*Article*

# Identifying Precarious Settlements and Urban Fabric Typologies Based on GEOBIA and Data Mining in Brazilian Amazon Cities

Bruno Dias dos Santos [1,2,*], Carolina Moutinho Duque de Pinho [2], Gilberto Eidi Teramoto Oliveira [1], Thales Sehn Korting [1], Maria Isabel Sobral Escada [1] and Silvana Amaral [1]

[1] Earth Observation and Geoinformatics Division, National Institute for Space Research (INPE), São José dos Campos 12227-010, Brazil; gilberto.oliveira@inpe.br (G.E.T.O.); thales.korting@inpe.br (T.S.K.); Isabel.escada@inpe.br (M.I.S.E.); silvana.amaral@inpe.br (S.A.)

[2] Center for Engineering, Modeling and Applied Social Sciences (CECS), Federal University of ABC (UFABC), Santo André 09210-580, Brazil; Carolina.pinho@ufabc.edu.br

[*] Correspondence: bruno.santos@inpe.br

**Abstract:** Although 70% of the Amazon population lives in urban areas, studies on the urban Amazon are scarce. Much of the urban Amazon population lives in precarious settlements. The distinctiveness and diversity of Amazonian precarious settlements are vast and must be identified to be considered in the development of appropriate public policies. Aiming at investigating precarious settlements in Amazon, this study is guided by the following questions: For the Brazilian Amazon region, is it possible to identify areas of precarious settlements by combining geoprocessing and remote sensing techniques? Are there different typologies of precarious settlements distinguishable by their spatial arrangements? Thus, we developed a methodology for identifying precarious settlements and subsequently classifying them into urban fabric typologies (UFT), choosing the cities of Altamira, Cametá, and Marabá as study sites. Our classification model utilized geographic objects-based image analysis (GEOBIA) and data mining of spectral data from WPM sensor images from the CBERS-4A satellite, jointly with texture metrics, context metrics, biophysical index, voluntary geographical information, and neighborhood relationships. With the C5.0 decision tree algorithm we carried out variable selection and classification of these geographic objects. Our estimated models show accuracy above 90% when applied to the study sites. Additionally, we described Amazonian UFT in six types to be identified. We concluded that Amazonian precarious settlements are morphologically diverse, with an urban fabric different from those commonly found in Brazilian metropolitan areas. Identifying and characterizing distinct precarious areas is vital for the planning and development of sustainable and effective public policies for the urban Amazon.

**Keywords:** Amazonian precarious settlements; Amazonian urbanization; GEOBIA; data mining; urban fabric typology

## 1. Introduction

According to estimates from the United Nations Human Settlements Program (UN-Habitat) [1], one in eight people in the world lived in slums in 2016. In developing countries, the number of urban residents living in slums has increased by 28% between 1990 and 2014, from 689 million to 881 million residents, representing 30% of the urban population of developing countries.

According to UN-Habitat estimates [2], the number of slum dwellers surpassed the 1 billion mark in 2018. The UN-Habitat declares that the growth trend in the number of slum dwellers is now increased by direct and indirect effects of the COVID-19 pandemic and admits that the achievement of the 11th Sustainable Development Goal - *"Make cities and human settlements inclusive, safe, resilient and sustainable"* by 2030 is unlikely.

In Brazil, slums are part of a larger category of inadequate housing—known as precarious settlements—a definition established in 2004 by the National Housing Policy [3]. The precarious settlements are segments of the urban territory of varying sizes and types, predominantly residential areas occupied by low-income populations, characterized by numerous shortages of public services and inadequate housing [3,4].

Dwellers of these settlements use multiple strategies to solve their own housing needs, as neither the State nor the formal market meets their demands [4]. This lack of public support results in low-quality self-constructed housing and, consequently, in irregular appropriation of areas in the cities [5].

These settlements encompass numerous typologies, especially tenements, slums, informal subdivisions, low-income irregular subdivisions, and degraded housing complexes [6]. Among these typologies, tenements and slums are the two types of settlements most studied in the literature [7–10].

Despite their relevance, the definition and characterization of precarious settlements established by the National Housing Policy [3] may not be sufficient to contemplate the particulars of most Brazilian municipalities, especially for periurban municipalities in the Amazon region [11–13]. The concept of "periurban" is similar to the definition of "peripheral urbanization pattern" developed by Erminia Maricato, Raquel Rolnik, Nabil Bonduki, and others during the 1970s and 1980s [12,14]. This term can also designate areas of urban–rural transitions that may or may not be in the outskirts of large cities [12,14].

To develop effective public housing policies for the Amazonian municipalities, there is an increased need to identify features of precarious settlements considering their regional, municipal, and urban diversities [13]. Understanding the characteristics of deprived areas in Amazonian cities is paramount to reveal social, economic, and regional inequalities. This study of identification, mapping, and characterization of precarious settlements in the Amazon allows the development of appropriate tools to support urban and housing policy in the region.

Although about 70% of the Brazilian population lives in urban centers, studies on Amazonian urbanization are still limited [15]. The highest relative percentage of households located in Brazilian precarious areas is in the metropolitan area of Belém, corresponding to more than 51% of households [15]. About 54% of Belém's Metropolitan area population resides in precarious areas [16,17]—that is, within the Legal Amazon. Furthermore, studies that use remote sensing and geoprocessing techniques to identify and characterize precarious settlements are restricted to the metropolitan areas of the southeast region of Brazil, such as the metropolitan area of São Paulo and the metropolitan area of Rio de Janeiro [18], with environmental and socioeconomic characteristics profoundly different from the Amazonian regions.

Commonly, the identification and characterization of precarious settlements nationwide uses techniques for extracting useful information from census data [17,19–21]. Mahabir et al. [22] listed some limitations in the use of census data to identify slums: (a) the intensive, time-consuming, and costly data collection; (b) the time between two surveys (10 year interval in Brazil); (c) the levels of aggregation of census data may restrain the capture of relevant heterogeneity present in slums [23,24]; (d) the fear that residents of irregular areas have to provide information and perhaps face the misuse of these data, facilitating evictions by public authorities [25–27]; (e) the use of census data to support the development of public policies is reduced with poor data collection quality [28].

Given the limitations of exclusively using data from the census, methodologies that include remote sensing data have become desired for large-scale studies [22]. The arising of free and widely available high spatial resolution images (<5 m) favored novel applications in studies of the urban fabric, replacing previous practices developed using less refined imagery [29]. Now it is possible to overhaul previous analysis at the household level and improve assessment of settlements, enabling a better understanding of the formation, status, and growth of slums [18,22].

Considering the above context, this work poses the following questions: In the Brazilian Amazon region, is it possible to identify areas of precarious settlements using geoprocessing and remote sensing techniques? Are there different typologies of precarious settlements distinguishable by their spatial arrangements?

To answer these questions, we first developed and implemented a methodology for identifying precarious areas in Amazonian municipalities. Then, we proposed the classification of the identified settlements, considering specific urban fabric typologies (UFT). We applied this methodology in the municipalities of Altamira, Cametá, and Marabá—cities located in the state of Pará, in the Brazilian Legal Amazon. This work proposes the identification of precarious areas using techniques of geographic object-based image analysis (GEOBIA) and data mining, and later classification by the C5.0 decision tree algorithm [30]. This methodology was entirely built with open source software and open access data.

This article has the following structure: Section 2 shows the importance of identifying precarious settlements in Brazilian cities by geotechnologies; Section 3 details the main concepts that guided the identification of these settlements; Section 4 proposes a novel methodology for identifying and characterizing precarious areas; Section 5 shows our methodology and accuracy assessment; Section 6 has the discussion of main results; and, finally, Section 7 shows final considerations.

## 2. Identification of Brazilian Precarious Settlements by Geotechnologies

The Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística, IBGE*) addresses informal housing through a category of census tracts called subnormal agglomerates. According to IBGE [17], subnormal agglomerates are illegally occupied areas, consisting of at least 51 housing units, with or without essential public services: water, sewage, waste and utilities, and urbanization patterns in discordance with current guidelines. Usually, subnormal agglomerates have narrow circulation routes and irregular house alignments (lots with unequal sizes and shapes), and construction not regulated by public bodies. This definition applies to several types of precarious settlements in Brazil, such as slums, occupations, floodplains, wetlands, villages, communities, huts, stilts, among others.

The 1990 and 2000 censuses held the same definition of subnormal agglomerates. The 2010 census relied on high spatial resolution images and on the support of municipal offices responsible for housing policies to improve mapping of subnormal agglomerates - the IBGE technical report [17] lacks to inform the type of sensor used in the census mapping, despite mentioning the use of high spatial resolution satellite images. According to 2010 census results, a total of 11.4 million Brazilians lived in 15,868 subnormal tracts—equivalent to 5% of all census tracts in Brazil [17].

However, some authors criticize the criteria used in the definition of subnormal agglomerates [20,31] and indicate that the minimum occupancy of 51 households excludes smaller arrangements, causing an underestimation of these tracts [16]. Fortunately, the limit of 51 households was not part of the guidelines for the delimitation of subnormal agglomerates in 2019 [32]. There are also criticisms regarding the requirement of land property analysis, which poses challenges in precarious areas; the derogatory nature of the term subnormal; and the use of census tracts as a territorial base [24]. Despite criticisms, IBGE have not attempted to distinguish precarious settlements from subnormal agglomerates or adapt census tracts to better represent different types of abnormal housing arrangements.

Subnormal agglomerates are the most similar regions to study precarious settlements at census tracts level, as defined by IBGE, and it is the only source of data in these areas at this level of detail nationwide [24]. Therefore, we use subnormal agglomerates [32] as samples of precarious settlements, even though we are following the definition of precarious settlements as established by the National Housing Policy [3].

Research developed by the Center for Metropolis Studies (CEM) shows that the 2000 and 2010 censuses underestimated housing precariousness numbers [20,21]. Using discriminant analysis techniques in census data, the CEM identified new subnormal agglomerates

tracts not classified as such by IBGE. Analyzing data from the 2010 Census [17], CEM and CEBRAPE [21] demonstrated an underestimation of approximately 30% of precarious settlements in the metropolitan area of São Paulo. Despite advances, the CEM data also have limitations, as the territorial database continues to be formed by census tracts, which does not allow for their disaggregation [24].

Kurkdjian [33] published the first study that identified Brazilian precarious settlements using remote sensing data, although their main goal was to estimate the size of urban populations. The authors identified slums in São José dos Campos, São Paulo as one of the homogeneous residential zones (ZRH). Homogeneous residential zones are areas with similar residential patterns identified by having similar texture in remote sensing data. In Brazil, previous reports [34,35] defined a methodology for identify ZRH and estimate the population size of São José dos Campos, São Paulo. The ZRH identification was based on visual interpretation of aerial photographs recorded in 1978 by an RC-10 camera, installed on the Bandeirante aircraft of the National Institute for Space Research (INPE), and Double-X Kodak panchromatic film, obtained at 1:10,000 scale. The authors also used a light magnifying glass with a scale and a stereoscope to perform sectoring of municipalities and estimate population sizes.

Studies with procedures for sectoring of Brazilian precarious areas by remote sensing were only published again in 2004, but still in the context of population size estimates through ZRH [36]. Matias and Nascimento [37] proposed the first study whose goal was to identify precarious settlements by mapping informal settlements in Ponta Grossa, Paraná, using a panchromatic image of the Ikonos-II sensor.

