

# Identification of gold mining vessels based on classification algorithms using Sentinel-2 images

*Diego Henrique Costa Pereira*<sup>1</sup> 

*Roberto Arnaldo Trancoso Gomes*<sup>2</sup> 

*Osmar Abílio de Carvalho Júnior*<sup>3</sup> 

*Renato Fontes Guimarães*<sup>4</sup> 

## Keywords

Gold mining vessels  
Classifiers  
Machine learning

## Abstract

Artisanal and small-scale gold mining can occur on land or in riverbeds. However, the activity needs to be supported by a Mining Permit, issued by the Agência Nacional de Mineração, which is the Brazilian institution responsible for mining management, and the appropriate environmental license from the competent environmental agency. The use of images from Sentinel-2 satellites presents itself as a potential tool for identifying gold mining vessels due to the temporal resolution, free imagery, global coverage, and more refined spatial resolution. So, this study aimed to identify gold mining vessels on the Madeira River near Porto Velho city, Rondônia state, located at Brazilian Amazon, in 13 Sentinel-2 images from 2018 to 2021 using the classifiers: Support Vector Machine (SVM), K-Nearest Neighbors (KNN) Random Forest (RF) and Spectral Angle Mapper (SAM). The results showed that machine learning classifiers obtained the best performance, especially the object-oriented SVM classifier, which had the best average F1 score (0.91). In addition, the detection percentage of gold mining vessels originated by this classifier was satisfactory, with only 0 to 4 active gold mining vessels with sediment plumes being omitted per image. Therefore, based on the results obtained, it was concluded that the use of machine learning classifiers proved to be effective in identifying gold mining vessels.

<sup>1</sup> Universidade de Brasília - UnB, Brasília, DF, Brazil. [diegohcpereira1988@gmail.com](mailto:diegohcpereira1988@gmail.com)

<sup>2</sup> Universidade de Brasília - UnB, Brasília, DF, Brazil. [robertogomes@unb.br](mailto:robertogomes@unb.br)

<sup>3</sup> Universidade de Brasília - UnB, Brasília, DF, Brazil. [osmarjr@unb.br](mailto:osmarjr@unb.br)

<sup>4</sup> Universidade de Brasília - UnB, Brasília, DF, Brazil. [renatofg@unb.br](mailto:renatofg@unb.br)

## INTRODUCTION

Artisanal and small-scale gold mining activities occur on land or in riverbeds and they can operate in possession of an Artisanal and Small-Scale Gold Mining Permit (PLG in Portuguese) issued by the Agência Nacional de Mineração, which is the institution responsible for mining management in Brazil (Brasil, 1989). Gold mining in rivers are vessels that dredge bottom sediments that can contain from 3 to 30 g/m<sup>3</sup> of gold (Balzino *et al.*, 2015). These vessels have all the necessary machinery to produce gold. Also, they are characterized by releasing tailing from the production process directly into the river and grouping it together during the operation (Figure 1). The tailings released have a higher sediment concentration than the natural water

and what can lead to the formation of sediment plumes (Fernandes *et al.*, 2014).

The plumes formation depends on the dredging capacity, type of sediment, watercolor, among others. (Cetem, 2018; Barbosa *et al.*, 2019).

The spectral response of water registered by optical sensors is a result of the interaction of electromagnetic radiation with the water column and its Optically Active Compounds (OAC), such as inorganic particles, organic matter and phytoplankton pigments (Lobo *et al.*, 2012).

Barbosa *et al.* (2019) compiled several studies that apply remote sensing in water resources, demonstrating the capability to differentiate watercolor, to estimate OAC and to evaluate reflectance of each type of OAC along the electromagnetic spectrum.

Figure 1 - Example of a Gold Mining Vessel.



Source: Balzino *et al.* (2015).

Several remote sensing methods are used to identify gold mining sites located on land (Asner *et al.*, 2013; Isidro *et al.*, 2017; Simionato *et al.*, 2021). Yet, methods dedicated exclusively to the identification of gold mining vessels are scarce due to: small size of the targets and its high locomotion capacity (Lobo *et al.*, 2018; Trindade; Barbosa Filho, 2002); the lack of automatic and real-time location systems (Dechesne *et al.*, 2019), either by the relevant level of informality or the precariousness and small size of the vessels (Cetem, 2018).

Despite the lack of methods related to gold mining vessels, there are several methods that use satellite images for vessels detection. Ciocarlan and Stoian (2021) used the U-Net deep learning architecture, achieving a mean F1 score of 53.1. They concluded that such methods show promising results, obtaining a lower incidence of false negatives, both for large (> 2,500 m<sup>2</sup>) and small (<2,500 m<sup>2</sup>) vessels.

However, they emphasized that these methods are sensitive to the quantity and quality of the training samples. Heiselberg (2016) proposed an algorithm for geometric and spectral evaluation of the vessels. The author was successful in detecting vessels, identifying the navigation orientation, and measuring the width and length of targets above 30 m. Vessels smaller than 30 m were only detected, without estimating the other parameters.

Heiselberg and Heiselberg (2017) developed a supervised object-oriented classifier for detecting vessels and differentiating them from other targets (ice, small islands, and clouds). They highlighted the loss of efficiency as vessel size decreases, since 32% of vessels smaller than 10 m were confused with clouds and other small objects.

Kanjir (2019) achieved satisfactory results related to vessels detection, including those smaller than 20 m, reaching a detection rate up

to 0.96. A decision tree was used after performing a pre-detection based on band subtraction. The author pointed out that the low presence or absence of vessels per image can lead to a deterioration in the efficiency of the method, such as the small size of the target.

So, one of the major challenges for these studies is to obtain enough quality training samples as there may be few vessels in the selected images (Ciocarlan; Stoian, 2021).

Therefore, the present study aims to evaluate the use of Sentinel-2 images to identify gold mining vessels using the machine learning classifiers Support Vector Machine (SVM) (Vapnik, 1995), K-Nearest Neighbors (KNN) (Fix; Hodges, 1951), Random Forest (RF) (Breiman, 2001), and the spectral classifier Spectral Angle Mapper (SAM) (Kruse *et al.*, 1993).

## MATERIALS AND METHODS

### Study area

The study area comprises the Madeira River, in the region of Porto Velho city, Rondônia state, located in Brazilian Amazon (Figure 2). The study area covers 80 km of the river, and it is one of the main gold mining spots through vessels in Brazil. By the time of this study, there

were 9,400 hectares of Small-Scale Gold Mining Permits (in Portuguese PLG) and 14,400 hectares of PLG Requests (Sigmine, 2022).

The Madeira River is a white-water river with a high concentration of suspended solids (Espinoza Villar *et al.*, 2013). The region's climate is classified as Awi and its rainfall ranges from 38.7 to 329 mm/month (Bezerra *et al.*, 2010). Its hydrological behavior is typically seasonal, with water flow from 5,000 m<sup>3</sup>/s in the dry season to 50,000 m<sup>3</sup>/s in the rainy season (Adamy, 2016).

In addition, it has good navigability conditions with a width of 700 to 2,000 m and fluvimetric level from 8 to 20 m (Rodriguez *et al.*, 2017).

