



Article

Land Use Land Cover Classification with U-Net: Advantages of Combining Sentinel-1 and Sentinel-2 Imagery

Jonathan V. Solórzano ¹, Jean François Mas ^{2,*}, Yan Gao ² and José Alberto Gallardo-Cruz ³

¹ Posgrado en Geografía, Centro de Investigaciones en Geografía Ambiental, Universidad Nacional Autónoma de México, Antigua Carretera a Pátzcuaro No. 8701, Col. Ex-Hacienda de San José de la Huerta, Morelia CP 58190, Mexico; jsolorzano@pmip.unam.mx

² Laboratorio de Análisis Espacial, Centro de Investigaciones en Geografía Ambiental, Universidad Nacional Autónoma de México, Antigua Carretera a Pátzcuaro No. 8701, Col. Ex-Hacienda de San José de la Huerta, Morelia CP 58190, Mexico; ygao@ciga.unam.mx

³ Centro Transdisciplinar Universitario para la Sustentabilidad, Universidad Iberoamericana Ciudad de México, Prolongación Paseo de la Reforma 880, Lomas de Santa Fe, Ciudad de México CP 01219, Mexico; jose.gallardo@ibero.mx

* Correspondence: jfmas@ciga.unam.mx; Tel.: +52-(443)-322-3865

Abstract: The U-net is nowadays among the most popular deep learning algorithms for land use/land cover (LULC) mapping; nevertheless, it has rarely been used with synthetic aperture radar (SAR) and multispectral (MS) imagery. On the other hand, the discrimination between plantations and forests in LULC maps has been emphasized, especially for tropical areas, due to their differences in biodiversity and ecosystem services provision. In this study, we trained a U-net using different imagery inputs from Sentinel-1 and Sentinel-2 satellites, MS, SAR and a combination of both (MS + SAR); while a random forests algorithm (RF) with the MS + SAR input was also trained to evaluate the difference in algorithm selection. The classification system included ten classes, including old-growth and secondary forests, as well as old-growth and young plantations. The most accurate results were obtained with the MS + SAR U-net, where the highest overall accuracy (0.76) and average F1-score (0.58) were achieved. Although MS + SAR and MS U-nets gave similar results for almost all of the classes, for old-growth plantations and secondary forest, the addition of the SAR band caused an F1-score increment of 0.08–0.11 (0.62 vs. 0.54 and 0.45 vs. 0.34, respectively). Consecutively, in comparison with the MS + SAR RF, the MS + SAR U-net obtained higher F1-scores for almost all the classes. Our results show that using the U-net with a combined input of SAR and MS images enabled a higher F1-score and accuracy for a detailed LULC map, in comparison with other evaluated methods.

Keywords: deep learning; multispectral and synthetic aperture radar (SAR) imagery; convolutional neural networks; tropical landscape mosaic; LULC mapping



Citation: Solórzano, J.V.; Mas, J.F.; Gao, Y.; Gallardo-Cruz, J.A. Land Use Land Cover Classification with U-Net: Advantages of Combining Sentinel-1 and Sentinel-2 Imagery. *Remote Sens.* **2021**, *13*, 3600. <https://doi.org/10.3390/rs13183600>

Academic Editors: Carlos Portillo-Quintero, José Luis Hernández-Stefanoni, Gabriela Reyes-Palomeque and Mukti Subedi

Received: 20 June 2021

Accepted: 2 September 2021

Published: 9 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Land use/land cover (LULC) classification has long been a topic of interest in Earth observation research [1–3]. These studies provide characterizations of large extents of the Earth surface by classifying the continuous variation of its attributes in discrete classes and contribute in the establishment of baselines in LULC change studies, which are essential for the management and monitoring of the land surface [4–7]. For this reason, LULC studies are crucial for the management and monitoring of the land surface.

A variety of LULC classification systems have been developed for different purposes. Particularly in tropical regions, several studies have emphasized the importance of discriminating between old-growth forests and plantations, as well as secondary forests, due to their differences in environmental management and biodiversity conservation [8–12]. Although these three classes may have similar canopy cover, secondary forests and plantations usually hold less above ground biomass, host less biodiversity and provide different ecosystem

services than old-growth forests [13–16]. Furthermore, the clearing of old-growth forests to establish plantations can lead to an increase in the carbon emissions from the dead biomass or soil [17,18].

Previous studies have relied on remote sensors to obtain reflectance or backscattering signals of the land surface to predict different LULC. Therefore, with the development of new sensors and other technological advances, novel methods for obtaining LULC classifications have been proposed [19–23]. Recently, deep learning approaches have gained a lot of popularity due to the increase in accuracy, in comparison with previous machine learning approaches [24–26]. Additionally, deep learning algorithms are capable of learning complex and nonlinear patterns in the spatial and temporal dimensions, and they do not require a previous transformation of the inputs, e.g., calculating spectral transformations such as vegetation indices [27–30]. Thus, deep learning algorithms are among the most explored types of algorithms for obtaining LULC classifications and other Earth observation applications [31–34].

Fully Convolutional Neural Networks are a particular type of Convolutional Neural Networks (CNN) that allow obtaining a class prediction for each pixel in an image [25,26,35–39]. These algorithms are capable of identifying patterns at different scales to produce classifications [32,37]. Thus, CNN architectures are nowadays among the most widely applied algorithms for classification tasks [20], and particularly, the U-net is one of the most popular algorithms in LULC studies [31,32,40] due to its capability of summarizing patterns in both the spectral and spatial domain (for additional details see Section 2.3).

Several studies have successfully used the U-net to evaluate forest disturbance and degradation, identify plantations or buildings, among others [41–48]. Furthermore, this architecture is designed to work with a small sample size, a common problem for LULC classification [25,40]. Nevertheless, the U-net has rarely been trained using multispectral bands (MS) besides RGB ones or in combination with synthetic aperture radar images (SAR) [31,32,37,49], even though the combination of MS and SAR imagery has provided more accurate results to generate LULC maps [7,50–53]. For example, the information of MS images can be very useful to differentiate among certain LULC classes (e.g., water, bare soil, vegetation); however, SAR data can interact with the structure of vegetation (i.e., branches, leaves, stems) and therefore can potentially discriminate between forests and plantations [54].

The U-net algorithm has been used in combination with very high spatial resolution (VHR) imagery because it was initially designed for biomedical image segmentation that intends to segment fine-grained class boundaries [31,40]. Nonetheless, for Earth observation applications, medium resolution images from satellite sensors such as Landsat or Sentinel are preferable due to their wider spatial coverage, sufficient resolution for land cover mapping, and most importantly, their free-of-cost availability [44,49,51,55]. All of these traits make them an excellent option for environmental monitoring.

In this context, the objective of this study is to evaluate the potential of the U-net in combination with Sentinel-1 and Sentinel-2 images to develop a detailed LULC classification in a tropical area in Southern Mexico, with a particular interest to differentiate young and old-growth plantations, as well as secondary and old-growth forests. In addition, this evaluation includes comparing the results obtained with the U-net and the random forests (RF) algorithm (a machine learning algorithm), as well as assessing the effect of image input over the classification accuracy. Because the U-net summarizes both spectral and spatial features to perform the LULC classification, we expect that it will obtain higher accuracy than the RF algorithm, which only uses spectral features. In addition, we assume that the combination of MS and SAR with the U-net will help differentiate natural forests from plantations because of their difference in spatial configuration (i.e., random vs. uniform).

2. Materials and Methods

2.1. Study Site

The study site is located in southeastern Mexico, in the municipalities of Marqués de Comillas, Benemérito de las Américas and Ocosingo, Chiapas (Figure 1). This area is part of the Selva Lacandona region, which holds one of the largest massifs of conserved tropical rainforest in North America [56,57]. Additionally, this region shows some of the highest deforestation rates in the country, which have been related to livestock ranching and, to a lesser extent, to agriculture or rubber/oil palm plantations [58–60]. This region exhibits a complex mosaic of LULC classes that includes tropical rainforest (old-growth and secondary), oil palm and rubber plantations, grasslands, agricultural fields, water, roads, areas with no or scarce vegetation, human settlements and aquatic vegetation.