Hofmann [38] proposed one of the first studies using GEOBIA for slums identification. For a region of Cape Town, South Africa, with Ikonos images, he generated geographic objects after multiresolution segmentation at different spatial scales, which were hierarchically linked. Precarious settlements were super objects that had several features regarding their shape and context, identified by fuzzy analysis techniques. Subsequently, Hoffman carried out studies for mapping slums in Rio de Janeiro [39]. The author developed a spatial ontology technique for identifying slums in high spatial resolution images. Further, he translated the spatial ontology model into an identification framework that uses GEOBIA techniques. After this study, several others have investigated the potential of ontology to map slums [40–46]. The study by Hofmann et al. [39] also inspired Kohli et al. [47] in the development of an ontology applied to slums that concentrates the physical characteristics of these inadequate housing arrangements, being a reference for the identification of precarious areas by remote sensing techniques [18].

More recently, Feitosa et al. [48] identified precarious settlements in Baixada Santista metropolitan area, São Paulo, using the IMMerSe methodology (integrated methodology for mapping and classifying precarious settlements). The IMMerSe Project identified precarious areas and classified them into urban fabric typologies. This study was conducted in collaboration between academia and government with the goal of measuring housing deficit and inadequacy in the Baixada Santista metropolitan area. Despite not employing digital image processing, IMMerSe extracted information from high spatial resolution remote sensing images, such as density and urban organization level, to characterize the urban environment.

In addition to the studies mentioned above, Costa et al. [49] and [50–54] developed maps of elements that compose the intraurban land cover (concrete, different types of roofs, water, and vegetation, among others). The authors applied the GEOBIA approach to classify the land cover of Brazilian cities using images with extremely high spatial resolution (<1 m). These studies greatly contributed to the use of remote sensing applications in the identification of precarious settlements, by allowing the classification of intraurban land cover given high spatial resolution images.

## 3. Precarious Settlements in the Brazilian Amazon

Cardoso et al. [55] described the morphological patterns of informal settlements, among other morphological typologies, in six cities in the State of Pará: Marabá, Parauapebas, Canaã dos Carajás, São Felix do Xingu (all in Southeast Pará), Santarém (Lower Amazonas, Southeast Pará), and Altamira (Southwest Pará).

Traditional Amazonian settlements have an organic layout and are modeled accordingly with the natural site, with a few urban voids, and a strong connection between the built-up areas and open spaces [56]. The open areas in traditional spaces function as house extensions, interweaving the urban economy with everyday life. On the contrary, informal settlements can be understood as reduced reinterpretations of these traditional settlements [55] and are divided into *spontaneous* informal and *organized* informal settlements.

The early transformation of rural areas with agrarian fragility systems in urban areas results in organized informal settlements, following the same patterns of territorial rationalization that the private sector pursues [55]. Both organized and spontaneous informal settlements are subject to poverty by reducing the size of residential lots, which restricts the use of land and hinders food self-sufficiency of low-income dwellers [55].

Some informal settlements are in lowlands [55] where seasonal floods pose a greater challenge to an informal settlement resident than to a riverside dweller [56]. Unlike in informal areas, in the riverside floods may not pose an environmental risk to inhabitants and buildings are designed to tolerate floods when they occur [56].

The migration of *caboclo* peasants (caboclo is the designation given in Brazil to those born from the miscegenation between Brazilian indigenous and white people) to urban wetlands was one of the main factors that led to the formation of precarious settlements in Amazonian cities. Population density was gradually increasing in wetlands and these areas transformed into informal settlements [55]. The destruction of livelihoods and exogenous control of the means of production, either through private appropriation of natural resources or through the control of the means of production [55] intensifies the migration from the countryside to the city. Another fact that explains the formation of precarious settlements is the irregular occupation of public or private land designated for "rural use" during periods of migratory flows to the Amazon region, especially by low-income population.

## 4. Materials and Methods

The study area has 622.50 km$^2$ and encompasses the municipal seats of Altamira (152.74 km$^2$), Cametá (44.18 km$^2$), and Marabá (425.47 km$^2$), located, respectively, in the Southwest, Northeast, and Southeast of the State of Pará, in the Brazilian Legal Amazon (Figure 1). We created a rectangular boundary around the study areas, overlaying it with urban census tracts of the municipal seats [19]. These municipalities were chosen because they have distinct patterns of urban occupation [55,57] and are part of a research study from the Laboratory for Investigation of Socioenvironmental Systems [58].

### 4.1. Preprocessing

In this work, we analyzed images from the CBERS-4A satellite. The CBERS (China–Brazil Earth Resources Satellite) is a series of satellites utilized in the partnership between Brazil and China for studies in technical–scientific space matters. The CBERS Program marked Brazil's entrance into space technology and allowing the collection of primary remote sensing data [59,60].

The CBERS-4A satellite is the fifth satellite of the Sino–Brazilian program, launched in December 2019, providing images from across the globe to researchers in several fields. In Brazil, many programs to control deforestation in the Legal Amazon use images from the CBERS series. The CBERS-4A is a satellite equipped with optical payloads operating in visible spectrum with spatial resolution in the range of 2 to 60 m [59,60].
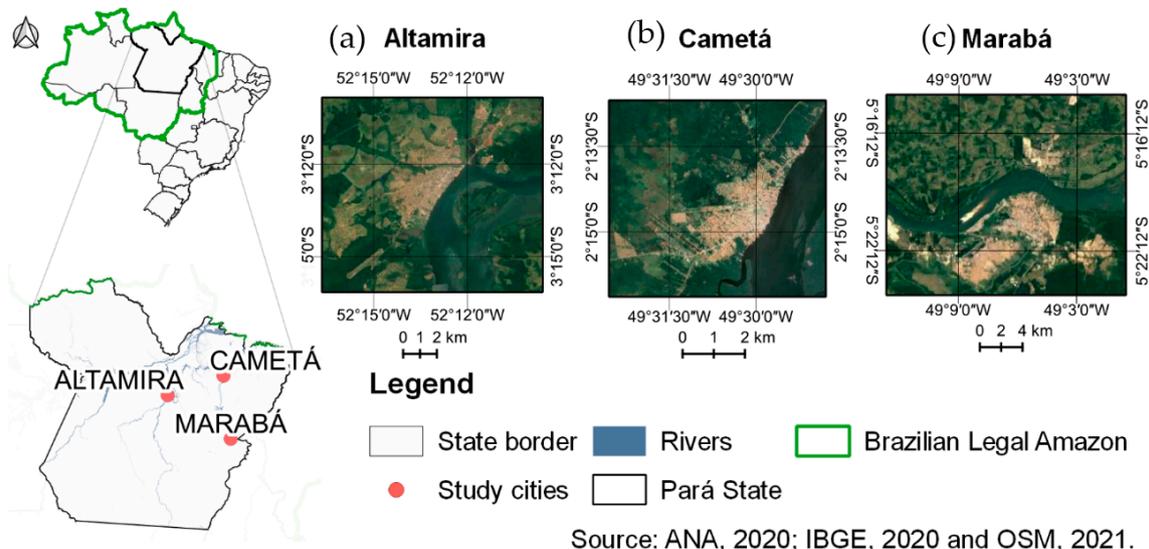
**Figure 1.** Study areas or municipalities: (**a**) Altamira; (**b**) Cametá; and (**c**) Marabá.

The WPM sensor provides panchromatic and multispectral images simultaneously. Panchromatic images have 2 meters of spatial resolution, with a spectral range between 0.45 and 0.90 μm. Multispectral images have 8 meters of spatial resolution with the spectral bands: blue (0.45–0.52 μm), green (0.52–0.59 μm), red (0.63–0.69 μm), and near-infrared (NIR) (0.77–0.89 μm). The radiometric resolution of the images is 10 bits. The imaged swath width is 92 km, and the revisit period is 31 days [59]. INPE makes the CBERS series imagery freely available in its catalog [61].

We used orthorectified images from the WPM sensor of the CBERS-4A satellite. For Altamira, the chosen image is from 31 January 2020 (path: 215, row: 117), for Cametá, the image is from 15 September 2020 (path: 212, row: 116). For Marabá, the image is from 20/08/2020 (path: 211, row: 120). The following software was used:

- QGIS [62]: for image clipping, PCA fusions, image registration, calculation of bio-physical index and textural index, creation of hexagon grids, and production of thematic maps;
- Orfeo ToolBox [63]: segmentation of WPM images;
- Geographical Data Mining Analyst (GeoDMA) [64]: for feature extractions and estimation of the classification model by applying the C5.0 algorithm.

First, we cropped the images to match the boundaries of the study areas. Then, we proceeded with a fusion of the panchromatic image with the multispectral images by using the principal component analysis (PCA) method. After the fusion, we registered the images using a third-degree polynomial algorithm with cubic resampling, having Google Earth images as a reference. The mean error was less than 3 pixels for each of the three images.

After preprocessing the images, we calculated the following indices:

(a) **Normalized difference vegetation index (NDVI)**: division of the near-infrared (NIR) band by the red band, normalized, to analyze the presence and condition of vegetation [65] (Equation (1)):

$$\text{NDVI} = \frac{(\text{NIR} - \text{RED})}{(\text{NIR} + \text{RED})} \quad (1)$$

(b) **Normalized difference roof index (NDRI)**: division of the red band by the blue band, normalized, to identify the presence of ceramic roofs and areas with exposed soil [66] (Equation (2)):

$$NDRI = \frac{(RED - BLUE)}{(RED + BLUE)} \tag{2}$$

(C) **Bare soil area index (BAI)**: the normalized division of the blue band by the near-infrared band, to identify exposed soil [46] (Equation (3)):

$$BAI = \frac{(BLUE - NIR)}{(BLUE + NIR)} \tag{3}$$

We also computed texture metrics from the gray level co-occurrence matrix (GLCM) by Haralick [67], GLCM variance, entropy, contrast, textural correlation, uniformity, and measures of correlation (MOC) among features using a $25 \times 25$ window size for all directions. After preprocessing steps (Figure 2), the database was composed of three layers corresponding to multispectral bands, NDVI, NDRI, BAi, and layers of textural metrics (variance, entropy, contrast, correlation, uniformity, and MOC).
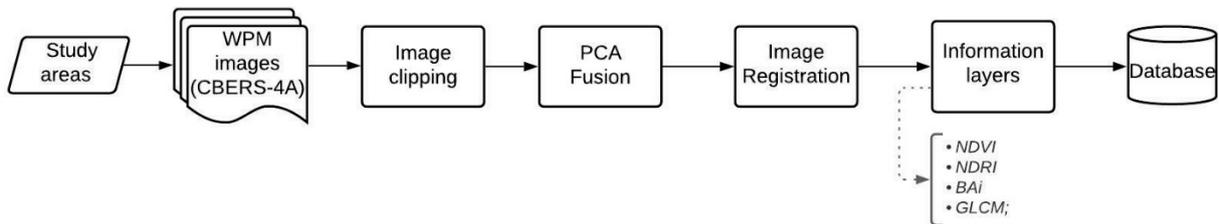


**Figure 2.** Preprocessing steps.

The urban fabric typologies (UFT) identification process took place after the preprocessing stage and the generation of informational layers (NDVI, NDRI, BAi, and GLCMs) from WPM images. This methodology can be summarized (Figure 3) in three main steps: (1) creation of the land cover base and generation of the first level of segmentation; (2) identification of precarious areas; (3) identification of typologies of precarious areas (UFT). After identifying precarious areas, the databases with images of the three municipalities were appended to carry out their typologies identification.
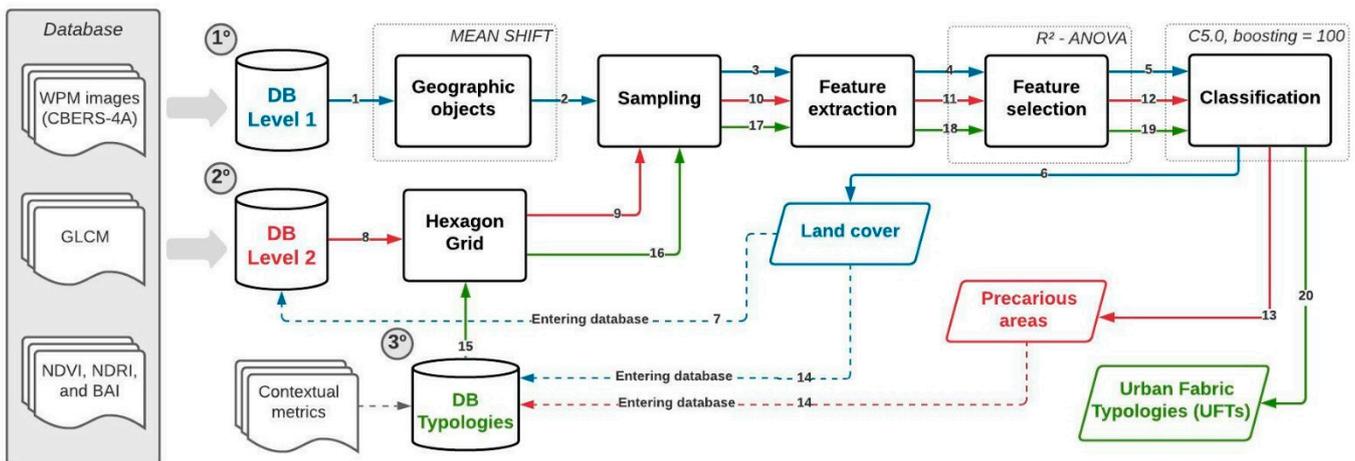


**Figure 3.** Methodology for identification of precarious areas and classification of urban fabric typologies (UFT). Arrow numbering indicates the order of steps taken.