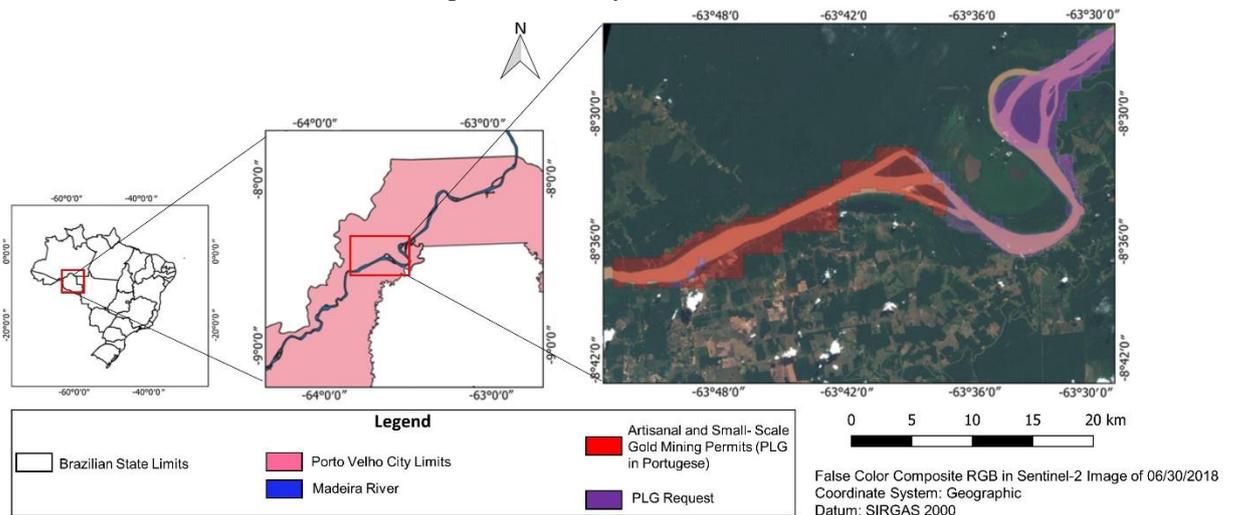
### Methodology

The methodology was divided into five steps, as shown in Figure 3.

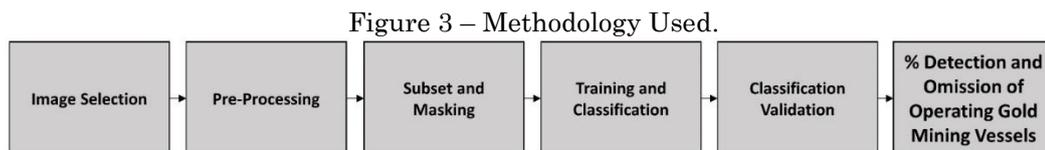
#### Step 1 – Image selection

The Sentinel-2 mission is composed by the 2A and 2B satellites with *Multispectral Instruments* (MSI). The MSI registers information in 13 spectral bands from blue to shortwave infrared with a revisit time of 5 days, when considering both satellites, and 12-bit radiometric resolution.

Figure 2 – Study Area.



Source: The authors (2023).



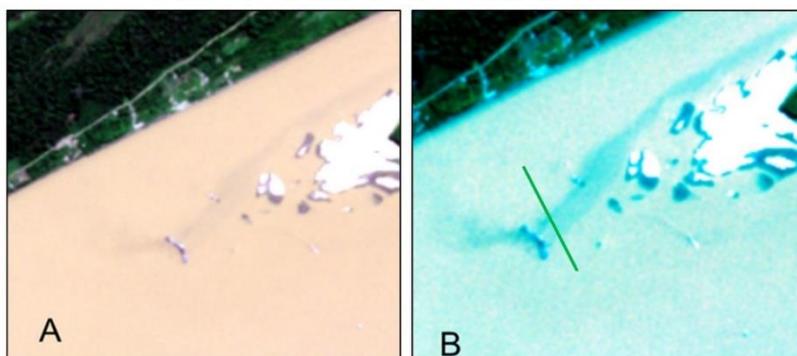
Source: The authors (2023).

The images are free and georeferenced, and the bands present the following spatial resolutions: i) 10 m for Blue, Green, Red and Near Infrared (NIR); ii) 20 m for Red Edge (RE) 1 to 4 and Shortwave Infrared 1 and 2 (SWIR); iii) 60 m for Aerosol, Water Vapor and Cirrus (ESA, 2015). Images from the 20LMR orbit between the years 2018 and 2021 were

evaluated to select those with the presence of gold mining vessels that coincided with the dates of the images that were used to validate the classifications.

The evaluation was performed visually by applying histogram enhancement in a false-color composite using Red, Green and Blue bands (Figure 4).

Figure 4 – Sediment Plumes from Gold Mining Vessels in False Color Composite (Blue, Green and Red Bands) in 06/14/2021 Sentinel-2 image: a) No enhancement; b) With Enhancement of the Red band. Green Line is the Sediment Plume.



Source: The authors (2023).

Thirteen images between June and September from 2018 to 2021 were selected (Table 1). This is the period of lowest rainfall

and cloud cover from 0 to 40% in the Amazon region (Martins *et al.*, 2018).

**Table 1 – Date of Sentinel-2 Images Used**

Date	Gold Mining Vessels Without Sediment Plume	Gold Mining Vessels With Sediment Plume
06/30/2018	Yes	Yes
07/25/2018	Yes	Yes
08/19/2018	Yes	No
06/20/2019	Yes	No
07/09/2020	Yes	Yes
09/22/2020	Yes	No
06/14/2021	Yes	Yes
07/14/2021	Yes	Yes
07/24/2021	Yes	Yes
07/29/2021	Yes	Yes
08/08/2021	Yes	No
08/13/2021	Yes	Yes
08/18/2021	Yes	Yes

Source: The authors (2023).

### Step 2 - Pre-Processing

Images were acquired at the L1C processing level without atmospheric correction. They were converted to the L2A level to discount atmospheric effects using the Sen2Cor plugin, developed for Sentinel-2 images (ESA, 2015; Louis *et al.*, 2016).

Subsequently, the infrared bands were resampled to 10 m, except for the Near Infrared, which already has this resolution.

### Step 3 – Subset and Masking

The images were subsetted with the limits of the study area and a mask was applied in the continental portion. The masks were generated from manual editing of the vector file of water bodies of Brazil – produced in scale 1:100,000 by the Agência Nacional de Águas e Saneamento, which is the federal agency responsible for the implementation and regulation of Brazilian water resources and sanitation policies (Brasil, 2022).

### Step 4 – Training and Classification.

The training samples were collected in the thirteen Sentinel-2 images used. The number of samples was not the same for each image, since the number of vessels, gold mining vessels and sediment plumes varies between different dates. Also, sometimes there are few targets in the image.

3,000 pixels were collected for water (in the absence of sediment plumes), vegetation and soil. Vegetation and soil occur on the islands and on the sandbanks along the river. The collected number of pixels reached the maximum possible when there were sediment plumes in the image. In these cases, the amount of water pixels was the same. Also, the maximum possible number of pixels per image were collected for gold mining vessels (Table 2 and Figure 5).