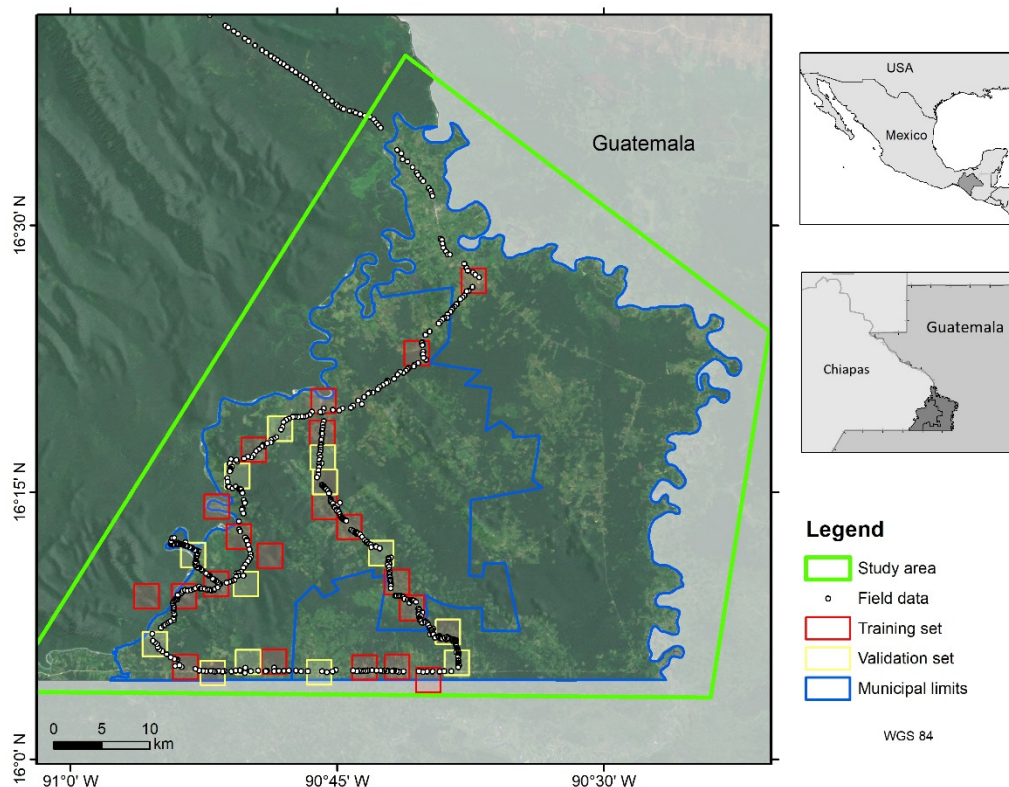


Figure 1. Study site location. Additionally, the in-field LULC data locations, as well as the training and validation areas are shown.

The complete method is divided in three sections: (1) satellite imagery and LULC classes acquisition and preprocessing, (2) algorithm training and validation (with two subsections: U-net and RF) and (3) most accurate architecture selection, complete study area LULC classification and accuracy assessment (Figure 2). The following paragraphs will describe each one of these sections.

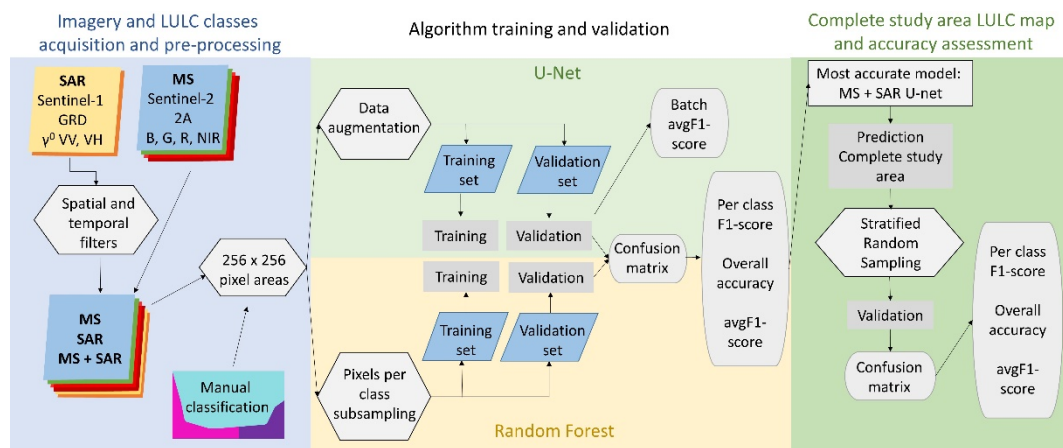


Figure 2. Schematic representation of the complete procedure used in the study.

2.2. Imagery and LULC Classes Acquisition and Pre-Processing

2.2.1. Remote Sensing Input Imagery

Sentinel-1 (SAR) and Sentinel-2 (MS) images were the imagery inputs used to train the U-net and RF algorithms. For Sentinel-1, the Ground Range Detected (GRD) collection was selected and only images acquired in the interferometric wide swath mode and ascending orbit were consulted. The two available SAR bands, vertical transmission and reception (VV) and vertical transmission and horizontal reception (VH), were converted from σ^0 to γ^0 by correcting the backscatter coefficient to the angle of acquisition of the image [61,62]. For the Sentinel-2 images, only the bands with the finest resolution were used, i.e., bands B, G, R and NIR with a pixel size of 10 m. These images corresponded to the Sentinel-2 2A collection, i.e., bottom of the atmosphere reflectance.

We first selected the Sentinel-2 image that had the lowest cloud cover percentage and was closest to the acquisition date of the field data (recorded in March 2019). This corresponded to an image acquired on 4 July 2019. Afterwards, the mean backscattering coefficient of nine Sentinel-1 images that were acquired up to one month prior to this date, i.e., from 4 June 2019 to 4 July 2019, was calculated to reduce the speckle noise typical of SAR images. In addition, a circular filter with radius = 3 pixels was applied to the mean backscattering image to further reduce the speckle noise.

Finally, three different remote sensing imagery datasets were constructed to test the performance of the U-net: (1) MS (4 bands), (2) SAR imagery (2 bands) and (3) MS + SAR imagery (6 bands). All the image processing was carried out in Google Earth Engine using the Javascript API (<https://code.earthengine.google.com/>, accessed on 16 May 2021) [63].

2.2.2. Acquisition of Training and Validation Data

Training data were manually classified through visual interpretation by an interpreter who performed the field data acquisition in 33 systematically selected areas of 256×256 pixels, where LULC information was gathered in the field (Figure 1). The training data were created using these in-field observations (300 points) and remote sensing imagery which included Sentinel-2 images from 2016–2020, Planet images from January and March 2019 (pixel size = 3.12 m) [64] and VHR images provided by Google Earth [65], Yandex [66] and Bing [67] (pixel size < 1 m) as XYZ tiles in QGIS 3.16 [68]. The 33 areas were visually classified into 10 classes: old-growth forest, secondary forest, old-growth plantations, young plantations, grasslands/agricultural fields, roads, soil or areas with no vegetation, water, human settlements and aquatic vegetation. Additionally, the training dataset for MS and MS + SAR included two additional classes, clouds and shadows, which represented areas for which a LULC class could not be obtained based on the MS data, thus these areas would be masked out in the final LULC map. These two classes were not included in the calculation of the final performance metrics (see Section 2.3.1 and Section 2.3.2).

The resulting classifications were rasterized with a pixel size equal to the Sentinel-1 and Sentinel-2 resolution (i.e., 10 m). Finally, the LULC manual classification maps consisted of 12 bands, one for each class, with binary values of 1: presence and 0: absence. Finally, to estimate the error associated with the visual interpretations, a stratified random sampling procedure was implemented, which resulted in 119 points. These points were interpreted without the field acquired information, and these interpretations were contrasted with the field information to calculate the error present in the training data.

In addition, several pre-processing operations were performed on the U-net's input information. First, imagery data were standardized by subtracting each band's mean and divided by its standard deviation. Moreover, both the imagery and the LULC manual classification data were augmented by subsampling each 256×256 -pixel area into 128×128 pixel tiles using a 64 pixel offset and mirroring these areas in both a vertical and horizontal directions. After this procedure, the total augmented dataset consisted of 891 sample units. From this dataset, 621 observations, from 23 images (256×256 pixel tiles) were used as the training set and 270 observations, from 10 images (256×256 pixel tiles) were used as the verification set. This procedure was performed using R 4.0.3 [69] and the raster [70], rray [71] and reticulate [72] packages.