### 4.2. Preparation of the Land Cover Database

The segmentation of WPM images using the mean shift cluster algorithm [68] resulted in the first level of information, represented by a land cover base. The mean shift algorithm first defines a window around each point and averages out the gray tones of the image. Then, it changes the center of the windows for each centroid and repeats the procedure

until it iteratively converges. As a result, the algorithm segments the image into several pixel clusters.

The mean shift cluster algorithm is available in the Orfeo Toolbox plugin (OTB) [63] and, during initial tests it performed better relative to processing time and segmentation of objects compared to the other four algorithms available. Chosen values of spatial radius, range radius, maximum number of interactions, and minimum region size control the size and shape of segments. In this case, we adopted: spatial radius of 5 pixels, range radius of 100 pixels, maximum number of interactions of 100 times, and minimum region size of 15 pixels.

We defined classes of objects by size and shape as follows: fiber cement roofing, ceramic roofing, high gloss roofing, exposed soil, asphalt, forest vegetation, herbaceous vegetation, dirt road, clouds, shadow, and water. About 2255 samples of segments from Altamira, 1607 from Cametá, and 2424 from Marabá. We performed random and stratified sampling, following the natural prevalence of classes.

Afterward, we extracted features from the spectral layers; the GLCMs layers; the biophysical index layers; and spatial metrics of objects—totalizing 121 features for each segment. We extracted spectral and shape features [69] of objects using GeoDMA software [64]. Using ANOVA's $R^2$, the features were ranked according to their explanatory strength to distinguish classes in the sample. We eliminated features with $R^2$ below 0.1 or with absolute linear correlation above 0.75 with other variables. After this stage, 24 variables remained for Altamira, 24 for Marabá, and 26 for Cametá. We classified all segments using the machine learning algorithm C5.0 [30], an improved version of the C4.5 algorithm [70,71].

The C5.0 algorithm has a two-step classification process [70,71]. The first step is training and consists of building a set number of decision trees in the training samples. The second step is classification, consisting of classifying the segments in the testing samples by applying the decision tree built in the previous step. In this algorithm, an entropy metric called information gain determines the choice of variables so that the features with greater information gains are chosen to be kept in the final decision tree. Each segment receives the class with the most votes from the decision trees. To run C5.0 in the GeoDMA, the following parameters are required:

- *Boosting value*: number of trees. We adopted a value of 100 trees for boosting.
- *Minimum cases*: the minimum number of samples required to split an internal node. We chose a minimum value of 2.
- *Training samples proportion*: the proportion of samples labeled for training. We adopted 70% of the labeled samples for training and 30% for validation.

### 4.3. Identification of Precarious Areas

We built a second level of information, hierarchically superior to the first, represented by a grid of regular hexagons with 100 m diagonal and 8661 m$^2$ area. Hexagons are spatial units of analysis used as a reference for merging data from different sources, such as environmental characterization data and data collected from remote sensing. We adopted a hexagonal spatial unit because it allows for better understanding of spatial phenomena through its neighborhood structure and cell connectivity [72,73]. We defined the dimensions of the hexagon grid based on studies that identified precarious settlements in Brazilian municipalities using cell grids [24,45,48] and considered 100 m the length of a regular urban block in Brazil. Furthermore, this cell size allows for capturing neighborhood differences without compromising internal spatial arrangements within hexagons (Figure 4).

As previously performed for the first level, we extracted features from spectral layers, GLCMs, and biophysical index. We also extracted landscape metrics for the land cover classes present in the first level, totalizing 344 features for each hexagon.
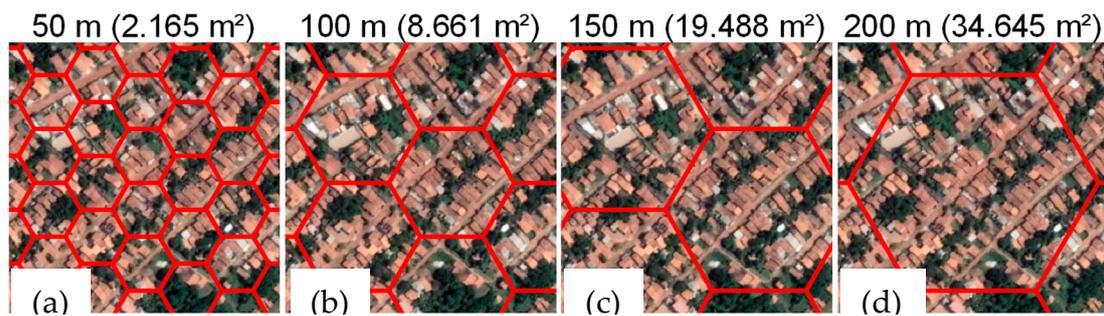
**Figure 4.** Comparison of hexagon sizes given the length of their main diagonal in meters: (**a**) 50; (**b**) 100; (**c**) 150; and (**d**) 200.

Further, we created spatial variables from neighborhood connections between hexagons. Since $X_1, X_2, X_3, X_4, X_5,$ and $X_6$ are neighbors and adjacent hexagons of a Y hexagon, for each variable "a" among the 344 existing (features) variables, we estimated the mean value ($MV_a$) of the neighbors for variable "a", the normalized difference ($DV_a$) between cell Y and the mean of its neighborhood, and the normalized global difference ($DG_a$) between Y and the mean of the variable "a" given a grid of hexagons (Equations (4)–(6)).
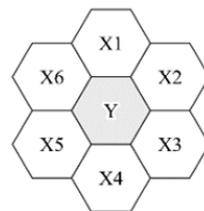


$$MV_a = \frac{\sum_{i=1}^{6} (X_a)_i}{6} \tag{4}$$

$$DV_a = \frac{(Y_a - MV_a)}{(Y_a + MV_a)} \tag{5}$$

$$DG_a = \frac{(Y_a - MG_a)}{(Y_a + MG_a)} \tag{6}$$

For the second level, we performed random and stratified sampling of precarious and nonprecarious settlements. In total, we selected 3427 samples for Altamira, 1229 for Cametá, and 7082 for Marabá. For Altamira and Marabá, we adopted the IBGE subnormal agglomerates as the reference or original data of precarious settlements to be identified [32]. For Cametá, we adopted references from a research study conducted by Sakatauska [13]. These data for precarious settlements needed preprocessing procedures. We redefined the boundaries of settlements when they had uninhabited portions, removing large areas of vegetation and roads. This cleaning process was necessary especially in areas classified as subnormal agglomerates [32] since these data come from a category of census tracts and may not necessarily represent the mapping of precarious settlements.

Subsequently, we eliminated variables with low explanatory strength ($R^2 < 0.1$) and with high linear correlation (> 0.75 absolute) with others. Thus, 26 variables remained for Altamira, 55 for Cametá, and 73 for Marabá. We classified the second level of data by applying the C5.0 algorithm with a boosting value of 100. The algorithm classified precarious and nonprecarious areas. We set 70% of the samples for training and the remaining 30% for validation.

### 4.4. Characterization and Identification of Urban Fabric Typologies (UFT) in Precarious Areas

With the identification of precarious areas, the next step consisted of the characterization of urban fabric typologies (UFT) based on visual analysis of satellite images. An urban fabric can have a variety of forms and dimensions of urban elements such as buildings, roads, public spaces, lots, and other infrastructure, among other components, and

a specific relationship among these elements and with the geographic base that supports them [74,75].

Our theoretical reference comes from Cardoso et al. [55] and Sakatauskas [13] to understand relevant physical characteristics of precarious settlements in the cities of Altamira, Cametá, and Marabá. We considered the following elements to identify fabric typologies: (a) the dimensions and shapes of urban elements; (b) built-up density and the presence or absence of unoccupied areas within settlements; (c) presence, shape, importance, and condition of pathways; (d) aspects of organized or spontaneous occupation, such as new construction and block shapes; and (e) existence of designated residential areas or guidelines for housing—data extracted from satellite images. We divided precarious areas into six different urban fabrics (UFT). Table 1 shows key features to identify urban fabric typologies.

**Table 1.** Key features to identify urban fabric typologies (UFT) in the areas of study.

| UFT | Full and Empty Scheme | WPM RGB (3,2,1) | Class Description |
|---|---|---|---|
| UFT1 Fibre cement | | | Organized occupation with planned blocks and structured roads. The roads have sections with poor quality asphalt and sections of unpaved ground. High built-up density, without unbuilt areas inside settlements or space within lots. Fiber cement is the predominant material for roofs. |
| UFT2 Ceramics | | | Organized occupation with planned blocks and structured roads. The roads have sections with poor quality asphalt and sections of unpaved ground. High built-up density, without unbuilt areas inside settlements or space within lots. The predominance of ceramic roof is what distinguishes UFT2 from UFT1. |
| UFT3 Vegetated | | | Organized occupation with planned blocks and structured roads. The roads are unpaved and longer than 400 m. Medium built-up density, with unbuilt areas and vegetation inside blocks. Ceramic is the predominant material for roofs. |
| UFT4 Sparse | | | Spontaneous and not consolidated occupation. Low built-up density, without well-defined blocks and with unpaved streets. There is a predominance of empty spaces between dwellings, with herbaceous vegetation. Recent occupations resulting from unmediated expansion of organized precarious settlements toward environmentally sensitive areas or land designated for rural purposes. |
| UFT5 Roads | | | Occupation located near highway BR-230 (*Transamazônica*) in Marabá and Altamira, BR-222 in Marabá, and BR-422 in Cametá. Organized settlements located far from central areas, in regions with periurban characteristics. Roads are unpaved (except highways). |
| UFT6 Wetlands | | | Spontaneous occupation, located in wetlands or in areas subject to flooding. Low built-up density and similar to UFT4, some are the result of the expansion of large settlements. It may not have well-defined street layouts and has shrubby vegetation between houses. |

Schemes of full and empty spaces—with "full" symbolizing constructions and "empty" symbolizing streets, squares, green areas, and other open spaces—are a way of representing these two morphological elements that cannot be seen or analyzed individually, as they are interdependent [76]. Additionally, full and empty schemes may represent the porosity of urban fabrics (related to the circulation of winds) and built-up density.

