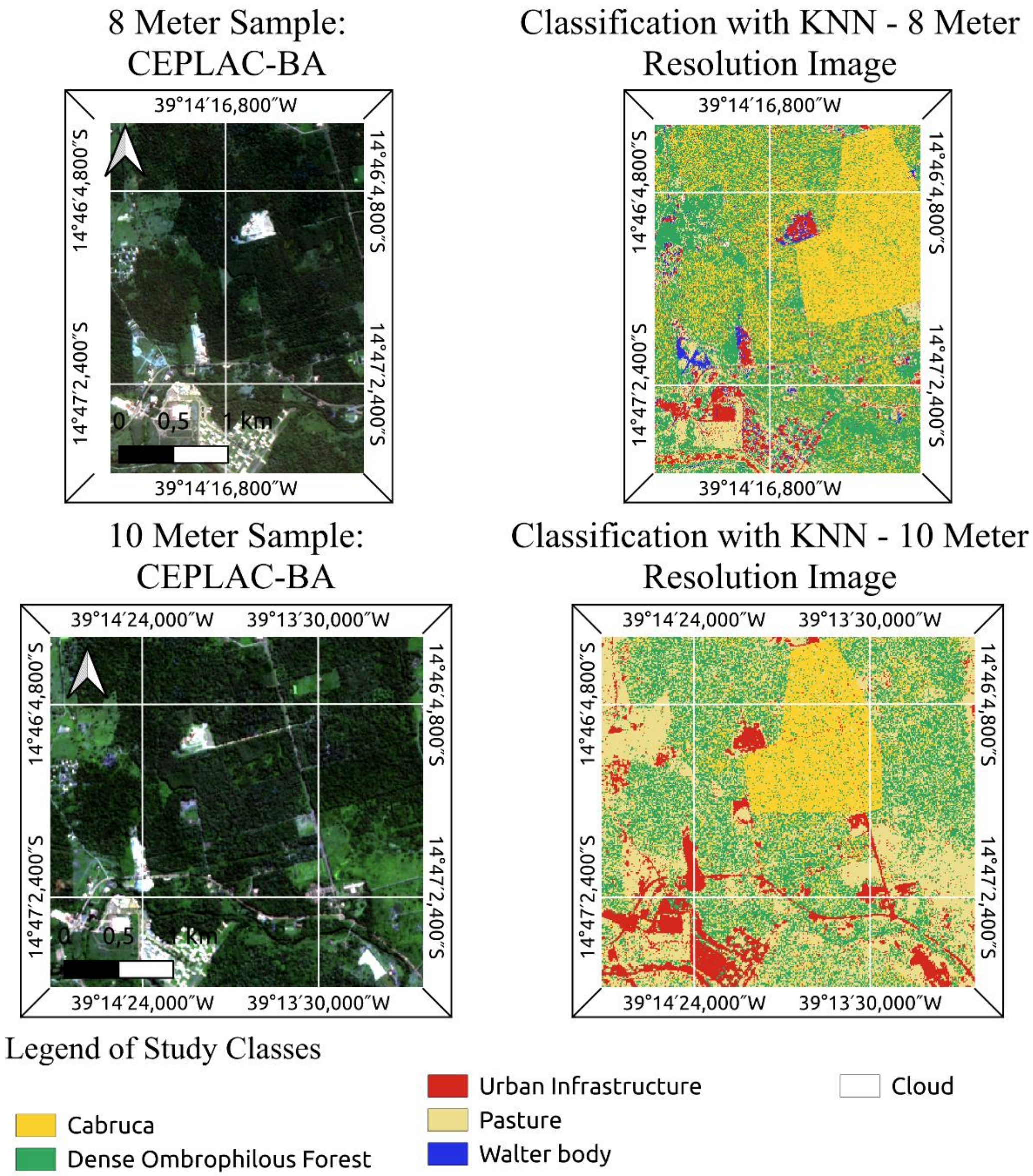

Figure 4 – Classification of land cover and use of the area near CEPLAC-BA using the KNN model. Source: Elaborated by the authors, 2024.

For the Neural Network model, the results demonstrate that the algorithm is effective in classifying the different classes, with consistent and balanced performance across all categories. This indicates that the Neural Network, compared to the KNN method, is an advantageous choice for classifying cabruca and DOF areas. However, it is necessary to evaluate the neural network structure; increasing the number of hidden layers may not significantly improve accuracy, resulting in computational costs without substantial gains (Moraitis et al., 2021).

RA'EGA
O ESPAÇO GEOGRÁFICO EM ANÁLISE
https://revistas.ufpr.br/raega
ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

In the test data, accuracy reached 92%, with a precision of 84% for the cabruca class, although recall suggests a loss of areas in this category, unlike the DOF class, which shows a balanced F1-Score (Table 3).