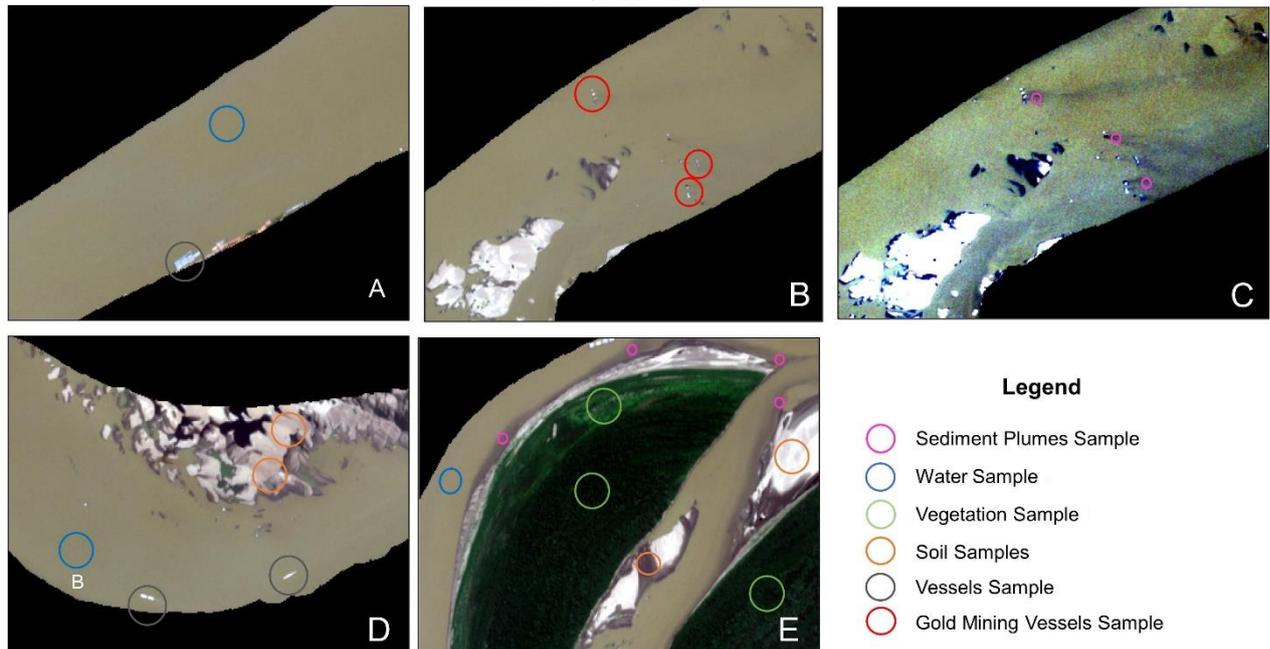
RF, KNN, SVM and SAM were used to classify the images. RF randomly generates decision trees based on a subset of data and search for the best result from a voting process (Breiman, 2001). KNN evaluates the distance of the sample in relation to its “k” nearest neighbors (Fix; Hodges, 1951).

**Table 2 - Number of Training Pixels per Image.**

Date	Classes					
	Vegetation	Soil	Water	Sediment Plumes	Vessels	Gold Mining Vessels
06/30/2018	3000	3000	1000	1000	250	95
07/25/2018	3000	3000	1200	1200	330	145
08/19/2018	3000	3000	3000	None	230	100
06/20/2019	3000	3000	3000	None	380	80
07/09/2020	3000	3000	1500	1500	390	100
09/22/2020	3000	3000	3000	None	160	45
06/14/2021	3000	3000	1000	1000	130	100
07/14/2021	3000	3000	2000	2000	300	200
07/24/2021	3000	3000	1000	1000	270	225
07/29/2021	3000	3000	1500	1500	450	160
08/08/2021	3000	3000	3000	None	285	120
08/13/2021	3000	3000	1000	1000	420	100
08/18/2021	3000	3000	1000	1000	255	135

Source: The authors (2023).

Figure 5 - Examples of Samples: a) Water and Vessels; b) Gold Mining Vessels; c) Sediment Plumes with Enhancement; d) Vessels and Water; e); Vegetation, Water and Sediment Plumes Without Enhancement.



Source: The authors (2023).

SVM plots samples as points in a  $n$ -dimensional space and searches for the best way to separate classes by building a hyperplane that acts as a boundary between classes (Vapnik, 1995). The SAM is based on the similarity of the spectral responses and calculates angles between the samples to classify the pixels (Kruse *et al.*, 1993). RF, KNN and SVM are machine learning classifiers. They may have a pixel-based approach or object-oriented (OB) approach. The pixel-based approach uses the spectral information of the pixels as the object-oriented approach group pixels to create objects that have a certain similarity. The segmentation is a crucial step within this approach (Simionato *et al.*, 2021).

The segmentations were performed in the ArcGIS Pro software using the Mean Shift Segmentation algorithm (Comaniciu; Meer, 2002).

The pixels are plotted in a  $n$ -dimensional space and the algorithm builds an interest area and then calculates the centroid and the direction in which the area of interest should move. After, the interest area changes its centroid and repeats the process until convergence is reached and all pixels are grouped into segments. ArcGIS Pro software requires three parameters to perform the segmentation: spectral and spatial detail (unit from 1 to 20); and minimum segment size (unit in pixels) (ArcGIS PRO, 2023). Spectral detail refers to the level of importance given to spectral

differences between pixels. Spatial details control the level of relevance given to the proximity between the pixels. Minimum size segment one is the minimum size of the segment that will be created (Lodi *et al.*, 2019). As for spectral detail, larger values are appropriate for separating objects that have similar spectral characteristics, while smaller values are appropriate for object generalization. For spatial detail, larger values are appropriate when objects are small and form clusters (ArcGIS PRO, 2023).

Different combinations were tested, and the segmentations were visually evaluated. Two main questions were observed: if the generated segments were consistent to the size of the gold mining vessels: if it was possible to separate water from sediment plumes. By the end, it was decided to use 20 for spectral detail and 18 for spatial detail. This combination generated segmentations composed by several small segments suitable to separate gold mining vessels and sediment plumes. The minimum size segment used was 1 pixel (10 m). It is the smallest size of a gold mining vessel mine according to the resolution of the Sentinel-2 images.

Finally, the training pixels of the pixel-based approach were plotted on the segmented image to collect the training segments.

### Step 5 – Classification Validation.

300, 150 or 100 pixels were used to validate the classifications. The quantity of pixels varied depending on the quantity and size of vessels, gold mining vessels and sediment plumes present in the images.

The pixels were collected using Planet images with 3 m of spatial resolution that were made available by the “Brasil MAIS” Programme, coordinated by the Polícia Federal, which is the Federal Policy of Brazil (Redemais/MJSP © 2020 Planet, 2022). The Kappa coefficient and F1 score were used to validate the classification (Cao *et al.*, 2020; Salgado *et al.*, 2019).

The F1 score (Equation 1) is calculated based on the precision (Equation 2) and recall (Equation 3) that indicate the accuracy of target detection and discrimination capability.

$$\text{Score F1} = 2 * \frac{P * R}{P + R} \quad (1)$$

$$P = \frac{\Sigma TP}{\Sigma TP + \Sigma FP} \quad (2)$$

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

P = precision; R = recall; TP = true positive; FP = false positive; FN = false negative.

Regarding F1 results, a balance between precision and recall must be sought, since a high result of precision and low recall means that few targets were properly classified. However, these few targets are very similar to the control data. It has relations to false negatives. On the other hand, low precision and high recall indicate that the targets were widely classified. But the targets are generalized. It is related to false positives (Cao *et al.*, 2020; Barbosa *et al.*, 2021).

Finally, the percentage of detection and omission of operating gold mining vessels were calculated according to Equations 4 and 5.

Gold mining vessels that originate sediment plumes were considered as operating gold mining vessels (Figure 4).

$$\% \text{ Detection} = \left( \frac{G_d * 100}{G_t} \right) \quad (4)$$

$$\% \text{ Omission} = \left( \frac{G_o * 100}{G_t} \right) \quad (5)$$

G<sub>d</sub> = operating gold mining vessels detected;  
G<sub>t</sub> = total of operating gold mining vessels in the image; G<sub>o</sub> = operating gold mining vessels omitted.

## RESULTS AND DISCUSSION

The targets in the images were water, sediment plumes, gold mining vessels, vessels (various sizes), soil, sandbanks, and vegetation (Figure 6a). The waters of the Madeira River are turbid, and they present high concentration of suspended sediments with naturally higher reflectance than water with a low concentration of sediments (Espinoza Villar *et al.*, 2013).