To identify typologies of precarious settlements, we created a database with land cover landscape metrics (level 1), classification of precarious areas (level 2), and other context metrics (Table 2). We generated context metrics using data from volunteered geographic information (VGI). We used the OpenStreetMap [77] road infrastructure and hydrography database to obtain the *"Euclidean distance between the nearest road and the centroid of a cell"* and the *"Euclidean distance between a municipal seat and the centroid of a cell"*. In addition, we computed neighborhood metrics, totalizing 984 variables for each cell at the end of the feature extraction process.

**Table 2.** Reasons for the choice of the following context metrics.

| Context Metric | Reason |
| --- | --- |
| Euclidean distance between the nearest road and the centroid of a cell | Roads are vectors of urban growth in Amazonian cities [55,57]. Additionally, they are related to the formation of precarious settlements in their surrounding areas [13]. |
| Euclidean distance between the municipal seat and the centroid of a cell | The central region of Amazonian cities in this study concentrates health and educational services and has better infrastructure and offer of basic services [13] than other regions. Some precarious settlements are far from the center of the municipality, as they occupy regions close to roads and areas in recent expansion. |
| Euclidean distance between the centroid of a cell and the height above the nearest drainage (HAND) | Some informal settlements are in flatlands [55]. However, unlike traditional Amazonian populations, here housing is not well designed to survive seasonal floods. Furthermore, Sakatauskas [13] identified a typology of precarious settlements located in wetlands. We built HAND according to the methodology developed by Rennó et al. [78]. |
| Average of the shape index of blocks | Informal settlements tend to have greater inconsistency of block designs when compared to traditional Amazonian settlements [57]. This value allows the comparison of geometric shapes of blocks in regular areas (rectangular blocks with similar sizes) and irregular shapes or the absence of subdivisions within blocks, a common characteristic of precarious settlements [75]. We computed shape index [79] from the road base [77] accordingly to the methodology described by [75]. |

Performing random and stratified sampling, we selected 41 samples for UFT1—Fiber Cement, 113 for UFT2—Ceramics, 53 for UFT3—Vegetated, 198 for UFT4—Sparse, 157 for UFT5—Roads, and 113 for UFT6—Wetlands. Finally, we eliminated variables with low explanatory strength ($R^2 < 0.2$) and high linear correlation (>0.75 absolute) with others, keeping only 28 variables for each cell. The identification of typologies was also performed by the C5.0 decision tree algorithm, with a boosting value of 100. We set 70% of samples for training and the remaining 30% for validation.

### 4.5. Validation

We built a decision tree based on the training sample (70%) and evaluated the classification of segments based on the validation sample (30%). We assessed accuracy through two confusion matrices: one to evaluate the classification of precarious areas and another to evaluate UFT classification.

## 5. Results

By qualitatively analyzing the classification of precarious areas by our methodology (Figure 5), we notice that, in general, they overlap with polygons of real precarious settlements used as references (polygons shaded black). Despite this, there are commission errors in all cities where nonprecarious areas were classified as precarious by our methodology.
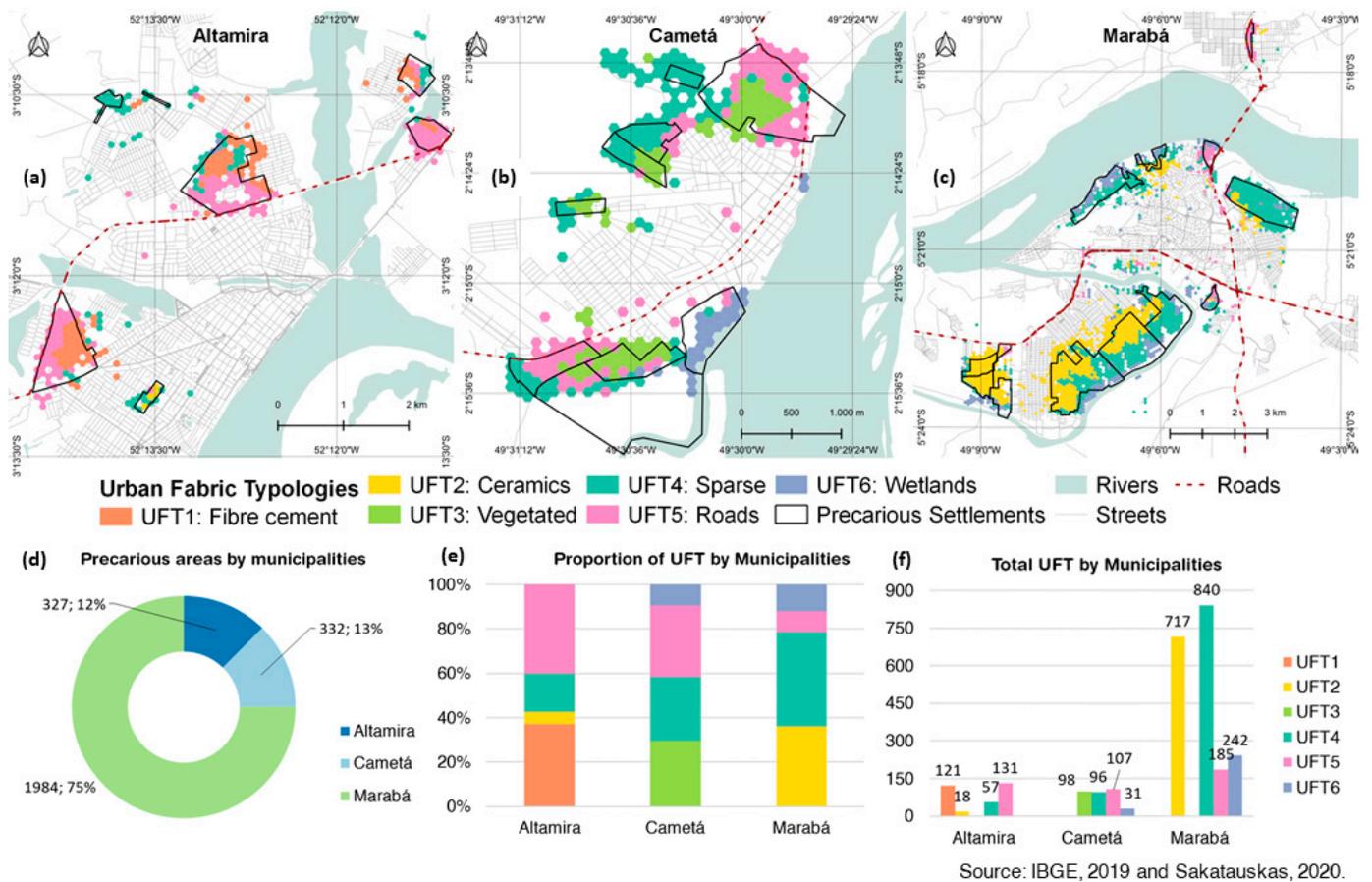
**Figure 5.** Urban fabric typologies (UFT) classified by the proposed methodology for (**a**) Altamira, (**b**) Cametá, and (**c**) Marabá. Graphs of (**d**) location of identified precarious areas, (**e**) the proportion of UFT per municipality, and (**f**) the total number of UFT per municipality.

Statistically, the classification of precarious areas resulted in the following global accuracy levels (GA) and F1-score (F1): 0.97 (GA) and 0.78 (F1) for Altamira; 0.94 (GA) and 0.81 (F1) for Cametá; and 0.97 (GA) and 0.91 (F1) for Marabá. Table 3 shows the outcome confusion matrix from our classification of precarious areas, computed in the validation samples. Among the three cities, Marabá had the highest F1-score value. About 40% of Marabá's area of study is uninhabited vegetated areas, which reduces the region size to be classified and mitigates classification errors. The large size of uninhabited areas in Marabá may be one of the reasons for high accuracy in classifying nonprecarious areas.

**Table 3.** Confusion matrix of the precarious area validation samples.

| | | Classification | | | Producer's Accuracy | User's Accuracy | Area |
|---|---|---|---|---|---|---|---|
| | | Nonprec. | Prec | Total | | | |
| **Reference** | Nonprec. | 1054 | 11 | 1065 | 0.99 | 0.98 | Altamira (F1: 0.78; GA: 0.97) |
| | Prec | 26 | 65 | 91 | 0.71 | 0.86 | |
| | Total | 1080 | 76 | 1156 | | | |
| | Nonprec. | 350 | 15 | 365 | 0.96 | 0.96 | Cametá (F1: 0.81; GA: 0.93) |
| | Prec | 14 | 62 | 76 | 0.82 | 0.81 | |
| | Total | 364 | 77 | 441 | | | |
| | Nonprec. | 1887 | 52 | 1939 | 0.97 | 0.98 | Marabá (F1: 0.91; GA: 0.97) |
| | Prec | 31 | 444 | 475 | 0.93 | 0.90 | |
| | Total | 1918 | 496 | 2414 | | | |

Marabá, the city with the largest area (425.47 km$^2$), concentrates most of the identified precarious areas (75% of the total) (Figure 5d). However, Cametá has the highest percentage of urban areas classified as precarious, corresponding to 30% (Figure 6). According to our classification, Altamira has 4% of urban areas classified as precarious. Our classification resulted in a similar number of urban areas identified as precarious in Cametá and Altamira when compared to the original data [32]. There was an increase in the urban area classified as precarious in Marabá. When compared with IBGE data [32], the percentage of precarious urban areas in Marabá changed from 16% to 23%, an increase of 46%.
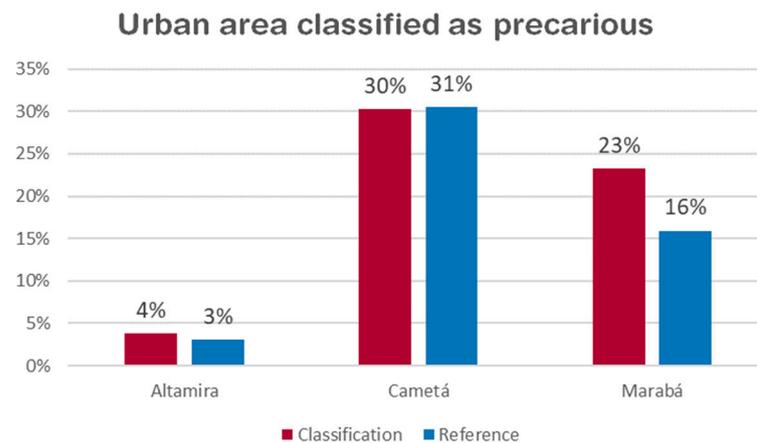


**Figure 6.** Comparison of urban areas classified as precarious by our decision tree classification and original data.

Concerning the typologies (Figure 5), only Altamira has UFT1—Fiber cement, representing 5% of the cells analyzed. UFT2—Ceramics is predominant in Marabá, but we can also find it in Altamira, which corresponds to 28% of the total. UFT3—Vegetated is exclusive to Cametá and is the least frequent typology (only 4% of the total). Most of the precarious areas identified are UFT4—Spaced type, accounting for 38%, and despite the predominance in Marabá, it is present in all cities. Similar to UFT4, UFT5—Roads is also present in all cities, corresponding to 16% of the total. Finally, UFT6—Wetlands is present in Marabá and Cametá, representing 10% of the total.