2.3. Algorithm Training and Validation

First, the U-net architecture with different inputs was trained. Afterwards, the imagery input that obtained the highest F1-score with the U-net was used to train a RF algorithm. The comparison among the three types of trained U-nets had the purpose of evaluating the effect of the imagery inputs on its capabilities to generate a LULC map, while the comparison of the most accurate U-net with the RF algorithm had the objective of providing a point of comparison for algorithm selection (i.e., U-net vs. RF).

2.3.1. U-Net

The U-net is a CNN based algorithm; therefore, like any CNN, it "learns" to recognize the classes of interest using a supervised scheme by fixing the weights of convolutional filters in an iterative process [27,28]. These convolutional filters are usually organized in a network-like structure that enables the CNN to recognize spectral and spatial patterns at different scales [32,37]. The U-net has two parts, an encoder and a decoder (Figure 3). In the encoder section, the input image is passed through several hidden layers. In each pass, the spatial resolution is reduced by the effect of the down sampling filters, while the "spectral" resolution is increased. On the contrary, in the decoder part, the image is passed through other hidden layers that perform the opposite process of the encoder part. Thus, in each pass, the image loses spectral resolution while it gains spatial resolution to obtain the final LULC classification. Two outputs are obtained from the U-net: the LULC classification map and the probability map. The typical U-net architecture comprises of five hidden layers [40]; however, in this study, due to memory restrictions, a simpler version of the U-net was used, with two to four hidden layers (Figure 3), which is expected to also reduce the chances of overfitting [73]. For a detailed description of the U-net architecture please consult [40].

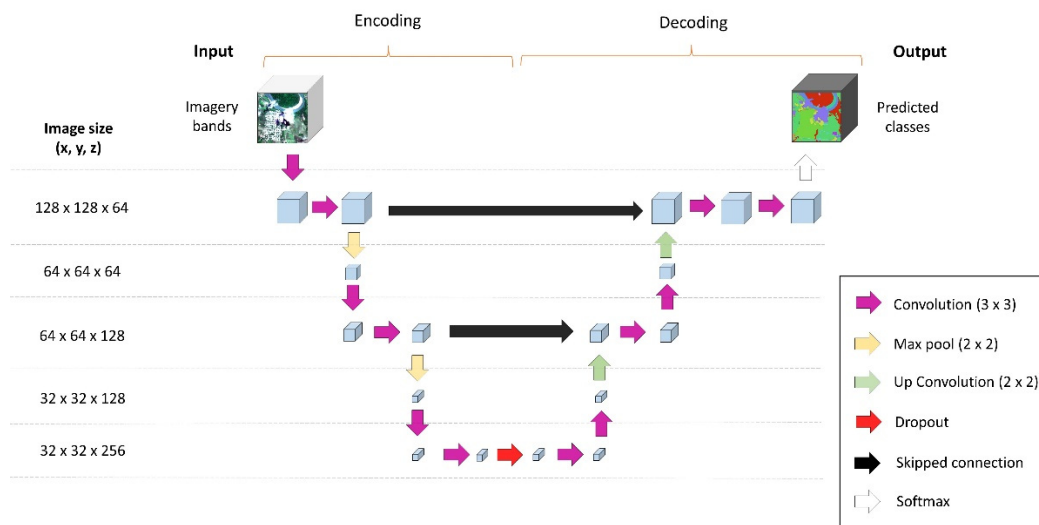


Figure 3. U-Net encoder/decoder architecture diagram. Image size (x, y, z) shows the dimensions of the images: width (x), height (y) and number of bands (z). This diagram shows a three hidden layer U-net, in which steps with the same number of bands correspond to a single hidden layer.

The whole training process was implemented in R 4.0.3, using the U-net and keras packages [74,75] to construct the U-net architecture and tensorflow [76] as the backend for keras. In turn, to search for the optimal hyperparameters, several iterations were fitted using an early stopping procedure to avoid overfitting. In this procedure, training was stopped if the loss metric over the validation set did not decrease in 0.01 in 10 iterations. The tested combination of hyperparameters included batch size (8, 16, 32), number of filters in the first layer (32, 64), number of hidden layers (2–4) and dropout probability (0, 0.1, 0.2, 0.3, 0.4, 0.5) [77]. In addition, cross-entropy was used as the loss function, while Adam was the selected optimizer [78]. Additionally, a batch normalization and a He initialization were implemented to set the initial weights [79].

In order to monitor the algorithm's training, two additional metrics, overall accuracy and average F1-score (avgF1-score; Equation (1)), were calculated in each iteration on both the training and validation sets. The overall accuracy was calculated as the number of correctly classified pixels over the total of pixels. Additionally, the F1-metric was calculated as the harmonic mean of precision and recall (Equation (1)). Afterwards, all the F1-scores were averaged to obtain an avgF1-metric [80]:

$$avgF_1 = \frac{1}{C} \sum_{c=1}^C 2 \frac{pr}{p+r} \quad (1)$$

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

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

where $avgF_1$ stands for the overall average F1-score, C for the number of classes (i.e., 10 classes), c for each class, p for precision, r for recall, TP for true positives, FP for false positives and FN for false negatives (Equations (2) and (3)). During the training phase, the F1-scores are calculated as an average of the F1-score of all the batches of the epoch. Thus, the U-net comparisons were made using this batch-averaged F1-score. However, because the RF training is not performed with batches, an avgF1-score was calculated for each U-net architecture, in order to enable F1-score comparisons with the RF algorithm. This avgF1-score was obtained directly from the confusion matrix observations (with the validation data), instead of an average of the per batch F1-score. Finally, the most accurate U-net architecture was selected as the one with the highest avgF1-score.

The hyperparameter exploration was done using the tfruns package [81]. In total, 90 architectures were generated for each type of input (i.e., MS, SAR and MS + SAR), which resulted in a total of 270 trained U-nets. All of the training and validation procedure was run in an NVIDIA RTX 2060 with 6 GB of memory. The code used to augment the training data, as well as the U-net hyperparameter exploration and training are available at <https://github.com/JonathanVSV/U-netR> (accessed on 26 April 2021).

2.3.2. Random Forests

The RF algorithm is one of the most popular machine learning algorithms in LULC classification [21,52,82]. This algorithm is an ensemble learning method based on decision trees that trains several trees on random subsets of the sample observations and predictive variables. Afterwards, this algorithm assigns the final class for each pixel as the one that had the most votes from the individual trees [83].

The RF classification was performed using the same input information as the most accurate U-net (i.e., MS + SAR, see Results section). In this classification, the same training and validation data as in the U-net were used, but without the augmentation procedure. Thus, the training data consisted of 23 areas of 256×256 pixels, while the validation data, of 10 areas of 256×256 pixels. Due to computing memory restrictions, the RF algorithm could not be trained using the complete training set; therefore, we decided to randomly sample the training set to obtain a balanced dataset with the same number of observations per class [84]. The training data for the RF algorithm consisted of 6057 points by class, because this was the smallest number of observations (i.e., pixels) in the rarest class (i.e., aquatic vegetation). The RF algorithm was trained using the random Forest R package [85] with 500 trees and two randomly selected variables at each split. Once the RF was trained, it was used to predict the LULC classification of the validation data, and the LULC classes were evaluated using a confusion matrix and the same metrics as the U-net results, i.e., overall accuracy and avgF1-score.

2.4. Complete Study Area LULC Classification and Accuracy Assessment

The most accurate U-net architecture (i.e., MS + SAR) was afterwards used to predict the LULC classification of the complete study area. A known problem in this step is that frequently the edges of the predicted tiles show results with less quality than the one in areas closer to the center of the tile [40,43]. This is caused by the fact that pixels found in the edges of the tiles have a smaller number of neighboring pixels than its counterparts farther from the edges; thus, the spatial information extracted by the convolutions is limited in areas closer to the edges. In order to reduce this effect, we divided the complete study area into two grids of 128×128 pixel tiles so that the edges of the predicted tiles in one grid overlapped with the center of the predicted tiles of the other grid. Afterwards, the predicted tiles in both grids were mosaicked, selecting the LULC class of each pixel as the one with the highest probability.