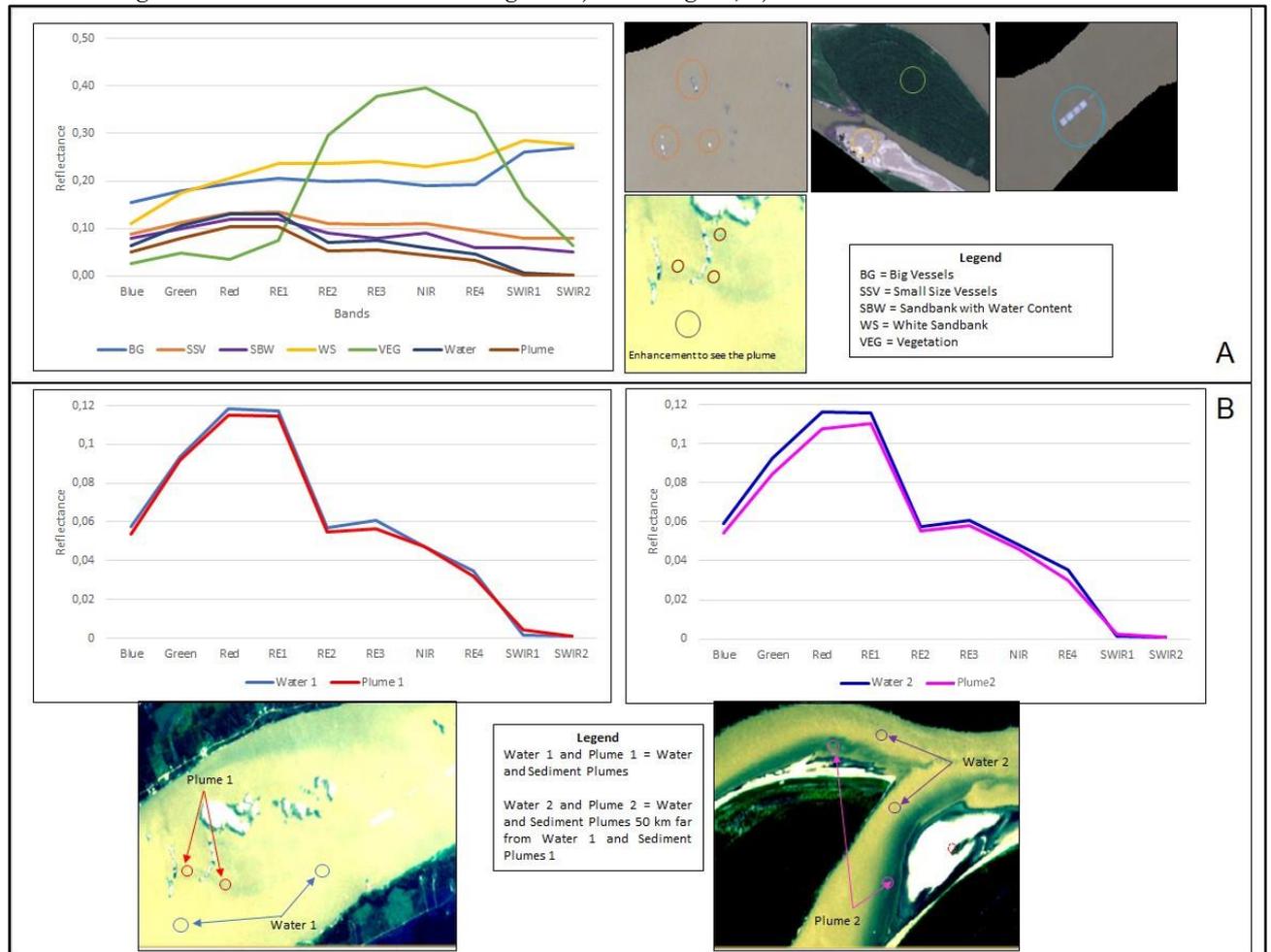
Water and sediment plumes showed very close reflectance, but sediment plumes reflectance is a little lower than the waters surrounding the gold mining vessels. The reflectance peak, as well as the greatest difference between these targets (0.03), occurred in the Green, Red and RE1 bands. The higher difference was realized in the Red band, but still very low. Higher reflectance in Green and Red bands occurs due to greater scattering of electromagnetic radiation by the presence of particulate matter and lower absorption of water in this range (Barbosa *et al.*, 2019; Goodin *et al.*, 1993).

The difference between water and sediment plumes, as well as their reflectance value, decreases from RE1, tending to zero in the SWIR region.

It is noteworthy that the natural concentration of OAC in Amazonian rivers varies depending on the hydrological period – dry or flood season (Barbosa, 2005).

Furthermore, the sediment plumes (Figure 6b) may show the same spectral behavior as the water in other parts of the river when the concentration of suspended sediments is similar. This was noticed in areas close to sandbanks.

Figure 6 – Reflectance of the Targets: a) All Targets; b) Water and Sediment Plume.



Source: The authors (2023).

Heiselberg (2016) identified that large vessels (>100 m) present higher reflectance, reaching higher values in SWIR bands. White sandbanks present similar behavior. These two targets have similar spectral response, with greatest difference in the RE1 band. Smaller vessels, gold mining vessels and sandbanks with water content, with muddy appearance, present inferior and close reflectance in all bands.

The smaller dimensions of the vessels make it difficult to register them into images due to the limitation of spatial resolution of the sensor and the cover type of the vessels. Straw, wood, and similar materials present low reflectance (Kanjir, 2019). Finally, the classes used in the study were: water; sediment plumes (if any); vessels (large, medium, and small); gold mining vessels (grouped medium and small vessels); soil; and vegetation.

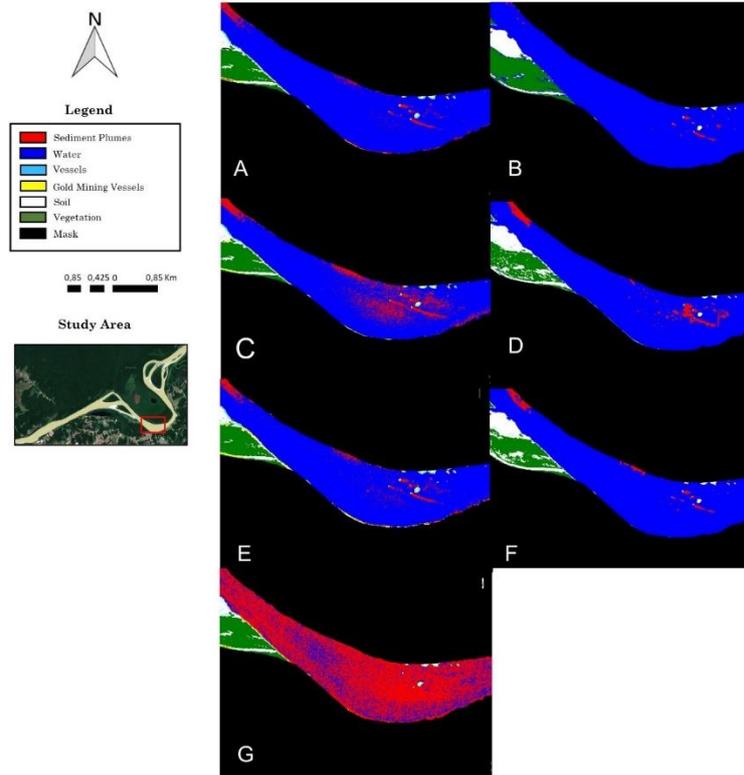
Figure 7 exemplifies the classifications of the 07/14/2021 image. The Figure highlights how the sediment plumes were detected according to

the classifier used. Table 3 shows that the Kappa coefficient ranged from 0.70 to 0.95 - results considered “very good” ( $0.6 < k < 0.8$ ) or “excellent” ( $0.8 < k < 1$ ) (Salgado *et al.*, 2019). Also, the F1 score ranged from 0.74 to 0.96. “Very satisfactory” results (Barbosa *et al.*, 2021; Cao *et al.*, 2020).