Regarding the identification of UFT, the variable selection process resulted in 28 features, all of them neighborhood features. The final decision trees have eleven features in common, with nine related to landscape metrics and the other two related to context metrics. Using $p$ the perimeter (in meters), $A$ the patch area (in m$^2$), and $n$ the number of patches, we show the 11 most relevant features found, starting with nine related to neighborhood mean (MV):

- **LSI Asphalt:** The landscape shape index (LSI) of the "Asphalt" class, which measures the complexity of shape by dividing the perimeter by twice the square root of the area multiplied by π (Equation (7)). This index quantifies the amount of edge present in a landscape relative to what would be present in a landscape of the same size but with a simple geometric shape (circle in vector, square in raster) and no internal edge. Equation (7) shows the LSI calculation:

$$\text{LSI} = \frac{\sum_{j=1}^{n} p_j}{2\sqrt{\pi A}} \tag{7}$$

- **PD Exposed Soil** and **PD Roads:** The patch density (PD) of the "Exposed Soil" and "Roads" classes, which represents the number of patches divided by the total patch area. Equation (8) shows the PD calculation:

$$\text{PD} = \frac{n}{A} \tag{8}$$

- *MSI Herbaceous Vegetation*: The mean shape index (MSI) of the "Herbaceous Vegetation" class. Equation (9) shows the MSI calculation:

$$\text{MSI} = \frac{\frac{\sum_{j=1}^{n} p_j}{2\sqrt{\pi A_j}}}{n} \tag{9}$$

- *NP Herbaceous Vegetation:* The number of patches (NP) of "Herbaceous Vegetation" class;
- *AWMPFD Herbaceous Vegetation*: The area-weighted average fractal dimension (AWMPFD) of "Herbaceous Vegetation" class; AWMPFD equals the average patch fractal dimension of patches of the corresponding patch type, weighted by patch area so that larger patches weigh more than smaller patches [79]. Equation (10) shows the AWMSI calculation:

$$\text{AWMPFD} = \sum_{j=1}^{n} \left[ \frac{2.\ \ln(p_j)}{ln(A_j)} \cdot \frac{A_j}{\sum_{i=1}^{n} A_{ij}} \right] \tag{10}$$

- *IJI Shrub Vegetation:* The interspersion and juxtaposition index (IJI) of the "Shrub Vegetation" class, which measures whether a particular patch type is adjacent to one or more classes. Equation (11) shows the IJI, with *m* being the perimeter of all edge segments involving the corresponding patch type:

$$\text{IJI} = \frac{-\sum_{j=1}^{n} \left( \frac{p_j}{\sum_{k=1}^{n} p_k} \right) \cdot\ ln\left( \frac{p_j}{\sum_{k=1}^{n} p_k} \right)}{\ln(m-1)} \cdot 100 \tag{11}$$

- *TABO Water* and *TABO Ceramic Roof:* Total area of the largest object (TABO) of the "Water" and "Ceramic Roof" classes that cross the landscape.

The two features from global difference (DG):

- Euclidean distance from a cell's centroid to the nearest river;
- Euclidean distance from a cell's centroid to the nearest road.

Figure 7 exemplifies three precarious original settlements and their UFT classifications. For *Loteamento Bela Vista*, an irregular allotment in Altamira (Figure 7a,b), the hexagons classified as UFT5—Roads are those located in the vicinity of the Transamazônica Highway, BR-230. Additionally, hexagons classified as UFT1—Fiber cement are located further away from the area of influence of highways. The same general pattern was observed for the *Invasão do Castanhal* settlement, in Cametá (Figure 7c,d), where the UFT5—Roads class hexagons are located near the Transcametá Highway, BR-422.

In the case of the Precarious Settlements *Bairro da Paz*, *Bairro Filadelfia*, and *São Miguel da Conquista*, in Marabá (Figure 7e,f), UFT2—Ceramics refers to a consolidated area, previously occupied. As the occupation advances toward the Itacaiúnas River, the level of consolidation decreases. Consolidated settlements are urbanized settlements that have already been integrated with the city and have basic infrastructure. Physical interventions that alter their morphology are not necessary (segmenting lots, redefining the road system), nor the implementation of basic infrastructure (such as water and sewage services). However, these settlements may be informal, and their population may have specific demands postoccupation, and needs for social programs, and other public services [80]. In this case, there is a transition from an organized occupied urban fabric to a spontaneous occupied urban fabric, motivated by unmediated precarious settlement growth.

Table 4 shows a confusion matrix from the UFT classification, computed from validation samples. The global accuracy and the average F1-score were above 0.91, which demonstrates the methodology's relevant potential.
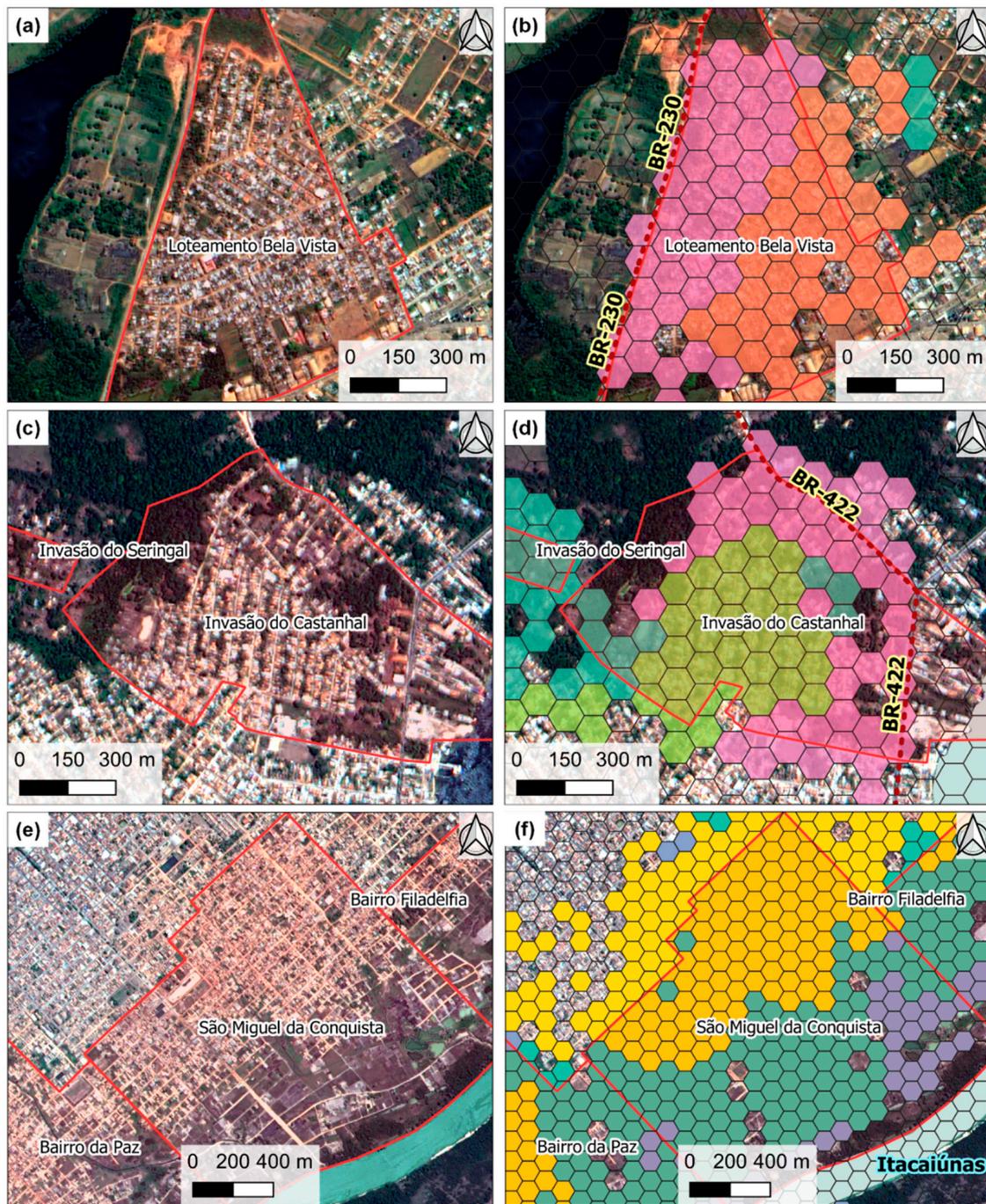
**Figure 7.** UFT classification of three precarious settlements: Loteamento Bela Vista, Altamira, in WPM true color image (**a**) and classification in UFT (**b**); Invasão do Castanhal, Cametá, in WPM true color image (**c**) and classification in UFT (**d**); Bairro da Paz, São Miguel da Conquista and Bairro Filadelfia, Marabá, in WPM true color image (**e**) and classification in UFT (**f**).

**Table 4.** Confusion matrix of the UFT validation samples.

| | | Classification | | | | | | | Producer's Accuracy | User's Accuracy | Class F1-Score | Global | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | UFT1 | UFT2 | UFT3 | UFT4 | UFT5 | UFT6 | Total | | | | | |
| **Reference** | UFT1 | 15 | 1 | | | 1 | | 17 | 0.88 | 0.83 | 0.86 | F1-Score | 0.91 |
| | UFT2 | | 32 | | | | 1 | 33 | 0.97 | 0.91 | 0.94 | Accuracy | 0.92 |
| | UFT3 | | | 19 | 1 | 1 | | 21 | 0.90 | 0.90 | 0.90 | | |
| | UFT4 | | | | 58 | | 1 | 59 | 0.98 | 0.92 | 0.95 | | |
| | UFT5 | 3 | 2 | 2 | 2 | 47 | 2 | 58 | 0.81 | 0.96 | 0.88 | | |
| | UFT6 | | | | 2 | | 34 | 36 | 0.94 | 0.89 | 0.92 | | |
| | Total | 18 | 35 | 21 | 63 | 49 | 38 | 224 | | | | | |

We obtained producer accuracy and F1-score levels above 0.9 for almost all classes except UFT1 and UFT5. The UFT1 also had the lowest user accuracy among the typologies. The smaller number of samples and the confusion with UFT5 explain the lower performance of our methodology for the UFT1 class. The UFT5 is a typology that includes areas near roads and does not necessarily have a specific morphological pattern, which increases the probability of classification errors. Although UFT5 does not have a well-defined morphological pattern, we identified it as one of our six categories because roads are vectors of urban growth in Amazonian cities. Roads replaced the role that rivers played in the formation of the first Amazonian urban centers [55], and they are also related to the formation of precarious settlements in the surrounding areas [13].

The UFT4—Sparse had the best classification performance according to F1-score. In addition to its very peculiar morphological pattern, which is easily distinguished from the other typologies, UFT4 had the largest number of samples.

## 6. Discussion

The identification of precarious settlements in Altamira and Cametá resulted in similar prevalence of precarious areas to the original data. Only in Marabá the area identified as precarious had greater prevalence than in the reference data. Therefore, there are some misclassifications. In all these cities, subnormal agglomerates have nonresidential areas. When we corrected the delimitation of subnormal agglomerates, removing areas of vegetation and nonresidential areas, our model underestimated the original prevalence of subnormal agglomerates in the IBGE data [32] (Figure 8).
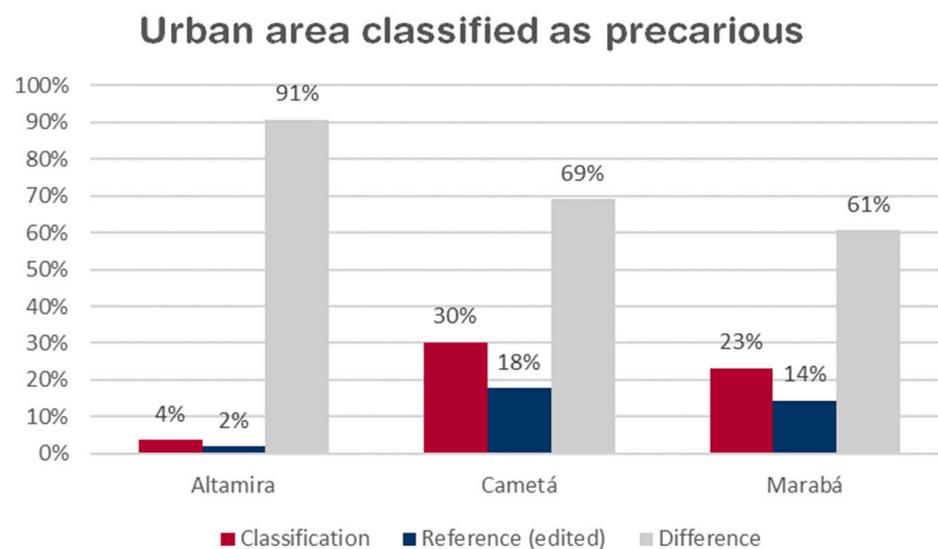


**Figure 8.** Comparison of urban areas classified as precarious by our methodology and as identified in the original (reference) data by IBGE.