Table 3 – Results of the Neural Network model evaluation using the database.

Class	Accuracy / %	Recall / %	F1-Score / %
Cabruca	84	65	73
Dense ombrophilous forest	94	92	93
Urban Infrastructure	69	93	79
Pasture	96	91	94
River, lakes and oceans	97	95	96
Cloud	95	86	90

Source: Elaborated by the authors, 2024.

In the classification of Condition A, it is evaluated that the 8-meter and 10-meter images exhibit similar characteristics to the KNN model, with inconsistent results in the classification of the water body and cloud classes, especially in the 10-meter images. It is identified that the images selected to compose the database did not meet the necessary requirement to avoid overlaps with other categories.

It is noteworthy to emphasize the importance of prior evaluation of the images before submitting them to the clipping algorithm. It becomes essential to establish criteria related to the level of roughness or the presence of elements from other classes in the images. Patel, Chatterjee, and Gorai (2017) and Nazmuzzaman and Anwar (2019) highlight factors such as texture, intensity, lighting conditions, and shadows. For example, the Cachoeira River, near the municipality of Itabuna-BA, presents cloud overlap and proximity to urban infrastructure and FOD, which negatively impacts classification (Figure 5).

ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

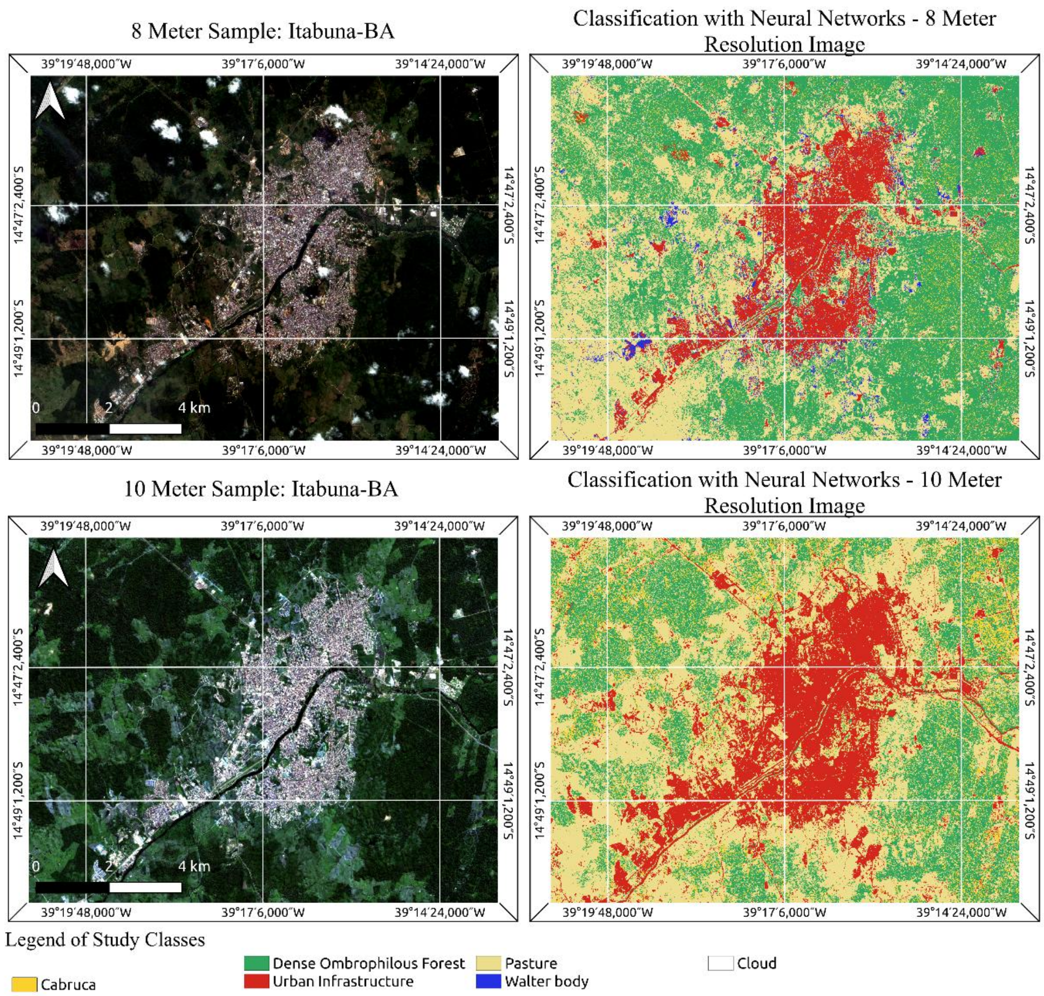

Figure 5 – Classification of land cover and use near the municipality of Itabuna-BA using the Neural Network Model. Source: Elaborated by the authors, 2024.