The LULC classification accuracy for the complete study area was evaluated following a stratified random sampling design [86–88] where the random points are distributed by the proportion of the area occupied by each class in the LULC map, except for the rarest classes, where a larger number of random points than its corresponding area proportion is assigned. The total number of sample units for the verification process was calculated as (Equation (4); [86]).

$$n = \frac{\left(\sum_i^q W_i S_i\right)^2}{\left(S(\hat{O})\right)^2 + (1/N) \sum_i^q W_i S_i^2} \approx \left(\frac{\sum_i^q W_i S_i}{S(O)}\right)^2 \quad (4)$$

where $S(\hat{O})$ is the desired standard error of the estimated overall accuracy. Here we adopted a slighter higher value than the one recommended by [87], 0.015. W_i is the proportion of area occupied by each class in the classification, while i stands for each class, while q stands for the number of classes. S_i is the standard deviation of class i ,

$S_i = \sqrt{U_i(1 - U_i)}$, where U_i stands for the a priori expected user's accuracy. In this case, we used the user's accuracy values obtained for the validation dataset (Tables A1–A3).

For the rarest classes (classes with less than 5% of the map, i.e., all classes except old-growth forest, grassland/agriculture and soil), 25 points were assigned per class, while, for the most abundant classes (old-growth forest, grassland/agriculture and soil), the number of verified points was assigned as a proportion of the area occupied by each class, giving a total of 448 points. The verification process was performed by visual interpretation using the same information as the one performed to create the manual classification of the 256×256 pixel tiles, i.e., Sentinel-2 2016–2019 images, Planet 2019 images, Google Earth, Bing and Yandex.

Finally, the Openforis accuracy assessment tool [89] was used to calculate unbiased area estimates, and its 95% confidence intervals. These estimates were based on the confusion matrix resulting from the accuracy assessment process and the proportion of area occupied by each class according to Equations (5) and (6). Additionally, the F1-scores of each class, overall accuracy and avgF1-score were calculated.

$$S(\hat{p}_{.k}) = \sqrt{\sum_i^q W_i^2 \frac{n_{ik}}{n_i} \left(1 - \frac{n_{ik}}{n_i}\right)} \quad (5)$$

$$95\% \text{ CI } \hat{A}_k = \left(A \sum_{i=1}^q W_i \frac{n_{ik}}{n_i} \right) \pm 1.96 (A S(\hat{p}_{.k})) \quad (6)$$

where $S(\hat{p}_{.k})$ is the standard error of the estimated area proportion for class k , W_i is the area proportion of map class i , n_{ik} stands for the sample count at cell (i,k) in the error matrix, while n_i is the row sum for class i . In addition, 95% CI stands for the 95% confidence intervals of \hat{A}_k , the estimated area of class k , A is the total map area, while q stands for the number of classes.

3. Results

3.1. U-Net

3.1.1. Input Imagery and Hyperparameter Exploration

The results for the hyperparameter exploration showed that MS + SAR had the highest overall accuracy (0.76) and avgF1-score (0.58) on the verification dataset, followed very closely by MS (overall accuracy = 0.75, avgF1-score = 0.55) and finally SAR (overall accuracy = 0.65, avgF1-score = 0.39; Tables A1–A3). A similar pattern was observed for the same metrics evaluated on the training dataset (MS + SAR: overall accuracy = 0.91, avgF1-score = 0.86; MS overall accuracy = 0.91, avgF1-score = 0.86; SAR: overall accuracy = 0.72, avgF1-score = 0.60). Thus, MS + SAR U-net was selected as the most accurate architecture. Two of the most accurate architectures, MS + SAR and MS, had three hidden layers, batch size = 16 and 64 filters in the first hidden layer (Table S1). However, the MS + SAR had a higher dropout (0.5) than the MS (0.1). In the case of the most accurate SAR U-net, it had two hidden layers, batch size = 8, 64 filters in the first hidden layer and an intermediate dropout (0.3; Table S1). Finally, an error of 0.15 was detected for the visual interpretations made for the training data without having access to the field information.

3.1.2. Image Input Comparison

MS+SAR U-net was the architecture that had the highest avgF1-score and overall accuracy. In comparison, MS U-net showed a slightly lower avgF1-score and overall accuracy. However, by comparing the F1-scores per class, it was evident that MS + SAR had higher scores in three classes (Δ F1-score ≥ 0.05 ; old-growth plantations, secondary forest and young plantations) and lower scores in a single class (Δ F1-score = 0.08; roads). Additionally, the other five classes had a similar F1-score in either MS + SAR or MS U-net (Δ F1-score ≤ 0.01 ; Table 1). In comparison with the SAR U-net, MS + SAR and MS had a higher F1-score for all the classes (Table 1); however, the F1-score for water was similar

in the three U-net architectures ($\Delta F1$ -score = 0.03). The differences in class identification among different U-net architectures were also visually evident by comparison with the original manual classification (Figure 4). Although there were particular differences among the three U-net architectures, in general, the classes with highest F1-score were old-growth forest, water and human settlements, while the poorly classified classes were young plantations, roads and aquatic vegetation (Tables 1 and A1–A3).

Table 1. F1-score and the difference from the highest score ($\Delta F1$ -score) for each class for the MS + SAR, MS, SAR U-net and MS + SAR Random Forests. MS: Multispectral bands, SAR: synthetic aperture radar bands, RF: Random forests.

Class	U-Net						RF	
	MS + SAR		MS		SAR		MS + SAR	
	F1-Score	$\Delta F1$ -Score	F1-Score	$\Delta F1$ -Score	F1-Score	$\Delta F1$ -Score	F1-Score	$\Delta F1$ -Score
Aquatic vegetation	0.15	0	0.10	0.05	0	0.15	0.04	0.11
Grassland/Agriculture	0.78	0	0.77	0.01	0.69	0.09	0.63	0.15
Human settlements	0.87	0	0.87	0	0.45	0.42	0.29	0.58
Old-growth forest	0.86	0	0.86	0	0.79	0.07	0.69	0.17
Old-growth plantations	0.62	0	0.54	0.08	0.47	0.15	0.43	0.19
Roads	0.35	0.08	0.43	0	0	0.43	0.23	0.2
Secondary forest	0.45	0	0.34	0.11	0.20	0.25	0.33	0.12
Soil	0.65	0	0.64	0.01	0.30	0.35	0.57	0.08
Water	0.96	0.01	0.97	0	0.94	0.03	0.94	0.03
Young plantations	0.11	0	0.03	0.08	0.01	0.1	0.11	0

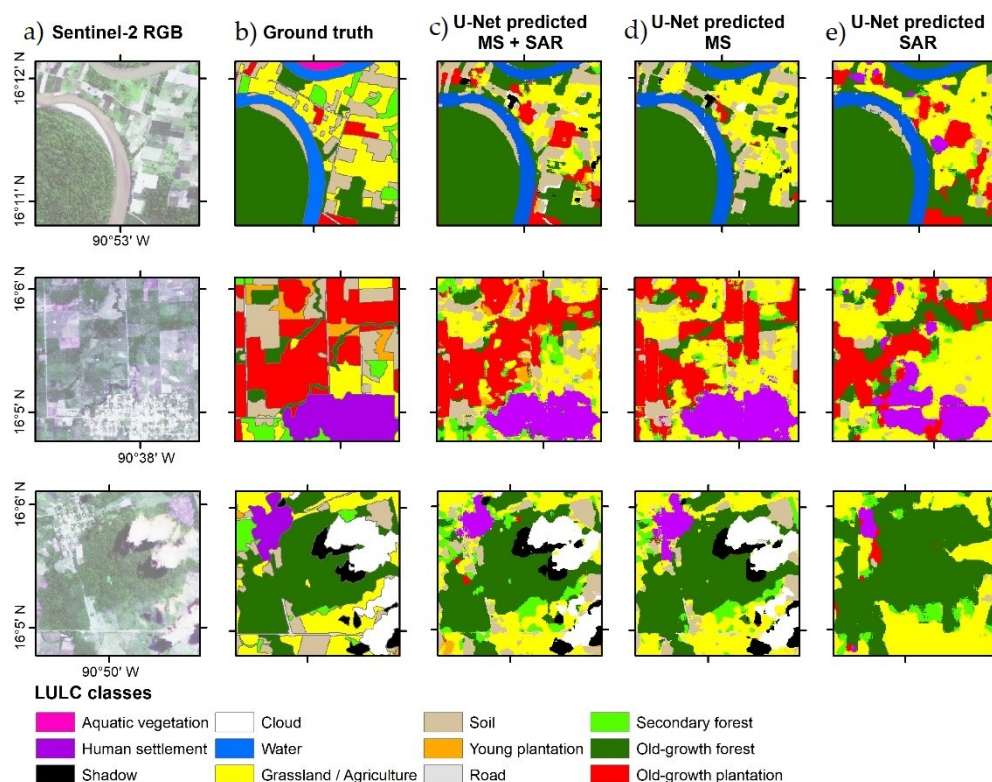


Figure 4. Evaluation of the U-net predictions in three different areas of the validation set. (a) RGB Sentinel-2 image, (b) Manually classified areas, (c) U-net prediction with multispectral (MS) + synthetic aperture radar (SAR) bands, (d) U-net prediction with MS, (e) U-net prediction with SAR.