Although Altamira has a low percentage of urban areas identified as precarious, our classification almost doubled the total number of precarious areas in the city when compared with the IBGE data [32]. In Cametá and Marabá, our classification overestimated in 69% and 61% the number of precarious areas in the reference data. Therefore, our work agrees with several studies that report an underestimation of subnormal agglomerates [16,20,21] by IBGE. Furthermore, our study shows that the underestimation of subnormal clusters by IBGE is not restricted to metropolitan areas, also occurring in Altamira, Cametá, and Marabá.

The characterization of urban fabric typologies reveals a diversity of morphological patterns existing in precarious settlements in the Amazonian region. This general characterization of urban fabric typologies demonstrates the need to build specific interpretation keys for Amazonian cities. Models based on generic samples neglect to identify these housing arrangements, especially for those located outside the metropolitan areas of Belém and Manaus.

Note that, none of the municipalities have all six UFT in its territory. However, the UFT4—Sparse and UFT5—Roads are present in all three cities. Slums and tenements are the two most studied precarious settlements in the Brazilian literature. However, we have not found slums and tenements in any of the three cities, reinforcing the need for more diverse studies on the singularities of precarious settlements in nonmetropolitan Amazonian cities.

In the Brazilian case, prior characterization of settlements is a mandatory step for local governments to access funding from the federal government's slum upgrading programs [4]. Furthermore, this type of characterization allows the understanding of internal structures in precarious areas to provide meaningful indicators to combat poverty through evidence-based policy-making toward more sustainable cities [81].

Feitosa et al. [48] (by IMMerSE methodology) and Georganos et al. [81] also identified morphological patterns in slums and other types of precarious settlements. Implementing the methodology in Nairobi, Kenya, Georganos et al. [81] assembled multiple WorldView-3 datasets with varying numbers of bands and image features, allowing for the exploration of computational efficiency, complexity, and scalability. The authors used a land cover classification generated by the GEOBIA approach to identify typologies of deprived urban areas (DUAs). The classification of DUAs typologies was performed by unsupervised k-means clustering. Similarly, both studies [48,81] identified settlements with different levels of consolidation, organization of the urban fabric, and built-up density. Note that these studies also used grids as spatial reference units.

The input variables in our final decision trees for UFT classification have theoretical relevance in the field. Figure 9 shows a boxplot of the "LSI Asphalt" and "IJI Shrub Vegetation" landscape metrics, and the contextual metrics of river and road distances. The "LSI Asphalt" has higher prevalence in UFT1—Fiber cement and UFT2—Ceramics, which are the two most consolidated types of settlement, with better conditions of road infrastructure and with the highest built-up density. In contrast, these two settlements have the lowest values of "IJI Shrub Vegetation" metric. As expected, UFT5—Roads has low distance values to highways, while UFT6—River has low distance values to the river (distance values appear negative because they are normalized).

The use of landscape metrics to classify urban environments was also experimented by Reis [82]. The author achieved reasonable accuracy and discriminatory power using landscape metrics as features, enabling the identification of intraurban patterns in the context of dengue fever surveillance and disease control in Rio de Janeiro, Brazil. The landscape metrics described in our methodology are consonant with Reis [82], demonstrating the potential of these metrics for urban and intraurban environmental studies.

From the methodological point of view, two decisions were central to obtaining high levels of global accuracy and F1-score. First, the use of two levels of segmentation, initially formed by ground cover base, and a second level hierarchically superior formed by hexagonal grids, where neighborhood connectivity allows better extraction of features than in rectangular grids [72]. Adopting these two levels enabled the extraction of landscape

metrics from first-level geographic objects, and the estimation of neighborhood features for second-level objects.
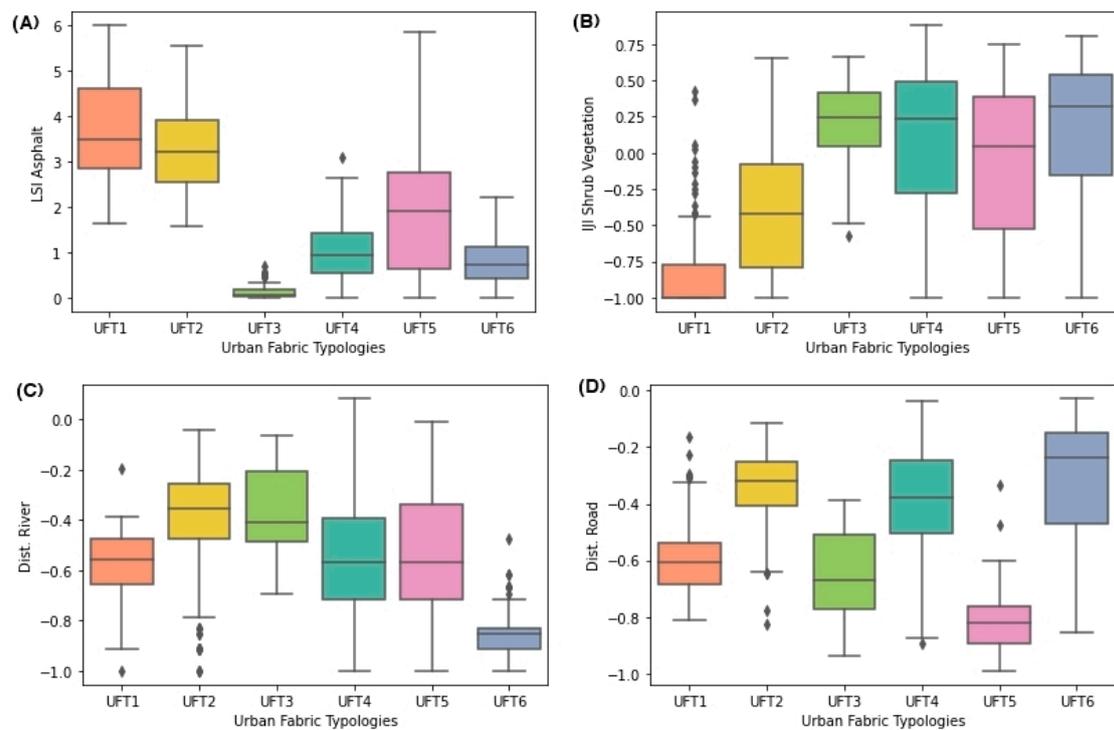


**Figure 9.** Variables included in the final decision tree for our classification model: (**A**) LSI Asphalt; (**B**) IJI Shrub Vegetation; (**C**) distance to the river, and (**D**) distance to the road.

The second main methodological decision was to classify cells using a machine learning algorithm. We applied the C5.0 algorithm to build optimal decision trees, which outperforms manual classification models. Additionally, regarding geographical data for identification and characterization of precarious settlements, two datasets were essential: (i) WPM images from the CBERS-4A satellite, in which spatial resolution of the panchromatic camera allowed the extraction of geographic objects for intraurban analysis scale; (ii) OSM road and hydrological database to calculate distance measures to the nearest river and nearest highway that were fundamental features for most of the UFT characterization.

Although our results show high global accuracy and F1-scores, there are a few methodological gaps that could be improved. Kuffer et al. [18] proposed the inclusion of socioeconomic census data jointly with features extracted by satellite images to improve the classification of precarious settlements. The last Brazilian census was carried out in 2010, and due to the COVID-19 pandemic the census scheduled for 2020 was postponed to 2022. Therefore, our proposed methodology sought to identify precarious settlements using only features from open data sources extracted from remote sensing and volunteered geographic information (VGI).

Our methodology for identifying precarious areas and characterization of UFT combined the GEOBIA approach with a classification using the C5.0 machine learning algorithm. Other algorithms can be assessed for the classification of geographic objects such as support vector machines (SVM) and random forests (RF), which have shown high accuracy in slum identification models [18,83]. However, notice that machine learning algorithms, although showing high accuracy [18], usually require great computational power and a large number of training samples to give useful models. These samples of precarious urban areas may not be available for many Brazilian municipalities and may be scarcer in Amazonian cities, impeding the use of these models for the distinction of intraurban classes. Furthermore, neighborhood features can be included as measures of spatial correlation in

traditional classification models that do not consider the spatial nature of the data, such as the C5.0 classifier and other traditional machine learning algorithms.

In this study, we analyzed images from the Sino–Brazilian Earth Resource Satellite CBERS-4A, launched in 2019 and in the process of commissioning. The CBERS-4A images are freely available in the INPE catalog [61]; however, the parameters necessary for the radiometric correction of the WPM sensor have not yet been published. This prevented us from performing radiometric correction in the images, so we processed the data for each city separately. The radiometric correction would facilitate image processing.

We chose the study areas from census tracts, considering urban those areas in the vicinity of municipal seats or to census tracts that are labeled as urban by IBGE [18]. A future work could include night light images that are usually indicators of built-up areas beyond the boundaries of municipal seats to indicate areas effectively occupied by the population [84]. Moreover, identifying study areas by night light images would not be restricted by the dichotomy of labeling regions as urban or rural areas, as assumed by census tracts.

A final methodological suggestion is the assessment of regression models to identify precarious areas. It allows the estimation of housing precariousness probability surfaces [24,48], showing cells most and least likely to be precarious. Regression models could also identify settlements with mixed characteristics that have lower probabilities to belong to each class.

Since this is the first study to identify and classify precarious settlements by remote sensing in Amazonian municipalities, these methodological suggestions could be tested in a future study. However, we bring to light the work developed by researchers from the *Instituto de Pesquisa Econômica Aplicada* (IPEA), who identified and characterized an informal urban nucleus for each of six Brazilian hubs, including the Marabá hub. Their work analyzed variables collected through remote sensing; IPEA researchers plan to publish a research book in 2022 [85].

Unlike many Brazilian metropolitan areas, nonmetropolitan Amazonian cities, as the study sites, have precarious areas that are not easily distinguishable from nonprecarious regions [13]. The difficulty of distinguishing between precarious and nonprecarious areas in Amazonian cities is mainly due to common poor city infrastructure, especially in environmental sanitation services—characteristics of regional inequality between Northern Brazil and other regions [13]. Another main difference of Amazonian cities is their distinct housing construction specifics: the house accesses (use of wooden bridges), the predominant type of material in the buildings, how the population relates to the place where they live, and the architectural design that carries local empirical knowledge inherited from traditional Amazonian residents, in the symbology of local culture [13].

In addition to methodological suggestions, future studies could broaden the geographical scope, deepening theoretical discussions to better identify precarious Amazonian areas and their typologies. The unique characteristics of urban fabric in Amazonian municipalities require a broader debate on the concept of precarious area, to inventory and extract relevant features that properly describe these regions.

## 7. Conclusions

This work identifies precarious areas in Amazonian municipalities and classifies them into urban fabric typologies (UFT). In addition, this methodology was entirely built using open source software analyzing free access data, using only data collected from remote sensing and volunteered geographic information (VGI). To the best of our knowledge, this work shows the first proposed technique to identify precarious settlements in the Amazonian territory using remote sensing data.