Furthermore, common classifications are observed for both algorithms, especially under Condition B, where the same cabruca area is represented by a cluster of pixels. It is noteworthy that the pixel arrangement in the 8-meter image resembles an area with low FOD presence, while the classification in the 10-meter image is more conservative, with pixels representing both cabruca and FOD areas. This is corroborated by field data validation, where there are cabruca areas alongside other areas of planted forest (Figure 6).

Similarly, the quality in the estimation of urban infrastructure and pasture areas stands out, with pixels consistently and accurately defined for the evaluated samples. This consistency is close to what was verified both in field data validation and with the use of Google Earth Pro.

RA'EGA
O ESPAÇO GEOGRÁFICO EM ANÁLISE
https://revistas.ufpr.br/raega
ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

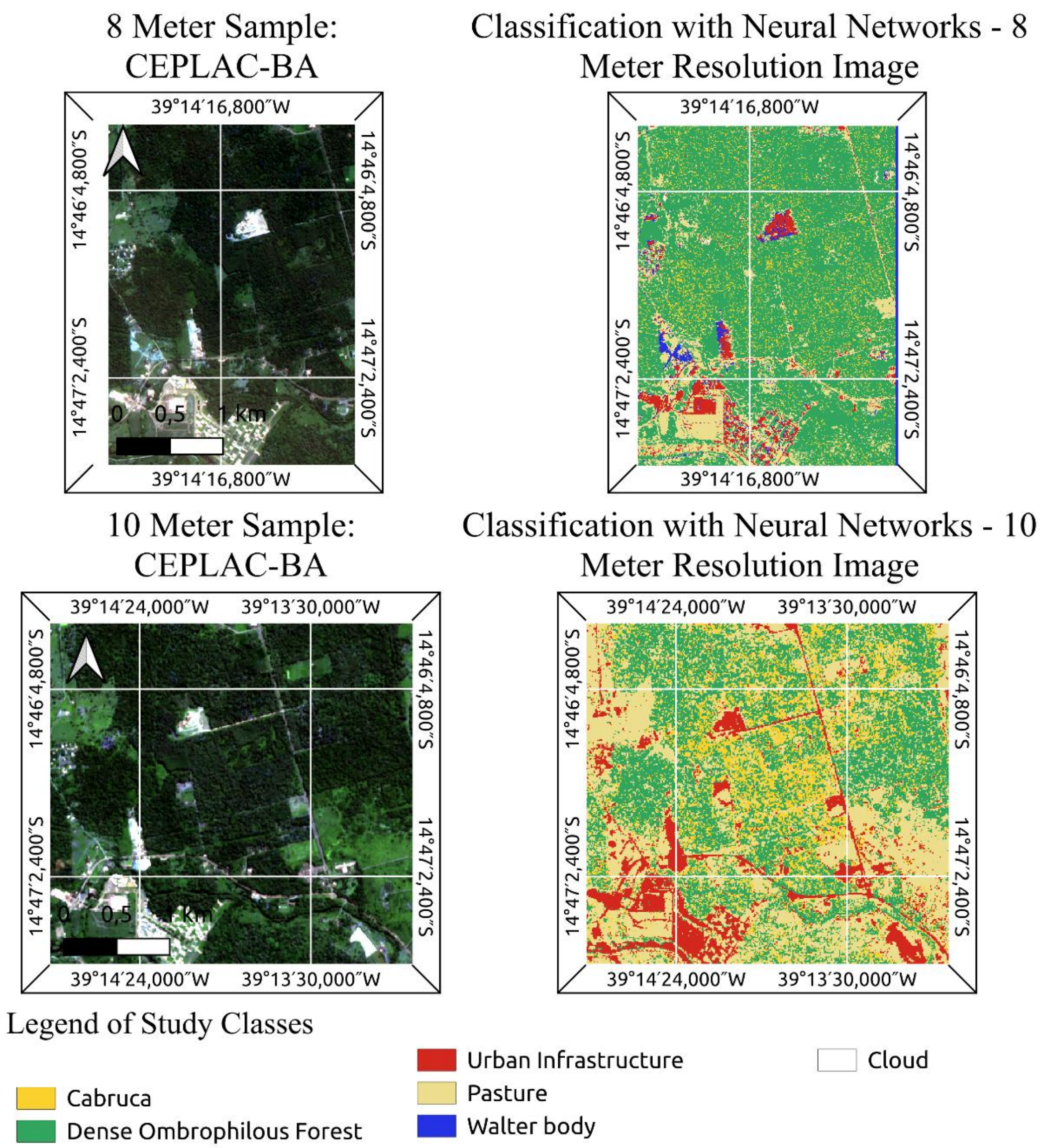

Figure 6 – Classification of land cover and use of the area near CEPLAC-BA using the Neural Network model. Source: Elaborated by the authors, 2024.