3.2. Algorithm Comparison

When comparing the results obtained by the MS + SAR U-net and the MS + SAR RF classification, it was obvious that the U-net had both higher avgF1-score and overall accuracy than its RF equivalent (Δ overall accuracy = 0.23 and Δ F1-score = 0.15; Table A4). Although the same pattern was observable by comparing the class F1-scores ($0.08 \leq \Delta$ F1-score ≤ 0.58), water showed a slightly lower F1-score using the RF algorithm (Δ F1-score ≤ 0.03), while young plantations showed the same F1-score in MS + SAR RF and MS + SAR U-net (Table 1).

3.3. Complete Study Area LULC Classification

The accuracy assessment of the complete study area classification showed an overall accuracy of 0.77 and avgF1-score of 0.68 (Table A5). In this verification, the classes that obtained the highest F1-score were water, human settlements, old-growth forests, grasslands/agriculture (Table 2). On the contrary, the classes with the lowest F1-scores were young plantations, secondary forest, soil and roads (Table 2). The complete LULC classification can be downloaded from <https://github.com/JonathanVSV/U-netR> (accessed on 26 April 2021).

Table 2. Area estimates and F1-score obtained for the U-net ensemble classification, as well as the unbiased area and 95% confidence intervals for the area occupied by each class. CI: confidence intervals.

Class	U-Net Study Area LULC Classification Accuracy Assessment			U-Net Validation Dataset	Area Estimates	
	Area (ha)	Proportion of Study Area (%)	F1-Score	F1-Score	Unbiased Area	95% CI
Aquatic vegetation	467.06	0.21	0.70	0.15	6184.96	3848.47
Grassland/Agriculture	84,572.15	38.03	0.79	0.78	76282.87	6516.50
Human settlements	2494.09	1.12	0.94	0.87	3322.42	1082.37
Old-growth forest	93,341.09	41.97	0.87	0.86	90772.02	5449.89
Old-growth plantations	8076.54	3.63	0.56	0.62	7511.87	3233.03
Roads	806.18	0.36	0.60	0.35	6111.42	2790.73
Secondary forest	6195.65	2.79	0.54	0.45	5824.51	2793.85
Soil	17,769.69	7.99	0.55	0.65	12,162.86	4080.16
Water	4617.35	2.08	0.96	0.96	4780.39	319.57
Young plantations	4048.08	1.82	0.31	0.11	9434.56	3823.09

Visually analyzing the probability map of the final classification, it became evident that most of the old-growth forest areas had high probability values corresponding to the assigned class (Figure 5). On the contrary, areas found in the limits of different LULC classes, as well as clouded areas with recently burned areas, had low probabilities of corresponding to the assigned class (Figure 5).

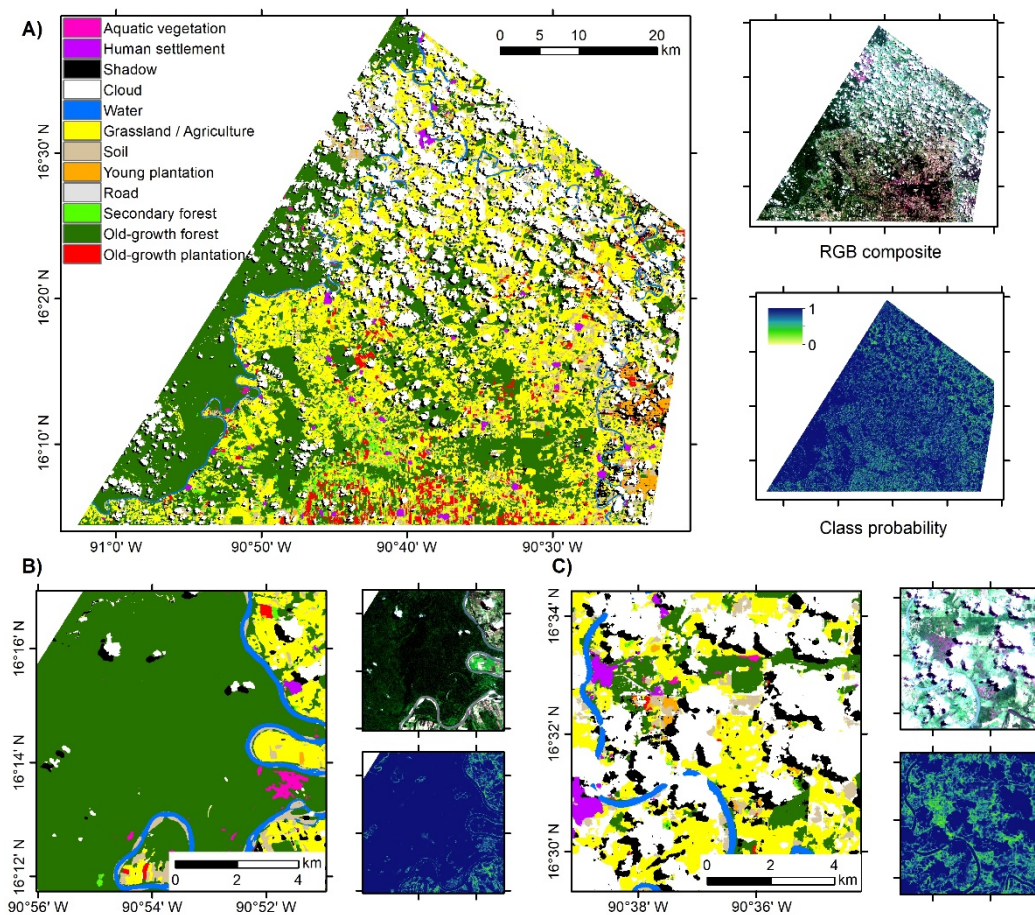


Figure 5. (A) Complete study area land use/land cover (LULC) classification using the MS + SAR U-net, along with its corresponding RGB composite and class probability maps. (B) Example of old-growth forest dominated areas with high class probability. (C) Example of areas with clouds and shadows that were recently cleared with low class probability.

4. Discussion

In this study we showed that deep learning algorithms with combined MS and SAR images can obtain a detailed LULC classification and discriminate among the forested classes of interest with promising results. Although the F1-scores obtained for secondary forests and old-growth plantations were not as high as the one of old-growth forests, the MS + SAR U-net indeed provided an increase between 0.08 and 0.25 in comparison with other tested methods (i.e., MS U-net, SAR U-net or MS + SAR RF). Thus, these results show that new classification methods can improve the capabilities of performing LULC maps from remote sensing information; however, they might still not meet the desired accuracies. It is worth mentioning that our results are relative to the comparisons made in regard to the type of images used (Sentinel-1 and Sentinel-2), as well as the algorithms compared (U-net and RF) and to the LULC classification system used. Future studies could help determine the difference in accuracy to perform LULC classifications in other study areas, with different LULC systems and comparing it with different supervised classification algorithms.

4.1. Algorithm Selection

Our results clearly showed the advantages of using deep learning methods, such as U-net, over traditional machine learning approaches, such as RF in LULC classification. The avgF1-score and accuracy obtained with the MS + SAR U-net was 0.15 and 0.23 higher than its RF equivalent (Tables A1 and A4). Additionally, 9 out of 10 classes obtained higher F1-scores in the MS + SAR U-net, in comparison with its RF counterpart, due to the

inclusion of features in the spatial domain. Previous studies comparing RF with a deep learning algorithm for mapping applications have reported similar outcomes [42,90–95].