Our methodology to identify precarious areas and classify typologies using GEOBIA and data mining techniques showed mapping results with an overall accuracy above 0.90, indicating that our methodology is a valuable tool for semiautomatic identification of precarious settlements in Amazonian cities. This methodology demonstrated that

precarious settlements in the Amazonian cities studied are diverse and can be grouped into different typologies of urban fabric, contributing to a methodological debate about precariousness in Amazonian cities.

Currently, there are millions of people living in precarious settlements in Brazil. Mapping these precarious settlements by presential site visits or visual analysis is time-consuming, costly, and demands the collaboration of researchers from several fields. The proposed methodology provides accurate information to support municipalities planning to assist in the identification of precarious settlement as another layer of relevant information.

## References

1. Un-Habitat. *World Cities Report 2016: Urbanization and Development–Emerging Futures*; United Nations Human Settlements Programme: Nairobi, Kenyav, 2016; Available online: https://unhabitat.org/world-cities-report (accessed on 29 October 2021).
2. Unstats. SDG Indicators: Make Cities and Human Settlements Inclusive, Safe, Resilient and Sustainable. Available online: https://unstats.un.org/sdgs/report/2019/goal-11 (accessed on 31 October 2021).
3. BRASIL. Lei No. 11.124. 2005. Available online: http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/lei/l11124.htm (accessed on 31 October 2021).
4. BRASIL. *Guia para o Mapeamento e Caracterização de Assentamentos Precários*; Ministério das Cidades: Brasilia, Brazil, 2010; Volume 1.
5. Maricato, E. Autoconstrução, a Arquitetura Possível. A Produção Capitalista da Casa (e da cidade) no Brasil industrial. 1982. Available online: https://erminiamaricato.files.wordpress.com/2012/03/a-produccca7acc83o-capitalista-da-casa-e-da-cidade-no-brasil-industrial.pdf (accessed on 31 October 2021).
6. Denaldi, R.; Rosa, J.S. Ações Integradas de Urbanização de Assentamentos Precários. Brasilia: Ministério das Cidades/Cities Alliance. 2010. Available online: https://bibliotecadigital.seplan.planejamento.gov.br/handle/iditem/272 (accessed on 31 October 2021).
7. Abreu, M.A. Reconstruindo uma história esquecida: Origem e expansão inicial das favelas do Rio de Janeiro. *Espaço Debates* **1994**, *14*, 34–46.
8. Vaz, L.F. Dos cortiços às favelas e aos edifícios de apartamentos—a modernização da moradia no Rio de Janeiro. *Análise Social* **1994**, *29*, 581–597.
9. Pasternak, S. São Paulo e suas favelas. *Pós Rev. Programa Pós-Grad. Arquitetura Urban. FAUUSP* **2006**, *19*, 176–197. [CrossRef]
10. Queiroz Filho, A.P. As definições de assentamentos precários e favelas e suas implicações nos dados populacionais: Abordagem da análise de conteúdo. *Urbe Rev. Bras. Gestão Urbana* **2015**, *7*, 340–353. [CrossRef]
11. Santana, J.V.; Holanda, A.C.G.; de Moura, A.D.S.F. A questão da habitação em municípios periurbanos na Amazônia. *Ed. UFPA* **2012**, *1*, 1–400.
12. Sakatauskas, G.D.L.B.; Santana, J.V. ST 2 Particularidades Das Habitações Nos Pequenos Municípios Paraenses. *An. ENANPUR.* **2015**, *16*, 1–14.
13. Sakatauskas Gdlb. *Especificidades da Precariedade Habitacional na Amazônia Ribeirinha: Um Olhar Sobre a Região do Baixo Tocantins*; Sakatauskas Gdlb: São Bernardo do Campo, Brazil, 2020.
14. Nakano, A.K. Desenvolvimento Urbano e Territorial em Municípios Periurbanos: Diferenciações no Universo dos Municípios brasileiros. *Proposta* **2012**, *35*, 1–60.
15. Becker, B. *A Urbe Amazônida*, 1st ed.; Garamound: Rio de Janeiro, Brazil, 2013; Volume 1, pp. 1–88.

16. Morais, M.D.P.E.; Krause, C.H.; Lima Neto, V.C. *Caracterização e Tipologia de Assentamentos Precários: Estudos de Caso Brasileiros*; Instituto de Pesquisa Econômica Aplicada; Ipea: Brasília, Brazil, 2016; Volume 1, pp. 1–548.

17. Kuffer, M.; Pfeffer, K.; Sliuzas, R. Slums from space—15 years of slum mapping using remote sensing. *Remote Sens.* **2016**, *8*, 455. [CrossRef]

18. Instituto Brasileiro de Geografia e Estatística. Censo 2000: A Divulgação dos Resultados. IBGE. 2010. Available online: https://www.ibge.gov.br/censo/divulgacao.shtm (accessed on 1 November 2021).

19. Instituto Brasileiro De Geografia E Estatística. Censo Demográfico 2010. Rio de Janeiro. 2011. Available online: https://biblioteca.ibge.gov.br/visualizacao/periodicos/93/cd_2010_caracteristicas_populacao_domicilios.pdf (accessed on 1 November 2021).

20. Centro de Estudos da Metropole-Cem, Cebrap. Assentamentos Precários no Brasil Urbano. Brasília. 2007. Available online: https://direito.mppr.mp.br/arquivos/File/relatorio-assentamentos-precatorios-no-brasil-urbano.pdf (accessed on 1 November 2021).

21. Centro de Estudos da Metropole-Cem, Cebrap. Diagnóstico dos Assentamentos Precários nos Municípios da Macrometrópole Paulista. São Paulo. 2013. Available online: https://centrodametropole.fflch.usp.br/pt-br/downloads-de-dados/assentamentos-precarios-2013 (accessed on 1 November 2021).

22. Mahabir, R.; Croitoru, A.; Crooks, A.T.; Agouris, P.; Stefanidis, A. A critical review of high and very high-resolution remote sensing approaches for detecting and mapping slums: Trends, challenges and emerging opportunities. *Urban Sci.* **2018**, *2*, 8. [CrossRef]

23. Duque, J.C.; Royuela, V.; Noreña, M. A Stepwise Procedure to Determinate a Suitable Scale for the Spatial Delimitation of Urban Slums. In *Defining the Spatial Scale in Modern Regional Analysis*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 237–254.

24. Gonçalves, G. *Identificação de Assentamentos Precários na Região do Grande ABC: Uma Abordagem Estatística*; UFABC: Santo André, Brazil, 2018.

25. Galeon, F. Estimation of population in informal settlement communities using high resolution satellite image. In Proceedings of the XXI ISPRS Congress, Commission IV, Beijing, China, 5–9 July 2008; pp. 1377–1381.

26. Karekezi, S.; Kimani, J.; Onguru, O. Energy access among the urban poor in Kenya. *Energy Sustain. Dev.* **2008**, *12*, 38–48. [CrossRef]

27. Alves, M.C.G.P.; Morais, M.D.L.S.; Escuder, M.M.L.; Goldbaum, M.; Barros, M.B.D.A.; Cesar, C.L.G.; Carandina, L. Household sampling in slums in surveys. *Rev. Saude Publica* **2011**, *45*, 1099–1109. [CrossRef] [PubMed]

28. Henderson, J.V.; Storeygard, A.; Weil, D.N. Measuring Economic Growth from Outer Space. *Am. Econ. Rev.* **2012**, *102*, 994–1028. [CrossRef] [PubMed]

29. Novo, E.M.L.D.M. *Sensoriamento Remoto: Princípios e Aplicações*; Editora Blucher: São Paulo, Brazil, 2010.

30. QUILAN, J.R. C5.0: An Informal Tutorial. Available online: http://www.rulequest.com/see5-unix.html (accessed on 1 November 2021).

31. Maricato, E.A. *Cidade do Pensamento Único: Desmanchando Consenso*, 3rd ed.; Revista do Programa de Pós-Graduação em Arquitetura e Urbanismo da FAUUSP: Petrópolis, Brazil, 2002; pp. 250–251.

32. Instituto Brasileiro De Geografia E Estatística—IBGE. Aglomerados Subnormais 2019: Classificação Preliminar e Informações de Saúde Para o Enfrentamento à COVID-19. Rio de Janeiro. 2020. Available online: https://covid19.ibge.gov.br/ (accessed on 1 November 2021).

33. Kurkdjian, M. Um método Para Identificação e Análise de Setores Residenciais Urbanos Homogêneos, Através de Dados de Sensoriamento Remoto, Com Vistas ao Planejamento Urbano. São Paulo. 158p. Tese (Doutorado em Arquitetura e Urbanismo)-FAU-USP. [São Paulo, Brazil]. 1986. Available online: http://mtc-m12.sid.inpe.br/col/sid.inpe.br/iris@1912/2005/07.18.21.29/doc/INPE%206358.pdf (accessed on 1 November 2021).

34. De Lourdes Neves de Oliveira, M.; Manso, A.P.; Barros, M.S.S. Setorização urbana através de Sensoriamento Remoto. Instituto Nacional De Pesquisas Espaciais. 1978. Available online: http://marte.sid.inpe.br/col/dpi.inpe.br/marte@80/2008/09.23.18.09/doc/436-451.pdf (accessed on 1 November 2021).

35. Manso, A.P.; Barros, M.S.S.; Oliveira, M.L.N. *Determinação de Zonas Homogêneas Através de Sensoriamento Remoto*; INPE: São José dos Campos, Brazil, 1978.

36. Gonçalves, C.; Souza, I.D.M.; Pereira, M.N.; Freitas, C.D.C. *Análise do Ambiente Urbano Visando a Inferência Populacional a Partir do uso de Dados de Sensoriamento Remoto Orbital de alta Resolução*; INPE: São José dos Campos, Brazil, 2004.

37. Matias, L.F.; Nascimento, E.D. Geoprocessamento aplicado ao mapeamento das áreas de ocupação irregular na cidade de Ponta Grossa (PR). *Geogr. Rio Claro* **2006**, *31*, 317–330.

38. Hofmann, P. Detecting informal settlements from IKONOS image data using methods of object oriented image analysis-an example from Cape Town (South Africa). *Remote Sens. Urban Areas/Fernerkund. Urbanen Räumen* **2001**, 41–42.

39. Hofmann, P.; Strobl, J.; Blaschke, T.; Kux, H. Detecting informal settlements from QuickBird data in Rio de Janeiro using an object based approach. In *Object-Based Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 531–553.

40. Kohli, D.; Sliuzas, R.; Stein, A. Urban slum detection using texture and spatial metrics derived from satellite imagery. *J. Spat. Sci.* **2016**, *61*, 405–426. [CrossRef]

41. Kohli, D.; Stein, A.; Sliuzas, R.V.; Kerle, N. *Identifying and Classifying Slum Areas Using Remote Sensing*; University of Twente Faculty of Geo-Information and Earth Observation (ITC): Enschede, The Netherlands, 2015.

42. Wurm, M.; Taubenböck, H. Detecting social groups from space–Assessment of remote sensing-based mapped morphological slums using income data. *Remote Sens. Lett.* **2018**, *9*, 41–50. [CrossRef]

43. REUß, F. Detection of Favelas in Brazil using Texture Parameters and Machine Learning. [Graz, Austria]. 2017. Available online: https://elib.dlr.de/115220/1/Reuss_MA.pdf (accessed on 1 November 2021).