The results of the models highlight the capability of classifying images of different resolutions. In the 8-meter resolution images, the model's inability to determine neighborhoods or patterns is evaluated due to the higher pixel density. Additionally, the incompatible details between the 2 and 8-meter images with the 55 and 60-meter resolution images interfere with the precise analysis of image classification.

Thus, in the second round of evaluation, an accuracy of 86% was observed for the Neural Network algorithm using 2-meter resolution images. The classification was balanced across all classes, except for the cabruca class, whose precision was similar to that of the KNN model. Furthermore, to maintain the hypothesis

ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

that resolution diversity provides complementary spectral features that enrich the analysis, no resampling techniques were applied (Table 4).

Table 4 – Evaluation results of the Neural networks models with 2- meter resolution imagens.

Class	Accuracy / %	Recall / %	F1-score / %
Cabruca	49	21	30
Dense ombrophilous forest	85	90	88
Urban infrastructure	88	84	86
Pasture	89	88	88
Walter body	82	93	87
Cloud	91	91	91

Source: the author, 2024.

In Table 4, it is highlighted that the pasture class shows robust metrics for precision, *recall*, and F1-score values, indicating consistent performance of the model in identifying this category, regardless of image resolution. For the other classes, it is emphasized that besides the spatial variables affecting classification quality, consideration of shared common characteristics among them is relevant.

In the application of the model, correct classifications are observed for the water body and cloud classes, due to the implementation of adjustments identified in previous evaluations. These adjustments included the establishment of criteria that highlighted the characteristics of the classes. According to Condé et al. (2023), such characteristics include color, hue or brightness intensity, reflectance, texture, size, shape, and surrounding influence.

Thus, the drainage of the Cachoeira River is identified, along with a conservative classification, where cabruca areas are associated with the presence of the FOD class (Figure 7).

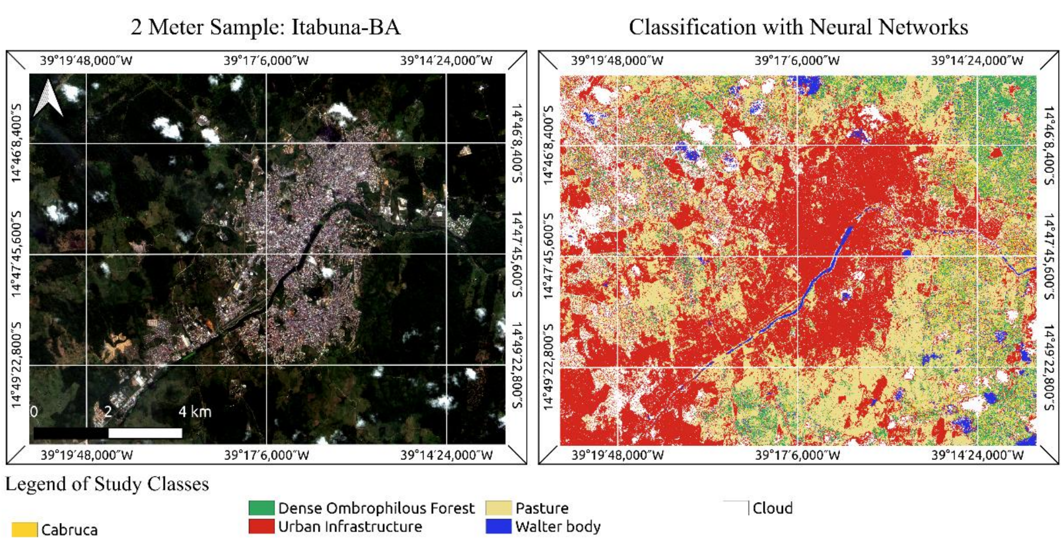

Figure 7 – Land cover classification by the Neural Networks model with 2-meter resolution. Source: Prepared by the authors, 2024.

RA'EGA O ESPAÇO GEOGRÁFICO EM ANÁLISE
https://revistas.ufpr.br/raega
ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

In the analysis of the confusion matrix (Figura 8), the cabruca class overlaps with other classes. Li, Ustin, and Lay (2005) and Shivakumar and Rajashekararadhya (2017) point out that spectral similarities between vegetation influence classification limitations and that the likelihood of producing incorrect classifications is higher for spectrally dependent classes. One condition reinforcing this assumption is that the tree canopies used for shading present challenges in obtaining distinct characteristics specific to cabruca areas.