The mean Kappa ranged from 0.76 to 0.89 and mean F1 from 0.80 to 0.91, with a standard deviation from 0.06 to 0.02. In this sense, SVM OB performed best, followed by SVM Pixel, KNN OB, KNN Pixel, RF OB, RF Pixel and SAM (Table 4).

These results were similar to the literature in which the performance factor varied from up to 0.68 to 1 (Heiselberg; Heiselberg, 2017). It is noteworthy that mean Kappa and F1 were not so different between the SVM, KNN and RF classifiers: Kappa from 0.81 to 0.89; F1 from 0.84 to 0.91

Figure 7- 07/14/2021 Image Classifications: a) SVM Pixel; b) SVM OB; c) KNN; Pixel; d) KNN OB; e) RF Pixel; f) RF OB; g) SAM.



Source: The authors (2023).

**Table 3 - Kappa Coefficient and Score F1 Results.**

Classifier	Metric	Data												
		06/30/18	07/25/18	08/19/18	06/20/19	07/09/20	09/22/20	06/14/21	07/14/21	07/24/21	07/29/21	08/08/21	08/13/21	08/18/21
SVM Pixel	Kappa	0,86	0,83	0,84	0,88	0,80	0,84	0,87	0,89	0,87	0,90	0,86	0,85	0,85
	F1	0,88	0,86	0,87	0,91	0,83	0,87	0,89	0,91	0,89	0,92	0,88	0,88	0,87
SVM OB	Kappa	0,88	0,84	0,93	0,93	0,87	0,87	0,95	0,89	0,91	0,91	0,87	0,89	0,89
	F1	0,89	0,86	0,94	0,94	0,88	0,89	0,96	0,91	0,92	0,92	0,89	0,90	0,91
KNN Pixel	Kappa	0,81	0,80	0,83	0,86	0,81	0,80	0,86	0,88	0,86	0,89	0,86	0,85	0,85
	F1	0,84	0,84	0,86	0,89	0,83	0,84	0,88	0,90	0,88	0,91	0,88	0,87	0,88
KNN OB	Kappa	0,81	0,80	0,87	0,88	0,81	0,82	0,84	0,86	0,87	0,88	0,86	0,86	0,85
	F1	0,84	0,83	0,89	0,90	0,83	0,85	0,86	0,89	0,89	0,90	0,88	0,88	0,87
SAM	Kappa	0,70	0,73	0,80	0,85	0,72	0,79	0,82	0,72	0,73	0,71	0,82	0,83	0,73
	F1	0,74	0,76	0,84	0,87	0,76	0,82	0,85	0,77	0,77	0,75	0,85	0,85	0,78
RF Pixel	Kappa	0,73	0,79	0,83	0,74	0,71	0,74	0,81	0,86	0,87	0,89	0,83	0,83	0,92
	F1	0,77	0,82	0,85	0,76	0,74	0,79	0,84	0,87	0,90	0,91	0,86	0,82	0,93
RF OB	Kappa	0,74	0,85	0,85	0,84	0,84	0,79	0,84	0,83	0,84	0,82	0,85	0,85	0,82
	F1	0,77	0,87	0,86	0,86	0,86	0,80	0,86	0,84	0,85	0,84	0,87	0,85	0,84

Note: pixel = pixel-based approach; OB = object-oriented approach.

Source: The authors (2023).

**Table 4** – Mean Kappa and F1 and Standard Deviation

Classifier	Mean Kappa	Standard Deviation	Mean F1	Standard Deviation
SVM OB	0,89	0,03	0,91	0,03
SVM Pixel	0,86	0,03	0,88	0,02
KNN OB	0,85	0,03	0,87	0,02
KNN Pixel	0,84	0,03	0,87	0,02
RF OB	0,83	0,03	0,84	0,03
RF Pixel	0,81	0,06	0,84	0,06
SAM	0,76	0,05	0,80	0,04

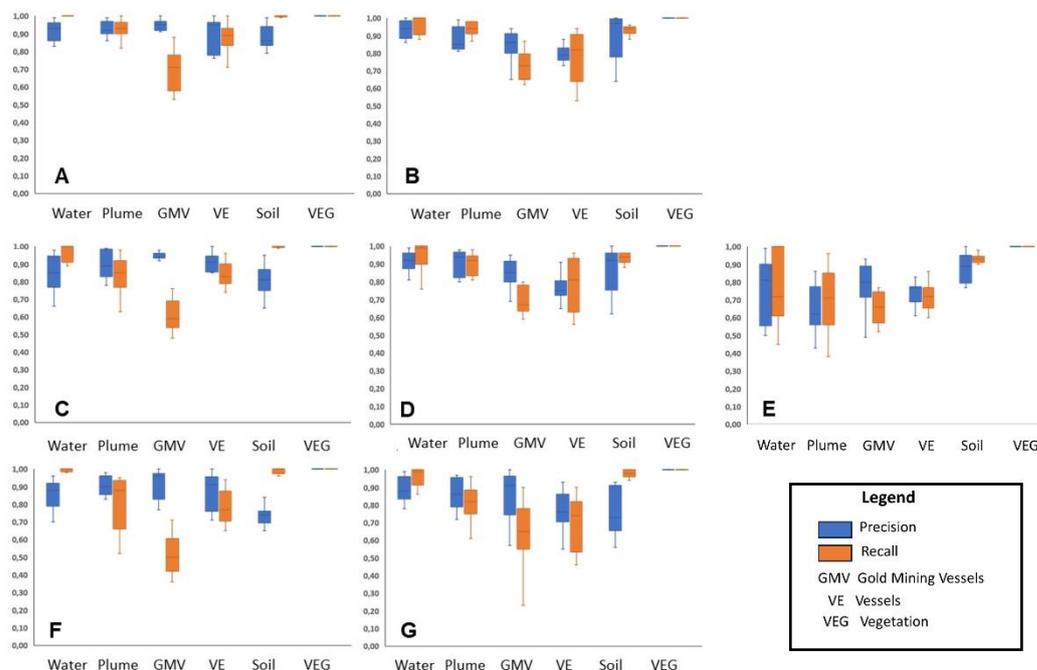
Source: The authors (2023).

Precision and recall values were more varied. Precision ranged from 0.43 to 1 and recall from 0.2 to 1 (Figure 8). Some classes were easily classified, like vegetation and water when there were no sediment plumes. In these cases, precision and recall were equal to 1. However, the classifiers presented difficulty in separating classes with close spectral responses, such as vessels and gold mining vessels. In these cases, precision ranged from 0.77 to 0.88 (gold mining vessels) and from 0.78 to 0.89 (vessels). Recall ranged from 0.58 to 0.71 (gold mining vessels) and 0.69 to 0.87 (vessels), which means that

there was a greater occurrence of false negatives.

This behavior is common when analyzing different types of vessels classification methods because the spectral response of the vessels does not show a clear pattern. Sometimes they are similar, sometimes they are very different, affecting the performance of the methods (Heiselberg, 2016). Furthermore, the small size of the vessels affects the methods. Heiselberg and Heiselberg (2017) identified a drop down in the performance factor for smaller vessels from 0.93 to 0.68 because of omissions and confusion with other small targets.

Figure 8 - Precision and Recall for Classifications: a) SVM OB; b) SVM Pixel; c) KNN OB; d) KNN Pixel; e) SAM; f) RF OB; g) RF Pixel.



Source: The authors (2023).