44. Gueguen, L. Classifying compound structures in satellite images: A compressed representation for fast queries. *IEEE Trans. Geosci. Remote Sens.* **2014**, *534*, 1803–1818. [CrossRef]

45. Dos Santos, B.D.; de Pinho, C.M.D.; de Jesus, T.B. Níveis De Consolidação De Assentamentos Precários A Partir De Dados De Sensoriamento Remoto. In *XIX Simpósio Brasileiro de Sensoriamento Remoto*; INPE: Santos, Brazil, 2019; pp. 3224–3227. Available online: http://marte2.sid.inpe.br/col/sid.inpe.br/marte2/2019/10.31.12.39/doc/98009.pdf (accessed on 2 January 2021).

46. Stark, T.; Wurm, M.; Zhu, X.X.; Taubenböck, H. Satellite-Based mapping of urban poverty with transfer-learned slum morphologies. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5251–5263. [CrossRef]

47. Kohli, D.; Sliuzas, R.; Kerle, N.; Stein, A. An ontology of slums for image-based classification. *Comput. Environ. Urban Syst.* **2012**, *36*, 154–163. [CrossRef]

48. Feitosa, F.D.F.; Vasconcelos, V.V.; de Pinho, C.M.D.; da Silva, G.F.G.; da Silva Gonçalves, G.; Danna, L.C.C.; Lisboa, F.S. IMMerSe: An integrated methodology for mapping and classifying precarious settlements. *Appl. Geogr.* **2021**, *133*, 102494. [CrossRef]

49. Da Costa, G.A.O.P.; de Pinho, C.M.D.; Feitosa, R.Q.; de Almeida, C.M.; Kux, H.J.H.; Fonseca, L.M.G.; Oliveira, D.A.B. InterIMAGE: Uma plataforma cognitiva open source para a interpretação automática de imagens digitais. *Rev. Bras. Cartogr.* **2008**, *60*, 331–337.

50. Alves, C.D.; Pereira, M.N.; Florenzano, T.G.; Souza, Í.D.M.E. Análise orientada a objeto no mapeamento de áreas urbanas com imagens Landsat. *Bol. Ciências Geodésicas* **2009**, *15*, 120–141.

51. Ribeiro, B.M.G. Avaliação das Imagens WorldView-II Para o Mapeamento da Cobertura do Solo Urbano Utilizando o Sistema InterIMAGE. São José dos Campos. 2010. Available online: http://mtc-m16d.sid.inpe.br/col/sid.inpe.br/mtc-m19@80/2010/08.03.12.52.20/doc/publicacao.pdf (accessed on 1 November 2021).

52. Kux, H.J.H.; Novack, T.; Ferreira, R.; Oliveira, D.A.; Ribeiro, B.M.G. Classificação da cobertura do solo urbano usando imagens ópticas de altíssima resolução e o sistema InterIMAGE baseado em conhecimento. In Proceedings of the Anais XV Simpósio Brasileiro de Sensoriamento Remoto–SBSR, Curitiba, Brasil, 30 April–5 May 2011; p. 822.

53. De Pinho, C.M.D.; Fonseca, L.M.G.; Korting, T.S.; de Almeida, C.M.; Kux, H.J.H. Land-cover classification of an intra-urban environment using high-resolution images and object-based image analysis. *Int. J. Remote Sens.* **2012**, *33*, 5973–5995. [CrossRef]

54. Kux, H.J.H.; Novack, T.; Fonseca, L.M.G. Mapeamento de favelas usando classificação orientada a objeto–estudo de caso em Paraisópolis, São Paulo (SP). *Simpósio Bras. Sens. Remoto* **2009**, *14*, 715–721.

55. Cardoso, A.C.D.; de Melo, A.C.; do Vale Gomes, T. O urbano contemporâneo na fronteira de expansão do capital: Padrões de transformações espaciais em seis cidades do Pará, Brasil. *Rev. Morfol. Urbana* **2016**, *4*, 5–28. [CrossRef]

56. Pontes, L.B. Cidade, Desenho e Natureza: Uma Reflexão Sobre os Espaços Livres de Marabá. [Belém do Pará]. 2015. Available online: https://bdtd.ibict.br/vufind/Record/UFPA_6aaf495909dd3b97f78cd452d6b7dc0b (accessed on 1 November 2021).

57. Cardoso, A.C.D.; Lima, J.J.F. Tipologias e Padrões de Ocupação Urbana na Amazônia Oriental: Para Que e Para Quem. In *O Rural e o Urbano na Amazônia: Diferentes Olhares em Perspectivas*; UFPA: Belém, Brazil, 2006.

58. LISS. Grupo De Pesquisa | Liss Inpe. Available online: https://www.lissinpe.com.br/ (accessed on 1 November 2021).

59. INPE. Câmeras Imageadoras CBERS-4A. INPE. 2019. Available online: http://www.cbers.inpe.br/sobre/cameras/cbers04a.php (accessed on 8 January 2021).

60. INPE. Sobre o CBERS. INPE. 2018. Available online: http://www.cbers.inpe.br/sobre/index.php (accessed on 2 January 2021).

61. Instituto Nacional de Pesquisas Espaciais. Divisão de Geração de Imagem: Catálogo de Imagens. Available online: http://www2.dgi.inpe.br/catalogo/explore (accessed on 1 November 2021).

62. Team, Q.D.; QGIS Geographic Information System. QGIS Association. 2021. Available online: https://www.qgis.org (accessed on 14 November 2021).

63. OTB. Segmentation—Orfeo ToolBox 7.4.0 Documentation. Available online: https://www.orfeo-toolbox.org/CookBook/Applications/app_Segmentation.html (accessed on 1 November 2021).

64. Körting, T.S.; Garcia Fonseca, L.M.; Câmara, G. GeoDMA—Geographic Data Mining Analyst. *Comput. Geosci.* **2013**, *57*, 133–145. [CrossRef]

65. Ponzoni, F.J.; Shimabukuro, Y.E.; Kuplich, T.M. *Sensoriamento Remoto da Vegetação*, 2nd ed.; Oficina de Textos: Paulo, Brazil, 2012; Volume 1.

66. De Pinho, C.M.D.; Ummus, M.E.; Novack, T. Extração De Feições Urbanas Em Imagens De Alta Resolução Espacial A Partir Do Estudo Do Comportamento Espectral Dos Alvos. *Rev. Bras. De Cartogr.* **2011**, *63*, 439–448.

67. Haralick, R.M. Statistical and structural approaches to texture. *Proc. IEEE.* **1979**, *67*, 786–804. [CrossRef]

68. Michel, J.; Youssefi, D.; Grizonnet, M. Stable Mean-Shift Algorithm and Its Application to the Segmentation of Arbitrarily Large Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 952–964. [CrossRef]

69. INPE. GeoDMA Features. Available online: http://wiki.dpi.inpe.br/doku.php?id=geodma_2:features&s[]=landscape (accessed on 1 November 2021).

70. Quinlan, J.R. *C4. 5: Programs for Machine Learning*; Elsevier: Amsterdam, The Netherlands, 2014.

71. Pandya, R.; Pandya, J. C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. *Int. J. Comput. Appl.* **2015**, *117*, 18–21. [CrossRef]

72. Birch, C.P.; Oom, S.P.; Beecham, J.A. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecol. Model.* **2007**, *206*, 347–359. [CrossRef]

73. Pereira, R.H.M.; Kauê, C.; Braga, V.; Serra, B.; Nadalin, V.G. Desigualdades Socioespaciais De Acesso A Oportunidades Nas Cidades Brasileiras—2019. Brasília. 2020. Available online: http://www.ipea.gov.br/portal/publicacoes (accessed on 1 November 2021).

74. Denaldi, R.; Petrarolli, J.G.; Gonçalves, G.; Moraes, G. Tecidos Urbanos E A Identificação De Assentamentos Precários Na Região Metropolitana Da Baixada Santista. In *III URBFAVELAS*; Observatório das Metrópoles: Salvador, 2018; pp. 1–21. Available online: http://lepur.com.br/wp-content/uploads/2018/12/TECIDOS-URBANOS-E-A-IDENTIFICA%C3%87%C3%83O-DE-ASSENTAMENTOS-PREC%C3%81RIOS-NA-REGI%C3%83O-METROPOLITANA-DA-BAIXADA-SANTISTA-.pdf (accessed on 1 November 2021).

75. Cdhu, ufabc. Desenvolvimento e aplicação de metodologia para identificação, caracterização e dimensionamento de assentamentos precários. São Bernardo do Campo. 2018.

76. Ribeiro, R.J.D.C. Índice Composto de Qualidade de vida Urbana: Aspectos de Configuração Espacial, Socioeconômicos e Ambientais Urbanos. 2008. Available online: file:///C:/Users/dias-bruno/Downloads/2008_RomuloJoseCostaRibeiro.pdf (accessed on 15 November 2021).

77. OSM. OpenStreetMap. 2021. Available online: https://www.openstreetmap.org/#map=4/-15.13/-53.19 (accessed on 1 November 2021).

78. Rennó, C.D.; Nobre, A.D.; Cuartas, L.A.; Soares, J.V.; Hodnett, M.G.; Tomasella, J.; Waterloo, M.J. HAND, a New Terrain Descriptor Using SRTM-DEM: Mapping Terra-Firme Rainforest Environments in Amazonia. *Remote Sens. Environ.* **2008**, *112*, 3469–3481. Available online: https://www.sciencedirect.com/science/article/pii/S003442570800120X (accessed on 15 November 2021). [CrossRef]

79. McGarigal, K.; Marks, B.J. Spatial pattern analysis program for quantifying landscape structure. In *General Technical Report*; PNW-GTR-351; US Department of Agriculture, Forest Service, Pacific Northwest Research Station: Portland, OR, USA, 1995; pp. 1–122.

80. De Cidades, A.; das Cidades, M. Ações Integradas de Urbanização de Assentamentos precários: Síntese do Curso à Distância. 2010. Available online: https://antigo.mdr.gov.br/images/stories/ArquivosSNH/ArquivosPDF/SinteseAcaoUrbanizacao_WEB.pdf (accessed on 15 November 2021).

81. Georganos, S.; Abascal, A.; Kuffer, M.; Wang, J.; Owusu, M.; Wolff, E.; Vanhuysse, S. Is It All the Same? Mapping and Characterizing Deprived Urban Areas Using WorldView-3 Superspectral Imagery. A Case Study in Nairobi, Kenya. *Remote Sens.* **2021**, *13*, 4986. [CrossRef]

82. Reis, I.C. Caracterizaçao de Paisagens Urbanas Heterogêneas de Interesse para a Vigilância e Controle da Dengue com o uso de Sensoriamento Remoto e Mineraçao de Padroes Espaciais: Um Estudo para o Rio de Janeiro. Master in Remote Sensing, National Institute of Spatial Research, São José dos Campos, Brazil. 2010. Available online: http://mtc-m16d.sid.inpe.br/col/sid.inpe.br/mtc-m19/2010/12.06.15.22/doc/publicacao.pdf?metadatarepository=sid.inpe.br/mtc-m19/2010/12.06.15.22.45&mirror=sid.inpe.br/mtc-m19@80/2009/08.21.17.02.53 (accessed on 15 November 2021).

83. Leonita, G.; Kuffer, M.; Sliuzas, R.; Persello, C. Machine Learning-Based Slum Mapping in Support of Slum Upgrading Programs: The Case of Bandung City, Indonesia. *Remote Sens.* **2018**, *10*, 1522. [CrossRef]

84. Amaral, S.; Câmara, G.; Monteiro, A.M.V.; Quintanilha, J.; Elvidge, C.D. Estimating population and energy consumption in Brazilian Amazonia using DMSP night-time satellite data. *Comput. Environ. Urban Syst.* **2005**, *29*, 179–195. [CrossRef]

85. Instituto de Pesquisa Econômica Aplicada. Webinar Destacou Principais Resultados da Pesquisa Sobre núcleos urbanos Informais. Available online: https://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=38607&catid=1&Itemid=7 (accessed on 1 November 2021).