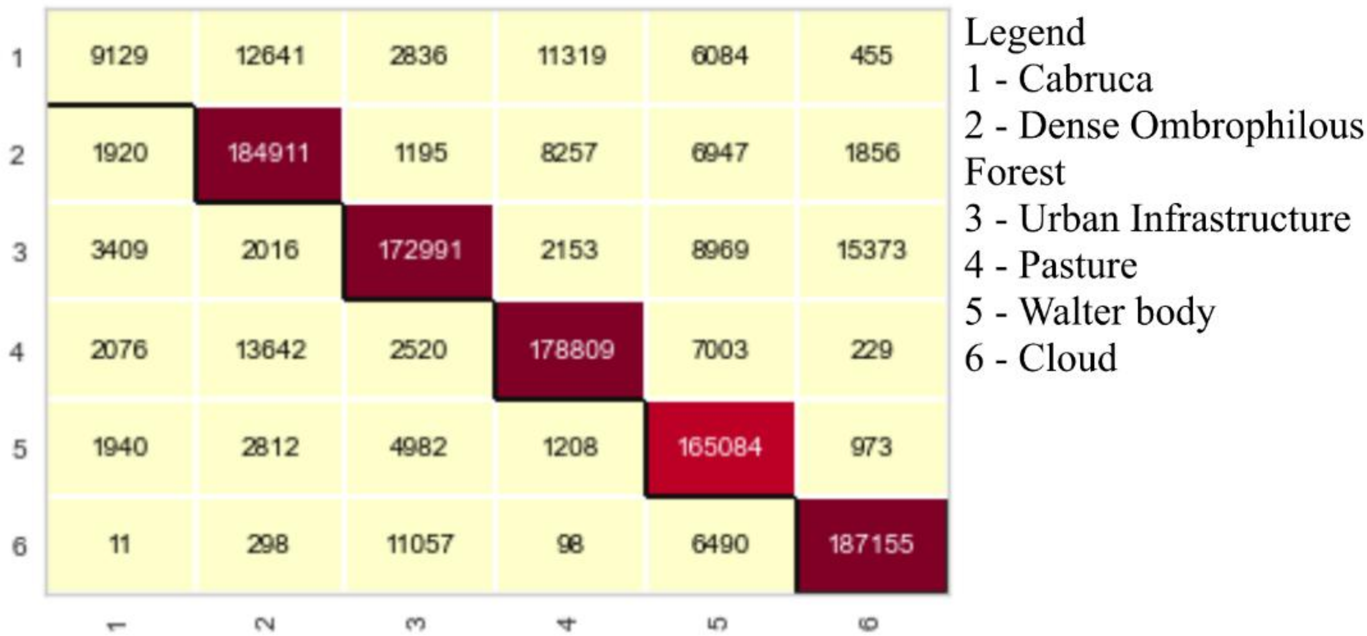

Figure 8 – Confusion Matrix for the Neural networks model with 2- meter resolution. Source: the author, 2024.

Furthermore, to address the interference of uneven lighting and noise, Shi and Pun (2018) propose a new percentage-based loss function based on spectral and structural similarity, while Liu et al. (2024) present three new indices for more precise vegetation extraction. These approaches allow for the correction of such parameters in the images used for training. However, due to the strong spectral similarity between cabruca and FOD, differentiating them remains the main challenge. According to Reuss-Strenzel and Faria (2023), the spectral responses of cabruca are similar to those of Atlantic Forest fragments.

In summarizing the results, it is evaluated that the Neural Network model classification, using the 8-meter image, approximates the estimated area for each class. The manual estimation was carried out by overlaying data from the MapBiomas Collection Eight, due to the difficulty in obtaining a reliable classification for the study area. In this approach, it was considered that 0.05% of the FOD class corresponds to the cabruca area, while 55 km² was allocated to the FOD area, based on field validations (Figure 9).