Comparing the results obtained for each class of the MS + SAR U-net and its RF counterpart, three notable cases were detected. The first one consisted of classes that showed higher F1-score using the U-net, while its RF counterpart obtained Δ F1-scores between 0.08–0.19 (Table 2). Most classes are included in this category (i.e., all classes except human settlement, water and young plantations); thus, the identification of these classes benefited from including information of the spatial domain. However, most of the potential to correctly identify these classes seems to derive from the spectral domain (Figure 6).

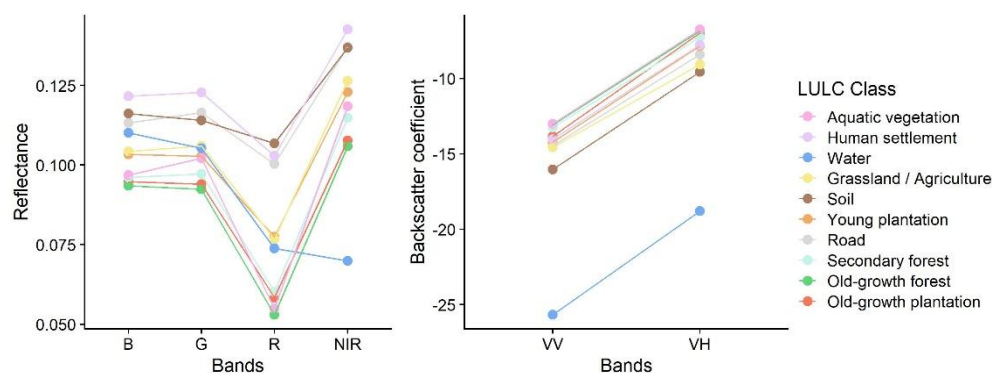


Figure 6. Mean spectral signal of the twelve classes included in the land use/land cover (LULC) classification system.

The second case refers to human settlements which shows that incorporating the spatial domain clearly boosted the performance for its identification (Δ F1-score = 0.58; Table 2). Because individual pixels corresponding to human settlements can have very different reflectance responses (e.g., buildings, trees, grassland: Figure 6), its correct detection strongly benefits from the information available in the spatial domain. Thus, using the U-net clearly outperforms the results obtained by RF, which essentially relies only on spectral information.

The third case consisted of two classes, water and young plantations, where the Δ F1-score between U-net and RF was minimal (Table 2). For the water class, because it is easily differentiated using spectral or backscattering information (Figure 6), including the spatial domain features only cause a minimal increment in its correct detection. In the case of young plantations, MS + SAR U-net and RF obtained the same F1-score. Thus, this class does not benefit from the spatial features of the image. We further discuss this topic in the Error analysis section.

4.2. U-Net: Imagery Input

4.2.1. Class Patterns

Among the different U-net architectures, the one that included MS + SAR as image input gave the highest avgF1-score; however, MS gave almost equally accurate results. This means that the classification capabilities of the MS + SAR U-net derive mainly from the MS imagery. This result is similar to previous studies, where using a combined input of MS + SAR outperforms the results obtained separately with MS or SAR to perform LULC classifications or detect LULC changes [52,96–99].

When comparing the F1-scores obtained for each class using the MS, SAR and MS + SAR U-nets, we discovered five different groups of responses. The first one was characterized by a lower F1-score in the MS + SAR U-net in comparison with the MS one (Table 1). Roads was the only class found with this type of response. Because this class consists of narrow lines of pixels, its classification is easier done using only the MS data, which has the finest spatial resolution.

The second group of classes were those that obtained intermediate to high F1-scores, and obtained higher scores with MS + SAR than only MS or SAR. The two classes in this category were old-growth plantations and secondary forest (Table 1). This was probably due to the ability of SAR data, in this case the Sentinel-1 C-band, to slightly penetrate the canopy cover and acquire information about the geometric arrangement and texture of the old-growth plantations and secondary forests. As previous studies have shown, the use of SAR bands in addition to MS help discriminate old-growth plantations from forests, particularly when the temporal dimension is included [10,11]. Although the penetration of Sentinel-1 C-band is not as deep as those from L or P band [100], this information aids in the correct detection of plantations or secondary forest, particularly when used with the MS set.

The third group consisted of classes where the F1-score difference between MS and MS + SAR was negligible ($\Delta F1\text{-score} \leq 0.01$; Table 1). Most LULC classes showed this pattern: grassland/agriculture, human settlements, old-growth forest, soil and water. For these seven classes, it was evident that MS bands gave the U-net most of its abilities to correctly classify them; thus, adding the SAR band did not substantially improve the F1-score for these classes.

The fourth corresponded to the water class, where its F1-score was very similar among the three U-nets with different imagery inputs (Table 1). This seems to be related to the particular spectral signal of water, which makes it easily distinguished with either MS or SAR signals (Figure 6). A similar conclusion was reached in the comparison between U-net and RF.

The last one refers to two classes, aquatic vegetation and young plantations. Although its F1-score was higher for the MS + SAR U-net than the MS U-net, its F1-score was low (0.11–0.15). Therefore, we concluded that the U-net had small capabilities of identifying these two classes. We further discuss this topic in the Error analysis section and in the complete study area accuracy assessment.

4.2.2. Hyperparameters Exploration

The exploration of different hyperparameters of the U-net showed that although the different combinations of hyperparameters affect the overall accuracy and avgF1-scores, the magnitude of its effect is limited to a 0.05 and 0.11 interval difference between the highest and lowest scores (Table S1). The three most accurate U-nets trained with different imagery inputs tended to consist of relatively simple architectures (two-three hidden layers). This result may be related to the relatively small dataset with which the U-net was trained. Previous studies have reported that CNN with a larger number of filters or hidden layers are capable of identifying more complex patterns; thus, capable of resolving more complex tasks with higher accuracy [27,30,101]. Nevertheless, when a limited training set is available, such as the one used in this study, the main problem with these architectures is that they tend to overfit. Thus, choosing simpler architectures reduces the chances of overfitting although it might limit the abstraction capabilities of the CNN [31,47,102].

4.3. Error Analysis

The incorrect discrimination of certain classes by the U-net consisted of five possibilities: (1) similarity in spectral/spatial information, (2) a similar conceptual class definition, (3) spatial errors probably caused by the U-net architecture, (4) small number of observations and (5) possible errors in the training data caused by the date difference in the VHR images. The first three were related to a limitation in the spatial resolution of the Sentinel-1 and Sentinel-2 images.

The first case was the most dominant source of confusion. Although the field data and VHR images helped determine the LULC of each polygon, the resolution of the input images (10 m) was too coarse to distinguish certain classes [49]. For example, regardless of the imagery used to train the U-net, at 10 m resolution, very young plantations (mainly rubber and oil palm) cannot be distinguished from herbaceous cover (grassland/agriculture) or

plantations at an intermediate growth stage with relatively larger individuals are practically indistinguishable from old-growth plantations (Table A5). On the other hand, most errors among the tree-dominated classes can be related to this type of confusion, where old-growth forests and plantations, as well as secondary forests, have very similar spectral/spatial signals (Table A5). This source of confusion was also evident in the error assessment procedure performed on the training dataset without having access to the field registered data, which obtained an accuracy of 0.85. This procedure showed that certain errors could be associated with the visual interpretation and that for certain classes, such as young plantations, the field data was essential to correctly classify these areas. Considering the insights provided by previous studies [103], the error associated with each LULC would likely increase in comparison with the ones reported here. Thus, the use of multiple interpreters (at least three) could help reduce visual interpretation errors [103].

The second circumstance is closely related to the arbitrary decisions that had to be made to manually classify the training and validation datasets, particularly to define LULC classes limits. Although the same guides and criteria were followed to manually classify the training and validation data, in some cases, both the conceptual and physical limits between certain classes were not completely clear. For example, the delimitation of water bodies and sand banks (i.e., soil) or any vegetated class and roads, where the limits were established mainly on mixed pixels which corresponded to neither one class nor the other. Thus, a small amount of error can be attributed to these decisions.