A similar behavior was noticed for gold mining vessels and sandbanks with content of water. It is justified by the low and close spectral response of these targets and the reduced size of

some sandbanks that present a similar response to small non-water targets (Kikaki *et al.*, 2022).

By analyzing the mean F1 scores it is possible to confirm the better behavior of the

SVM OB. Besides achieving a better F1, SBM OB presented greater constancy, with a lower quantity of false negatives and false positives. SVM OB was followed by the SVM Pixel (Table 5). KNN and RF were inferior, and RF presented greater inconstancy, mainly regarding the occurrence of false negatives. SAM obtained the worst results, with a greater level of confusion, mainly between gold mining vessels and vessels, targets with similar spectral response (Figure 9).

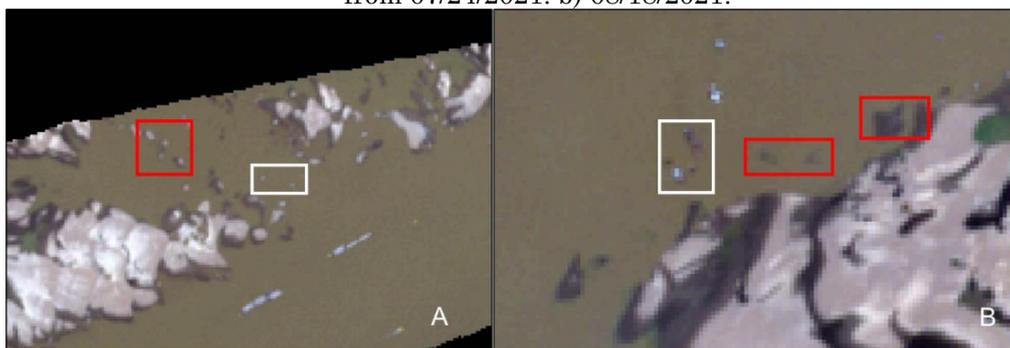
Still, SAM presented difficulty in classifying water and sediment plumes. In some cases, the confusion was so strong that it could not even delimitate the sediment plumes. This occurred due to the low variation in the concentration of OAC, since different types of water are better classified when the difference is more pronounced (Barbosa, 2005; Vantrepotte *et al.*, 2012).

**Table 5** - Precision, Recall and F1 Score of the Classifications.

Classes	SVM Pixel		SVM OB		KNN Pixel		KNN OB		RF Pixel		RF OB		SAM	
	Mean Precision	Mean Recall												
Water	0,88	0,96	0,90	0,99	0,92	0,95	0,85	0,96	0,89	0,95	0,88	0,91	0,75	0,78
Sediment Plume	0,88	0,94	0,91	0,94	0,90	0,87	0,90	0,88	0,82	0,81	0,82	0,82	0,65	0,70
Gold Mining Vessels	0,84	0,72	0,88	0,72	0,84	0,69	0,86	0,59	0,77	0,63	0,78	0,64	0,78	0,65
Vessels	0,80	0,78	0,89	0,88	0,76	0,79	0,77	0,83	0,78	0,69	0,81	0,72	0,73	0,72
Soil	0,89	0,93	0,87	0,99	0,87	0,93	0,84	1,00	0,86	0,96	0,85	0,93	0,88	0,93
Vegetation	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
F1 Water	0,92		0,94		0,93		0,90		0,92		0,89		0,77	
F1 Sediment Plume	0,91		0,92		0,88		0,89		0,82		0,82		0,68	
F1 Gold Mining Vessels	0,78		0,79		0,76		0,70		0,69		0,70		0,71	
F1 Vessels	0,79		0,89		0,77		0,80		0,73		0,76		0,72	
F1 Soil	0,91		0,93		0,90		0,91		0,91		0,89		0,90	
F1 Vegetation Score	1,00		1,00		1,00		1,00		1,00		1,00		1,00	
Mean F1	0,88		0,91		0,87		0,87		0,84		0,84		0,80	

Source: The authors (2023).

Figure 9 - Confusion between Small Vessels and Gold Mining Vessels with Other Targets. In Red Small Sandbanks with Water Content. In White Small Vessels and Gold Mining Vessels a) Image from 07/24/2021. b) 08/18/2021.



Source: The authors (ano).

When comparing pixel-based versus object-oriented approaches, it is clear the advantage for SVM OB that improved all validation metrics (Table 5). Keshtkar *et al.* (2017) and Noi and Kappas (2017) achieved similar results, since SVM presented the best metrics for identifying different types of water. Also, the performance improved when using an object-oriented approach (Hartoni *et al.*, 2022). The results for KNN and RF were varied. Although the mean F1 scores were very close, precision and recall were more random, improving or worsening the

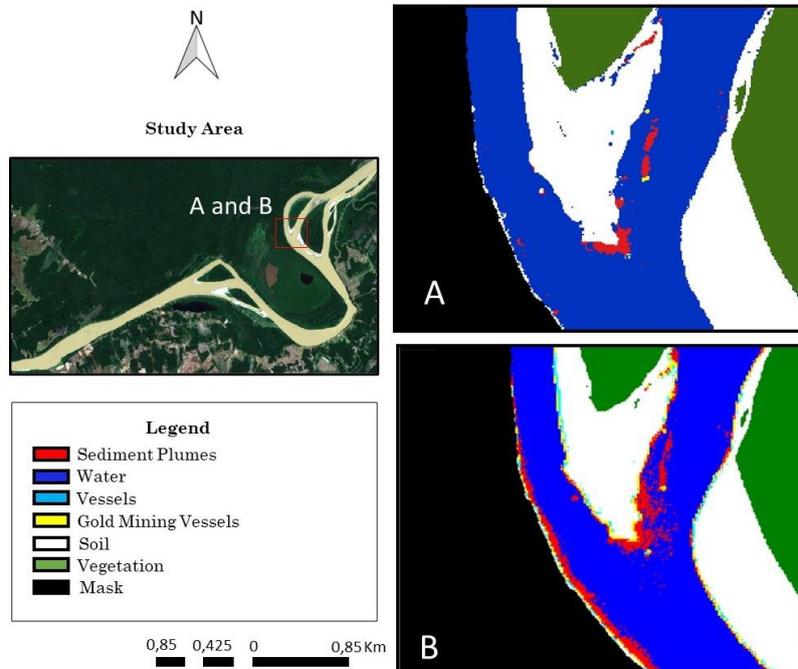
Kappa and F1 depending on the class. Besides, they still proved to be more effective than SAM. The pixel-based approach had many commission errors between vessels and gold mining vessels. It generated classifications with systemic errors (Figure 10). However, all object-oriented approach classifications significantly attenuated the systemic errors, producing cleaner classifications. It happened because the ability to distinguish between classes is improved when evaluating segments with a

certain spectral and spatial similarity and not just pixels (Simionato *et al.*, 2021).

Since SVM OB presented the best results, its classifications were used to evaluate the detection and omission percentage of the operating gold mining vessels. The detection percentage of gold mining vessels without

sediment plumes was 100%. While the detection percentage of operating gold mining vessels with sediment plumes ranged from 73 to 100%. The omission percentage ranged from 0 to 26.6%, failing to detect 0 to 4 operating gold mining vessels with sediment plumes per image (Table 6).

Figure 10 - Improvements in SVM Classifier in Image from 08/13/2021: a) SVM OB; b) SVM Pixel.



Source: The authors (2023).