RA'EGA O ESPAÇO GEOGRÁFICO EM ANÁLISE
https://revistas.ufpr.br/raega
ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

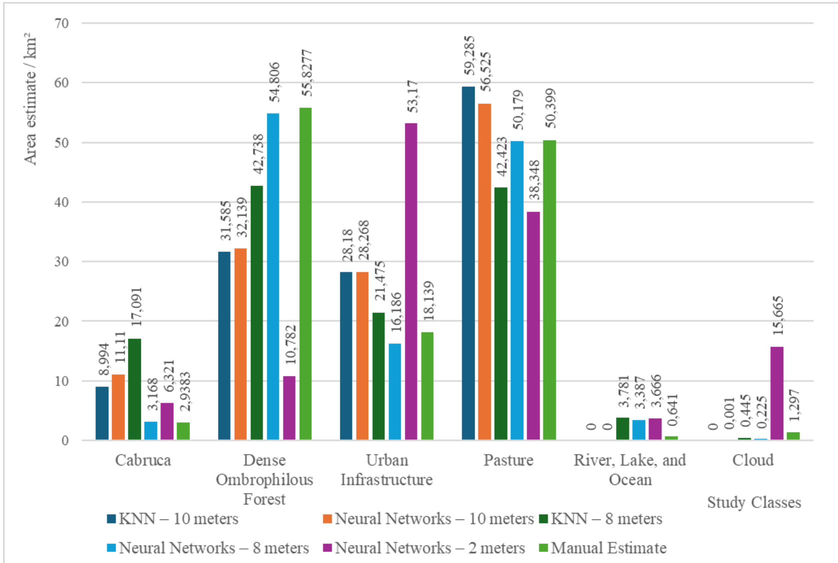

Figure 9 – Comparison between land cover and land use classifications by machine learning models. Source: the authors, 2024.

In Figure 9, it is observed that the percentage errors for the FOD and pasture classes were 1.83% and 0.44% lower than the reference value, respectively, highlighting the superiority of the model compared to the others. In contrast, the cabruca class, with a percentage error of 7.82%, indicates the need for corrections in its classification.

The results show synergy with Fonseca et al. (2023) in the classification of cocoa areas. The term "shade-grown cocoa" is highlighted, which includes both cabruca cocoa and consortium areas. However, this approach may hinder targeted interventions for the conservation and management of cabruca cocoa environments.

Additionally, the limitation of coverage in the BHRA, BHRC, and BHRU is noted due to the approximate extent of 764,000 km² and the need for robust computational processing to store the pixels in memory. For this reason, Mielikainen (2006) emphasizes strategies to store actual pixel values through schemes such as band interleaved by line, band interleaved by pixel, or band sequential.

IX. Conclusions

The results demonstrate the effectiveness of using the clipping algorithm for the development of the geographic database, adapting to the specific characteristics of the study area. The importance of establishing criteria that highlight the characteristics and attributes for each class is emphasized, aiding in the recognition of patterns that enhance classification performance.

ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

The classification of urban infrastructure and pasture areas highlights the potential of the database as an effective tool for training machine learning models. However, there is difficulty in accurately distinguishing the cabruca class, as shown in the tables, for both the KNN and Neural Networks algorithms with 2-meter resolution images. This reinforces that the results should be interpreted as estimates of the presence of cabruca.

On the other hand, for the other classes such as urban infrastructure, pasture, and DOF, it is evaluated that the Neural Networks model showed the best performance, regardless of the diversity of resolutions. Additionally, it is recommended to use 10-meter resolution images for classifying these classes, instead of the *pixel* density of 8-meter resolution images, which demonstrate the model's inability to accurately determine neighborhoods or patterns in land cover and use classification in the image.

Additionally, it is suggested to adopt effective strategies for reading and storing information in computer memory or servers, such as alternative methods for reading image information, especially for large areas that require robust computational resources, as exemplified by the study area of 8600 km². Finally, the importance of interdisciplinary collaboration and synergy with other works that address other areas of the Atlantic Forest biome, such as mangroves and coastal sand dunes (restingas), is emphasized.

Acknowledgments

To the National Council for Scientific and Technological Development (CNPq) for granting a research productivity fellowship to one of the authors. To the Coordination for the Improvement of Higher Education Personnel (CAPES) for granting a postgraduate scholarship to one of the authors. To the Research Support Foundation of Bahia (FAPESB) for providing a master's scholarship to one of the authors.

X. REFERENCES

ALBA, E.; ALEXANDRE, M. L. S.; MARCHESAN, J.; SOUZA, L. S. B.; BEZERRA, A. C.; SILVA, E. A. Comparação entre Algoritmos de Aprendizado de Máquina para a Identificação de Floresta Tropical Sazonalmente Seca. Anuário do Instituto de Geociências, vol. 45, no. 1982-3908, p. 1–10, 2022. DOI https://doi.org/10.11137/1982-3908_2022_45_4075.

BHOSALE, V.; PATANKAR, A. Performance Comparision of VGG-16 and ResNet-34 Algorithms for Supervised Classification of Landsat Images. 16 Jul. 2023. DOI https://doi.org/10.1109/igarss52108.2023.10281537.