A third case was related to a border effect in the LULC predictions caused by the U-net design. When analyzing the probability of each pixel corresponding to each class, low probability values tended to concentrate on the border of each class polygons (Figure 5). Admittedly, this aspect could be related to the arbitrary decisions on the limits of the classes; however, it is also associated with the effect of down- and up-sampling filters in the U-net, which cause a degradation in the spatial resolution of the image in each pass. Although the U-net foresees this aspect, by using the skipped connections to add spatial detail to the final result, it might not be enough to provide results with the same resolution as the original input, as previous studies have reported [40,44,95,104]. Thus, as confirmed by the verification of the LULC classification in the complete study area, certain errors were associated with the limits of the polygons, either by an increased size of the polygon in the classification in comparison with the input images (e.g., larger clouds) or by a small spatial offset of the borders of the polygons (e.g., limits between roads and plantations).

The fourth condition could help explain the low accuracy observed for identifying rare classes such as aquatic vegetation or young plantations. Although CNN are capable of identifying rare classes, the number of observations in the training data might have been too small for the CNN to correctly extract general patterns to identify this class [49]. Nevertheless, the case of aquatic vegetation is further discussed in the final part of the discussion.

Finally, another minor source of error might reside on the difference in acquisition date of the VHR imagery and the Sentinel-1 and Sentinel-2 used to obtain the LULC classification. In most cases, this date difference was negligible, but for three points in the accuracy assessment procedure, the U-net predicted young plantations as its LULC class; however, no VHR information was available to help identify these areas as young plantations or not. In these rare cases, the areas were manually classified as grassland/agriculture or soil, depending exclusively on the Planet and Sentinel-2 images. Although this might be a very small source of error, a sparse annotations approach can help reduce this error because not all the areas included in the training data need to be labeled [95].

4.4. Comparisons with Similar Studies

Although, the overall F1-score and accuracy reported in our LULC classification is admittedly lower than previous LULC studies using the U-net [42–49,73,91,94,95,105–107], this difference can be easily explained by the combination of three factors: (1) a much more detailed classification system, (2) a much coarser spatial resolution imagery and (3) a smaller number of observations in the training set.

In the first case, other studies have used simpler classification systems for which the discrimination is easier to achieve (e.g., forest/nonforest systems) [43,45,47]. Although, these studies had different research interests, the spectral and spatial discrimination of the LULC classes will also affect the performance of the deep learning algorithm to solve certain tasks (Figure 6). Other studies have tackled this problem by using hierarchical classification systems and calculating the performance of the algorithm on each level [49]. A similar approach could be adopted in our classification system to obtain broader classes with higher F1 scores, e.g., a single plantations or forest class, instead of separating old-growth and young ones.

In the second situation, other studies have mainly relied on VHR imagery (e.g., Worldview or aerial images) to obtain higher accuracy (e.g., Worldview, aerial images) [41,42,45–47,73,91,105–107]. This type of imagery allows a better discrimination of LULC classes, in comparison with high or intermediate resolution imagery, by providing more detailed information; thus, the potential of the U-net is limited by the spatial resolution of the input imagery [19,31,41]. For example, in this study, the use of VHR imagery would have helped in distinguishing poorly identified classes such as young plantations and roads. In addition, other studies have enhanced the potential of the U-net by incorporating the temporal dimension into the convolutional filters [44,95], fusing multi-resolution inputs [94] or using customized U-net architectures [51]. These alternatives might be interesting for future studies interested in performing LULC classification with the U-net architecture.

In the third circumstance, it is frequently mentioned that the full potential of deep learning algorithms is especially evident when there is a large volume of training data [28,30]. Nevertheless, frequently in Earth observations applications the data available for training are limited, due to the large quantity of time and resources required to obtain this data [20,37]. Therefore, it is not surprising that augmentation techniques are frequently used. For example, other studies have used more aggressive augmentation techniques, such as rotation in different angles (90° , 180° , 270°), as well as brightness, saturation and hue alterations [41,42,46]. Although augmentation techniques might enhance the generalization capabilities of the algorithm, they are usually computationally expensive, and therefore, they need to be adjusted to the available resources. In this case, we opted for an augmentation scheme that enabled a fast training with the available computational resources, mainly to shorten the hyperparameter exploration procedure.

Previous studies have relied on transfer learning to compensate for the small size of the available training set to obtain higher accuracies with deep learning algorithms [35,108,109]. In this case, we did not use transfer learning because we included NIR and two SAR bands, while most pre-trained CNN use exclusively RGB imagery. Additionally, many of the pre-trained CNN use images with very different viewing geometry, resolution and classification systems than remote sensing ones [20,101]. Thus, we opted to train the U-net with only our training data.

Although it is clear that these three factors play a role in determining the performance of the U-net, the interactions among them are unclear. Future studies should address this topic in order to understand the effects of each factor over the classification performance.

4.5. Methodological Highlights

We consider that the use of the accuracy assessment protocol developed by [86,87] is a relatively common approach used to assess the accuracy of a map; however, in studies that use deep learning algorithms, it is rarely used (but see similar approaches in [41,42]). We are aware that if the verification set is large enough, with a random spatial distribution in the study area and has not been exposed to the CNN during the training phase (i.e., such as in the early stopping procedure), it should give similar estimates to the ones obtained by the abovementioned protocol. Nevertheless, a comparison of the F1-score obtained for each class in the validation and accuracy assessment procedures evidenced that for three classes—aquatic vegetation, roads and young plantations—its $\Delta F1$ -score was higher

than 0.2 (Table 2). This large difference in the F1-score can be related to two situations that artificially inflated the error for these three classes: (1) their small sample size in the validation dataset, and (2) their spatial location in the validation dataset.

In the first case, the effect of inflating the error to identify a class seems to be clearer in the rare classes, probably due to its limited number of observations. Because the validation data come from an augmented procedure, the data are not completely independent; thus, many observations are in fact a mirrored or a subsample of another. A direct consequence of this condition is that if a pixel was wrongly classified in one sample, it will be probably classified wrongly in an augmented version; thus, the F1-score obtained in the validation procedure can be artificially decreased, in comparison with the one obtained for the accuracy assessment procedure ($0.20 \leq \Delta F1\text{-score} \leq 0.25$).

In the second situation, after comparing the F1-score obtained for the validation and accuracy assessment procedures for aquatic vegetation ($\Delta F1\text{-score} = 0.55$), it was evident that another factor was also responsible for the large difference. Thus, we noticed that aquatic vegetation class was systematically found in edges of the validation data, an aspect that was overlooked in the training and validation data acquisition. Therefore, due to its location in the edges of the validation dataset, the U-net obtained low quality predictions for this class, which inflated the error calculated for this class in the validation dataset in comparison with the accuracy assessment procedure.

We consider that the accuracy assessment results give a more precise estimate of the precision of the U-net to obtain LULC maps because they evaluate the classification over the complete study area and reduce the error caused by the low-quality predictions in the edges of the tiles. In addition, using this protocol allows obtaining an estimate of each class surface with confidence limits. Thus, we recommend future studies to use similar protocols to evaluate LULC maps.

Finally, it is worth mentioning that we opted for making a LULC map using a single date image to maximize the agreement between the field information and the MS image (as SAR images do not detect clouds) in a very dynamic landscape, instead of maximizing LULC coverage. Although other studies have addressed this issue by using multitemporal composites, few studies have analyzed the effect of the image composition over the abilities of an algorithm to perform the LULC classification task (but see [110,111]), less still using CNN-based algorithms. These studies report that better classification accuracies are usually obtained with composites constructed from images acquired inside a small temporal window (e.g., a season within a year). Nevertheless, in areas with high cloud coverage throughout the year, as the one studied here, even using multitemporal composites might not ensure a high-quality composite or a complete study area coverage. Future studies should address this trade-off between temporal agreement and study area coverage, especially using CNN-based algorithms that consider the spatial context of a pixel to determine its class.

5. Conclusions

The use of CNN for Earth observation applications has further improved the capabilities to generate detailed LULC classifications. Nevertheless, it is essential to evaluate the role of different imagery inputs and algorithms in LULC mapping with special focus on discriminating among forested classes. In this study we found that although we trained the U-net with a small dataset, it outperformed the random forests algorithm. Additionally, the LULC map with the highest accuracy was achieved using the U-net with the MS + SAR bands as inputs, followed very closely by MS U-net and lastly by SAR U-net. Furthermore, MS + SAR U-net obtained higher F1-scores for similar LULC classes such as old-growth forests and plantations. We conclude that the better performance of the U-net, in comparison with the RF, is mainly because of the incorporation of the spatial and spectral features in the LULC classification. In addition, the combined use of MS + SAR imagery helps in obtaining a detailed LULC map and especially in discriminating among forested

classes. This study demonstrates the capabilities of CNN for obtaining detailed LULC classifications with medium spatial resolution satellite images.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/rs13183600/s1>, Table S1: U-net Hyperparameter exploration results.