**Table 6-** Identification of Operating Gold Mining Vessels Using SVM OB Classifier.

Data	Operating Gold Mining Vessels (Total)	Operating Gold Mining Vessels Detected	% of Detection	Operating Gold Mining Vessels Omitted	% of Omission
06/30/2018	18	17	94,44	1	5,56%
07/25/2018	6	6	100	0	0,00%
08/19/2018	9	8	88,89	1	11,11%
06/20/2019	15	11	73,33	4	26,67%
07/09/2020	21	17	80,95	4	19,05%
09/22/2020	8	6	75	2	25,00%
06/14/2021	5	4	80	1	20,00%
07/14/2021	3	3	100	0	0,00%
07/24/2021	9	8	88,89	1	11,11%
Média	--	--	86,83	2	13,17

Source: The authors (2023).

Kanjir (2019) observed similar results with detection percentage ranging from 63 to 96% and omission percentage from 4 to 34%. Heiselberg and Heiselberg (2017) also obtained similar results, since the detection percentage ranged from 67 to 100% and omission percentage from 0 to 33%. In all cases the worst performances were caused by omissions of small vessels.

In the present study, the omissions of operating gold mining vessels did not occur due to the non-detection of the target itself, but because it was not possible to identify the origin of the sediment plumes on the vessel.

In the case of gold mining vessels without sediment plumes, visual assessment is extremely necessary to confirm that the targets are in fact gold mining vessels. By visual

evaluation, it is possible to identify the typical behavior of gold mining vessels groupings together (Figure 11).

Figure 11 – Grouped Gold Mining Vessel without Sediment Plume in Image of 07/29/2021.



Source: The authors (2023).

## CONCLUSION

The results showed that the SVM classifier presented the best evaluation metrics. When comparing the pixel-based versus object-oriented approach, SVM object-oriented results were noticeably better, achieving the highest F1 score in the study (0.91). 100% of the gold mining vessels without sediment plumes were detected, as well as most of the operating gold mining vessels. The omissions occurred because it was not possible to identify the origin of the sediment plumes on the vessel. Therefore, the methodology proved to be effective to detect gold mining vessels using machine learning classifiers in Sentinel-2 images.

The performance of the classifiers is influenced by several factors, such as the characteristics of the targets and the quantity and quality of the training samples. Therefore, the choice of classifier depends on the objectives sought, since a method that obtained satisfactory performance may have significantly altered results when applied to another context. The limited number of training samples and the spectral similarity of certain targets is a sensitive point for the methodology employed.

## REFERENCES

- ADAMY, A. Dinâmica fluvial do Rio Madeira. In: SILVA, R.G.C. (org.). *Porto Velho: cultura, natureza e território*. p. 120–147, 2016.
- ASNER, G. P.; LLACTAYO, W.; TUPAYACHI, R.; LUNA, E.R. Elevated rates of gold mining in the Amazon revealed through high-resolution monitoring. *Proceedings of the National Academy of Sciences*, v. 110, n. 46, p. 18454–18459, 2013. <https://doi.org/10.1073/pnas.1318271110>
- ARCGIS PRO. Segmentation. Available: <https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/segmentation.htm>. Accessed on: apr. 15, 2023.
- BALZINO, M.; SECCATORE, J.; MARIN, T.; DE TOMI, G. Gold losses and mercury recovery in artisanal gold mining on the Madeira River, Brazil. *Journal of Cleaner Production*, v. 102, p. 370–377, 2015. <https://doi.org/10.1016/j.jclepro.2015.05.012>
- BARBOSA, C. C. F. *Sensoriamento Remoto da Dinâmica da Circulação da Água do Sistema Planície de Curuai/Rio Amazonas*. 2005. São José dos Campos (SP). 281 p. (Tese de Doutorado, Instituto Nacional de Pesquisas Espaciais).
- BARBOSA, C. C. F.; NOVO, E. M. L. M.; MARTINS, V. S. *Introdução ao Sensoriamento Remoto de Sistemas Aquáticos: Princípios e Aplicações*. São José dos Campos. Instituto Nacional de Pesquisas Espaciais. 2019.
- BARBOSA, F. L. R.; GUIMARÃES, R.F.; CARVALO JÚNIOR, O.A.; GOMES, R.A.T. Classificação do uso e cobertura da terra utilizando imagens SAR/Sentinel 1 no Distrito Federal. *Sociedade & Natureza*, v. 32, p. e55954, 2 fev. 2021. <https://doi.org/10.14393/SN-v33-2021-55954>
- BEZERRA, R. B.; DANTAS, T. R.; TRINDADE, A. G. Caracterização Temporal da Precipitação Pluvial no Município de Porto Velho/RO no Período de 1945 a 2003. *Sociedade & Natureza*, v. 22, n. 3, p. 609–623, 2010. <https://doi.org/10.1590/S1982-45132010000300015>
- BRASIL. Lei No 7.805, de 18 de julho de 1989. Available: <https://presrepublica.jusbrasil.com.br/legislacao/103411/lei-7805-89/>. Accessed on: aug. 02, 2022.
- BRASIL. Catálogo de Metadados. 2022. Available: <https://metadados.snirh.gov.br/geonetwork/srv/por/catalog.search#/metadata/7d054e5a-8cc9-403c-9f1a-085fd933610c>. Accessed: sep. 04, 2022.

- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. <https://doi.org/10.1023/A:1010933404324>
- CAO, R.; TU, W.; YANG, C.; LI, Q.; LIU, J.; ZHU, J.; ZHANG, Q.; LI, Q.; QIU, G. Deep learning-based remote and social sensing data fusion for urban region function recognition. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 163, n. February, p. 82–97, 2020. <https://doi.org/10.1016/j.isprsjprs.2020.02.014>
- CETEM. **Inventário Nacional de Emissões e Liberações de Mercúrio pelos Garimpos de Ouro**. 2018. Available: <https://diretoriopre.mma.gov.br>. Accessed on: jul. 29, 2022.
- CIOCARLAN, A.; STOIAN, A. Ship detection in sentinel 2 multi-spectral images with self-supervised learning. **Remote Sensing**, v. 13, n. 21, 1 nov. 2021. <https://doi.org/10.3390/rs13214255>
- COMANICIU, D.; MEER, P. Mean shift: A robust approach toward feature space analysis. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 24, n. 5, p. 603–619, 2002. <https://doi.org/10.1109/34.1000236>
- DECHESNE, C.; LEFÈVRE, S.; VADAIME, R.; HAJDUCH, G.; FABLET, R. Ship identification and characterization in Sentinel-1 SAR images with multi-task deep learning. **Remote Sensing**, v. 11, n. 24, p. 1–18, 2019. <https://doi.org/10.3390/rs11242997>
- ESA. **Sentinel-2 User Handbook**. 2015. Available: [https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2\\_User\\_Handbook.pdf/8869acdf-fd84-43ec-ae8c-3e80a436a16c?t=1438278087000](https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook.pdf/8869acdf-fd84-43ec-ae8c-3e80a436a16c?t=1438278087000). Accessed on: aug. 03, 2022.
- ESPINOZA VILLAR, R.; MARTINEZ, J.M.; TEXIER, M.; GUYOT, J.L.; FRAIZY, P.; MENESES, P.R.; OLIVEIRA, E. A study of sediment transport in the Madeira River, Brazil, using MODIS remote-sensing images. **Journal of South American Earth Sciences**, v. 44, p. 45–54, 2013. <https://doi.org/10.1016/j.jsames.2012.11.006>
- FERNANDES, F. R. C.; ALAMINO, R. C. J.; ARAUJO, E. Recursos minerais e comunidade: impactos humanos, socioambientais e econômicos. Rio de Janeiro. CETEM. 2014.
- FIX, E; HODGES; J. L. JR. Discriminatory analysis, nonparametric discrimination. **USAF School of Aviation Medicine**, Randolph Field, Tex., Project 21-49-004, Rept. 4, AF41(128)-31, February 1951.
- GOODIN, D. G.; HAN, L.; FRASER, R.N.; RUNDQUIST, D.C.; STEBBINS, W.A.; SCHALLES, J.F. Analysis of suspended solids in water using remotely sensed high resolution derivative spectra. **Photogrammetric Engineering & Remote Sensing**, v. 54, n. 4, p. 505–510, 1993.
- HARTONI, H.; SIREGAR, V.P.; WOUTHUYZEN, S.; AGUS, S.B. Object based classification of benthic habitat using Sentinel 2 imagery by applying with support vector machine and random forest algorithms in shallow waters of Kepulauan Seribu, Indonesia. **Biodiversitas**, v. 23, n. 1, p. 514–520, 1 jan. 2022. <https://doi.org/10.13057/biodiv/d230155>
- HEISELBERG, H. A direct and fast methodology for ship recognition in sentinel-2 multispectral imagery. **Remote Sensing**, v. 8, n. 12, 1 dez. 2016. <https://doi.org/10.3390/rs8121033>
- HEISELBERG, P.; HEISELBERG, H. Ship-iceberg discrimination in Sentinel-2 multispectral imagery by supervised classification. **Remote Sensing**, v. 9, n. 11, 1 nov. 2017. <https://doi.org/10.3390/rs9111156>
- ISIDRO, C. M.; MCINTYRE, N.; LECHER, A.M.; CALLOW, I. Applicability of earth observation for identifying small-scale mining footprints in a wet tropical region. **Remote Sensing**, v.9, n.9, p. 945, 12 sep. 2017. <https://doi.org/10.3390/rs9090945>
- KANJIR, U. Detecting migrant vessels in the Mediterranean Sea: Using Sentinel-2 images to aid humanitarian actions. **Acta Astronautica**, v. 155, p. 45–50, 1 fev. 2019. <https://doi.org/10.1016/j.actaastro.2018.11.012>
- KESHTKAR, H.; VOIGT, W.; ALIZADEH, E. Land-cover classification and analysis of change using machine-learning classifiers and multi-temporal remote sensing imagery. **Arabian Journal of Geosciences**, v. 10, n. 6, 1 mar. 2017. <https://doi.org/10.1007/s12517-017-2899-y>
- KIKAKI, K.; KAKOGEORGIOU, I.; MIKELI, P.; RAITOS, D.; KARANTZALOS, K. MARIDA: A benchmark for Marine Debris detection from Sentinel-2 remote sensing data. **PLoS ONE**, v. 17, n. 1 January, 1 jan. 2022. <https://doi.org/10.1371/journal.pone.0262247>
- KRUSE, F. A.; LEFKOFF, A.B.; BOARDMAN, J.B.; HEIDEBRECHT, K.B.; SHAPIRO, A.T.; BARLOON, P.J.; GOETZ, A.F.H. The Spectral Image Processing System (SIPS) - Interactive Visualization and Analysis of Imaging spectrometer Data. **Remote Sensing of Environment**, v. 44, p. 145–163, 1993. [https://doi.org/10.1016/0034-4257\(93\)90013-N](https://doi.org/10.1016/0034-4257(93)90013-N)
- LODI, L. A; FRANZINI, M; CASELLA, V. Mapping Land Cover Types using Sentinel-2 Imagery: A Case Study. In: **GISTAM**. 2019. p. 242-249.
- LOBO, F. de L.; NOVO, E.M.L.M.; BARBOSA, C.C.F.; GALVÃO, L.S. Reference spectra to classify Amazon water types. **International Journal of Remote Sensing**, v. 33, n. 11, p. 3422–3442, 2012. <https://doi.org/10.1080/01431161.2011.627391>
- LOBO, F. de L.; SOUZA-FILHO, P.W.M.; NOVO, E.M.L.D.M.; CARLOS, F.M.; BARBOSA, C.C.F. Mapping Mining Areas in the Brazilian Amazon Using MSI/Sentinel-2 Imagery (2017). **Remote Sensing**, v. 10, n. 8, p. 1178, 25 jul. 2018. <https://doi.org/10.3390/rs10081178>
- LOUIS, J.; DEBAECKER, V.; PLUG, B.; MAINKORN, M.; BIENIARZ, J.; MUELLER-WILM, U. CADAU, E.; GASCON, F. SENTINEL-2 SEN2COR: L2A PROCESSOR FOR USERS. **Proceeding living planet symposium**, p. 1–8, 2016.
- MARTINS, V. S.; NOVO, E.M.L.D.M.; LYAPUSTIN, A.; ARAGÃO, L.E.O.C.; FREITAS, S.R.; BARBOSA, C.C.F. Seasonal and interannual assessment of cloud cover and atmospheric constituents across the Amazon (2000–2015): Insights for remote sensing and climate analysis. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 145, p. 309–327, 1 nov. 2018. <https://doi.org/10.1016/j.isprsjprs.2018.05.013>

- NOI, P. T.; KAPPAS, M. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. **Sensors (Basel, Switzerland)**, v. 18, n. 1, 22 dez. 2017. <https://doi.org/10.3390/s18010018>
- REDEMAIS/MJSP; © 2020 PLANET. PROGRAMA BRASIL MAIS, 2020. Available: <https://plataforma-pf.scon.com.br>. Accessed on: jun. 22, 2022.
- RODRIGUEZ, D. A.; LOPES, L.G.; CARIELLO, F.; LÁZARO, J. JÚNIOR, S.; PINTO, G.L. Previsões de cheias extremas nos horizontes sazonais e de curto e médio prazos na bacia do rio Madeira: estudo de caso da enchente de 30 de março de 2014 em Porto Velho. **Revista Brasileira de Cartografia**, v. 69, n. 4, p. 827–836, 2017. <https://doi.org/10.14393/rbcv69n4-44337>
- SALGADO, C. B.; CARVALHO JÚNIOR, O.A.; GOMES, R.A.T.; GUIMARÃES, R.F. Análise da interferência de nuvens na classificação de séries temporais MODIS-NDVI na região da Amazônia, município de Capixaba, Acre. **Sociedade & Natureza**, v. 31, p. 1–20, 2019. <https://doi.org/10.14393/SN-v31-2019-47062>
- SIGMINE. **Sistema de Informações Geográficas da Mineração**. 2022. Available: <https://geo.anm.gov.br>. Accessed on: out. 17, 2022.
- SIMIONATO, J.; BERTANI, G.; OSAKO, L. S. Identification of artisanal mining sites in the Amazon Rainforest using Geographic Object-Based Image Analysis (GEOBIA) and Data Mining techniques. **Remote Sensing Applications: Society and Environment**, v. 24, p.100633, 2021. <https://doi.org/10.1016/j.rsase.2021.100633>
- TRINDADE, R. B. E.; BARBOSA FILHO, O. **Extração de ouro - princípios, tecnologia e meio ambiente**. CETEM/MCT & PUC-RJ, 2002.
- VANTREPOTTE, V.; LOISEL, H.; DESSAILLY, H.; MÉRIAUX, X. Optical classification of contrasted coastal waters. **Remote Sensing of Environment**, v. 123, p. 306–323, ago. 2012. <https://doi.org/10.1016/j.rse.2012.03.004>
- VAPNIK, V. N. **The nature of Statistical learning theory**. New York. Springer-Verlag. 1995. <https://doi.org/10.1007/978-1-4757-2440-0>

## AUTHOR CONTRIBUTION

Diego Henrique Costa Pereira conceived the study, collected, processed and analyzed the information and wrote the text. Roberto Arnaldo Trancoso Gomes helped with the design of the study, guided the work from the initial stages, as well as helped with the analysis of information and writing of the text. Osmar Abílio de Carvalho Júnior assisted in analyzing the information and writing the text. Renato Fontes Guimarães assisted in analyzing the information and writing the text.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.