CHATTERJEE, S.; BHATTACHARYYA, S.; SAHA, S.; HALDER, A.; RANJAN, P. Detection of Dense Built-Up Area in Low-Resolution Satellite Images Using Deep Learning and DBSCAN Approaches. International Journal of Pat-

O ESPAÇO GEOGRÁFICO EM ANÁLISE
https://revistas.ufpr.br/raega
ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

tern Recognition and Artificial Intelligence, vol. 38, no. 04, 30 Mar. 2024. https://doi.org/10.1142/s0218001424510042.

CONDÉ, T. M.; HIGUCHI, N.; ADRIANO; ASSIS, A.; JACKELIN DIAS CONDÉ; CAMARGO, A.; CARNEIRO, L. Spectral Patterns of Pixels and Objects of the Forest Phytophysiognomies in the Anauá National Forest, Roraima State, Brazil. Ecologies, vol. 4, no. 4, p. 686–703, 28 Oct. 2023. DOI https://doi.org/10.3390/ecologies4040045.

DAS, P.; PANDEY, V. Use of Logistic Regression in Land-Cover Classification with Moderate-Resolution Multispectral Data. Journal of the Indian Society of Remote Sensing, vol. 47, no. 8, p. 1443–1454, 8 May 2019. https://doi.org/10.1007/s12524-019-00986-8.

FAGUNDES, W. S.; JÚNIOR, M. J. A. Evolução das Técnicas de Classificação de Imagens do Sensoriamento Remoto Utilizadas na Produção Científica Brasileira: Evolution of remote sensing image classification techniques used in Brazilian scientific production. Geosciences = Geociências, vol. 41, no. 3, p. 593–604, 2022. DOI https://doi.org/10.5016/geociencias.v41i03.16209.

FONSECA, M. G. et al. Mapbiomas Cacau: avanços e desafios no mapeamento do cultivo sombreado de cacau no Sul da Bahia. [S.l.]: [s.n.], 2023.

GAIAD, S.; CARVALHO, P. E. R. Glossário — Portal Embrapa. 22 dez. 2021. Disponível em: https://www.embrapa.br/.

GOOGLE COLABORATORY. Disponível em: https://colab.research.google.com/. Acesso em: 25 out. 2025.

GOOGLE EARTH ENGINE TEAM. Google Earth Engine: a planetary-scale geospatial analysis platform. [S.l.]: [s.n.], 2013. Disponível em: https://earthengine.google.com/. Acesso em: 25 out. 2025.

HAFSA, O.; ABDESSAMAD, A. E. B. Machine learning algorithms for satellite image classification using Google Earth Engine and Landsat satellite data: Morocco case study. IEEE Access, v. 11, p. 71127–71142, 2023. https://doi.org/10.1109/ACCESS.2023.3293828.

LI, L.; USTIN, S. L.; LAY, M. Application of multiple endmember spectral mixture analysis (MESMA) to AVIRIS imagery for coastal salt marsh mapping: a case study in China Camp, CA, USA. International Journal of Remote Sensing, vol. 26, no. 23, p. 5193–5207, 10 Dec. 2005. https://doi.org/10.1080/01431160500218911.

LIANG, J.; LIANG, G.; SUN, L. Using Sentinel images for analyzing water and land separability in an agricultural river basin. Environmental Monitoring and Assessment, vol. 195, no. 11, 13 Oct. 2023. https://doi.org/10.1007/s10661-023-11908-0.

LISBOA, G. P. Validação do Modelo Linear de Mistura Espectral e Índices De Vegetação para Identificação de Fragmentos Florestais de Mata Atlântica e Sistema Agroflorestal Cacau Cabruca Utilizando Imagens Orbitais. 2022.

LIU, X.; ZHENG, X.; WANG, Z.; LI, Z.; WANG, K.; ZHANG, H.; DUAN, J. Monitoring Wetland Changes and Analyzing the Spartina alterniflora Invasion in the Yellow River Delta over the Past 30 Years Based on Google Earth Engine. IEEE Geoscience and Remote Sensing Letters, vol. 21, p. 1–5, 1 Jan. 2024. DOI https://doi.org/10.1109/lgrs.2024.3400027.

MIELIKAINEN, J. Lossless compression of hyperspectral images using lookup tables. IEEE Signal Processing Letters, vol. 13, no. 3, p. 157–160, Mar. 2006. https://doi.org/10.1109/lsp.2005.862604.

MORAITIS, N.; TSIPI, L.; VOUYIOUKAS, D.; GKIONI, A.; LOUVROS, S. Performance evaluation of machine learn-

ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

ing methods for path loss prediction in rural environment at 3.7 GHz. Wireless Networks, vol. 27, no. 6, p. 4169–4188, 18 Jun. 2021. https://doi.org/10.1007/s11276-021-02682-3.

MOREIRA, M. A.; ADAMI, M.; RUDORFF, B. F. T. Análise espectral e temporal da cultura do café em imagens Landsat. Pesquisa Agropecuária Brasileira, v. 39, n. 3, p. 223–231, mar. 2004. https://doi.org/10.1590/S0100-204X2004000300004.

NAZMUZZAMAN, K. M.; ANWAR, S. Robust Weed Recognition Through Color Based Image Segmentation and Convolution Neural Network Based Classification. Nov. 2019.

PATEL, A. K.; CHATTERJEE, S.; GORAI, A. K. Development of machine vision-based ore classification model using support vector machine (SVM) algorithm. Arabian Journal of Geosciences, vol. 10, no. 5, Mar. 2017. https://doi.org/10.1007/s12517-017-2909-0.

ROCHA, J.; TENEDÓRIO, J. A.; ESTANQUEIRO, R.; SOUSA, P. M. Classificação de uso do solo urbano através da análise linear de mistura espectral com imagens de satélite. Finisterra, v. 42, n. 83, 13 dez. 2012. https://doi.org/10.18055/finis1438.

SAMBUICHI, R. H. R. Estrutura e dinâmica do componente arbóreo em área de cabruca na região cacaueira do sul da Bahia, Brasil. Acta Botanica Brasilica, vol. 20, no. 4, p. 943–954, Dec. 2006. https://doi.org/10.1590/s0102-33062006000400018.

SENA SOUZA, J. P. et al. Comparação dos métodos de classificação por ângulo espectral e distância euclidiana no mapeamento das formas de terreno. Revista Brasileira de Geomorfologia, v. 17, n. 3, 1 set. 2016. https://doi.org/10.20502/rbg.v17i3.846.

SHI, C.; PUN, C.-M. Perceptual Loss for Superpixel-Level Multispectral and Panchromatic Image Classification. 1 Apr. 2018. DOI https://doi.org/10.1109/icassp.2018.8461966. Available at: https://ieeexplore.ieee.org/document/8461966.

SHIVAKUMAR, B. R.; RAJASHEKARARADHYA, S. V. Spectral similarity for evaluating classification performance of traditional classifiers. 1 Mar. 2017. DOI https://doi.org/10.1109/wispnet.2017.8300111.

SMITH, A. R.; JONES, M. L. Optimization techniques in machine learning for satellite image analysis. Computational Intelligence in Remote Sensing, vol. 8, p. 125–138, 2021. https://doi.org/10.1016/j.cirs.2021.06.004.

SOUZA, C. M. P. de; MOREAU, M. S.; MOREAU, A. M. S. dos S.; FONTES, E. de O. Níveis de degradação de pastagens da bacia do Rio Colônia-BA com uso de imagens LANDSAT 5 TM. Revista Brasileira de Geografia Física, v. 3, n. 3, p. 228–243, 7 jul. 2011. https://doi.org/10.26848/rbgf.v3i3.232685.

TAN, Z.; LIU, C.; MA, S.; WANG, X.; SHANG, J.; WANG, J.; AI, W.; YAN, W. Detecting Multilayer Clouds From the Geostationary Advanced Himawari Imager Using Machine Learning Techniques. IEEE Transactions on Geoscience and Remote Sensing, vol. 60, p. 1–12, 2022. DOI https://doi.org/10.1109/TGRS.2021.3087714.

TEIXEIRA, A. C. de O. et al. Análise da dinâmica de uso e ocupação da terra na bacia hidrográfica do baixo Rio de Contas – Bahia, entre os anos de 1973 a 2001. Revista Geoaraguaia, 18 set. 2013.

VALADARES, J. O. Método de classificação especialista para imagens de sensoriamento remoto para o mapeamento de áreas de cultivo de cacau-cabruca. repositorio.ufba.br, 20 Jun. 2016.

XAVIER, L. B.; NASCIMENTO JR, F. D. C. do; CHIAPETTI, J. Da crise regional às novas dinâmicas de especialização da produção de cacau no Sul da Bahia. Caminhos de Geografia, v. 22, n. 79, p. 77–96, 1 fev. 2021b.

ISSN eletrônico 2177-2738
RA'EGA, Curitiba, PR, V.64, n.1, p. 71 – 91, 12/2025
http://dx.doi.org/10.5380/raega.v64i1.101334

https://doi.org/10.14393/rcg227953461.

XAVIER, L. B.; NASCIMENTO JR, F. D. C. do; CHIAPETTI, J. Verticalidades e horizontalidades na certificação da produção de cacau orgânico no Sul da Bahia. Revista Campo-Território, v. 15, n. 38, p. 361–390, 23 abr. 2021a. https://doi.org/10.14393/rct153814.