Author Contributions: Conceptualization and writing—review and editing, J.V.S., J.F.M., Y.G. and J.A.G.-C.; methodology, J.V.S. and J.F.M.; software, validation, formal analysis, data curation, investigation, visualization, project administration and writing—original draft preparation J.V.S.; supervision, J.F.M., Y.G., J.A.G.-C.; resources and funding acquisition, J.F.M. and J.A.G.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Project PAPIME PE117519 “Herramientas para la enseñanza de la Geomática con programas de código abierto” (Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México) and the Universidad Iberoamericana 14th DINV grant.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The first author wishes to thank the Consejo Nacional de Ciencia y Tecnología (CONACyT) for providing a PhD scholarship. The authors are also grateful to the project “Herramientas para la Enseñanza de la Geomática con programas de Código Abierto”, Programa de Apoyo a Proyectos para la Innovación y Mejoramiento de la Enseñanza (PAPIME # PE117519, <http://lae.ciga.unam.mx/proyectos/geomatica/index.php> accessed on 20 April 2021) for its support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

Table A1. Confusion matrix for the validation set using the MS + SAR U-net. Rows show predicted classes while column, ground truth classes. LULC abbreviations: AV, aquatic vegetation; G/A, grasslands/agriculture; HS, human settlements; OF, old-growth forests; OP, old-growth plantations; R, roads; SF, secondary forest; So, soil, W, water; YP, young plantations. User acc stands for user’s accuracy, while Prod acc for producer’s accuracy.

	AV	G/A	HS	OF	OP	R	SF	So	W	YP	User acc	
AV	594	115	0	240	0	0	0	0	0	0	0.63	
G/A	0	902,138	11,583	35,162	25,472	23,834	64,813	88,484	322	22,190	0.76	
HS	0	6917	159,892	627	1278	2810	2010	1926	0	13	0.91	
OF	6001	55,782	5081	1,372,026	116,552	3245	68,833	3980	1595	3938	0.82	
OP	0	22,157	238	26,673	246,789	3790	15,303	1667	128	26,695	0.72	
R	0	8587	3488	895	2364	21,636	986	10,066	0	653	0.44	
SF	0	67,145	8288	67,390	49,581	2463	143,531	1287	0	5531	0.41	
So	21	48,149	1745	3046	1413	12,314	1363	181,030	1041	2280	0.69	
W	170	855	1	775	44	0	9	272	59,215	0	0.97	
YP	0	8338	197	164	12,528	927	1104	2668	0	5159	0.17	
Prod acc	0.09	0.80	0.84	0.91	0.54	0.30	0.48	0.61	0.95	0.08		
											Overall accuracy	0.76
											Batch avgF1-score	0.72
											Overall avgF1-score	0.58

Table A2. Confusion matrix for the validation set using the MS U-net. Rows show predicted classes while column, ground truth classes. LULC abbreviations: AV, aquatic vegetation; G/A, grasslands/agriculture; HS, human settlements; OF, old-growth forests; OP, old-growth plantations; R, roads; SF, secondary forest; So, soil, W, water; YP, young plantations. User acc stands for user's accuracy, while Prod acc for producer's accuracy.

	AV	G/A	HS	OF	OP	R	SF	So	W	YP	User Acc		
AV	397	327	0	25	8	0	62	0	0	0	0.48		
G/A	109	945,076	12,870	38,984	58,029	24,740	105,371	103,450	45	33,760	0.71		
HS	0	8727	171,136	1782	4255	5038	4903	5070	0	130	0.85		
OF	5497	57,945	2401	1,386,121	136,575	3906	71,614	3827	324	14,346	0.81		
OP	1	23,140	381	34,138	206,132	2072	23,195	1857	0	10,017	0.68		
R	0	7045	1256	1393	1861	26,717	1382	11,403	0	561	0.52		
SF	317	43,758	1888	43,919	45,045	870	89,761	338	1	5930	0.39		
So	13	32,971	771	1631	899	8216	768	162,686	536	697	0.76		
W	451	1104	0	916	82	0	38	409	61,225	0	0.95		
YP	0	2316	1	171	3178	129	368	1180	0	1001	0.12		
Prod acc	0.06	0.84	0.90	0.91	0.45	0.37	0.30	0.55	0.98	0.02			
											Overall accuracy	0.76	
												Batch avgF1-score	0.72
												Overall avgF1-score	0.55

Table A3. Confusion matrix for the validation set using the SAR U-net. Rows show predicted classes while column, ground truth classes. LULC abbreviations: AV, aquatic vegetation; G/A, grasslands/agriculture; HS, human settlements; OF, old-growth forests; OP, old-growth plantations; R, roads; SF, secondary forest; So, soil, W, water; YP, young plantations. User acc stands for user's accuracy, while Prod acc for producer's accuracy.

	AV	G/A	HS	OF	OP	R	SF	So	W	YP	User Acc		
AV	0	0	0	0	0	0	0	0	0	0	0.00		
G/A	1	944,129	26,365	115,271	50,653	48,227	103,875	199,658	124	33,311	0.62		
HS	0	14,585	79,076	19,238	23,998	2662	7312	5583	0	3997	0.51		
OF	6561	150,845	70,460	1,512,101	183,508	13,216	138,688	24,125	1727	6467	0.72		
OP	1	25,098	12,542	36,230	182,438	6766	14,267	12,689	64	17,395	0.59		
R	0	103	16	1	32	7	1	30	0	6	0.04		
SF	0	41,897	2610	50,157	18,848	2383	48959	2739	0	2351	0.29		
So	1	16,572	514	2610	2725	1690	1030	60,535	552	2876	0.68		
W	222	1014	0	1291	25	0	4	3088	59,834	0	0.91		
YP	0	2382	150	77	2014	202	201	463	0	203	0.04		
Prod acc	0.00	0.79	0.41	0.87	0.39	0.00	0.16	0.20	0.96	0.00			
											Overall accuracy	0.65	
												Batch avgF1-score	0.57
												Overall avgF1-score	0.39

Table A4. Confusion matrix for the validation set using the MS + SAR random forests. Rows show predicted classes while column, ground truth classes. LULC abbreviations: AV, aquatic vegetation; G/A, grasslands/agriculture; HS, human settlements; OF, old-growth forests; OP, old-growth plantations; R, roads; SF, secondary forest; So, soil, W, water; YP, young plantations. User acc stands for user's accuracy, while Prod acc for producer's accuracy.

	AV	G/A	HS	OF	OP	R	SF	So	W	YP	User Acc
AV	551	3068	133	19,409	615	22	3474	6	0	174	0.02
G/A	16	93,773	2459	3373	1220	1385	5325	6354	12	1725	0.80
HS	2	10,147	9104	1760	539	1512	1185	2672	0	266	0.28
OF	296	1180	322	121,083	12,033	29	4389	10	0	411	0.85
OP	56	2552	1451	17,256	23,980	129	4370	126	1	2065	0.46
R	0	12,179	6247	525	620	5367	627	7774	3	229	0.15

Table A4. Cont.

	AV	G/A	HS	OF	OP	R	SF	So	W	YP	User Acc	
SF	56	20,566	2804	37,231	13,487	470	24,686	288	0	2146	0.24	
So	6	12,449	2015	322	244	1110	162	26,977	8	542	0.59	
W	73	439	1	215	16	0	67	127	8542	4	0.90	
YP	1	19,031	4343	3747	6912	933	3381	3575	0	3211	0.07	
Prod acc	0.49	0.53	0.30	0.58	0.40	0.48	0.51	0.55	0.99	0.29		
											Overall accuracy	0.53
											Batch avgF1-score	-
											Overall avgF1-score	0.43

Table A5. Confusion matrix resulting from the accuracy assessment procedure of the complete study area LULC classification obtained with the MS + SAR U-net. Rows show predicted classes while columns, ground truth classes. LULC abbreviations: AV, aquatic vegetation; G/A, grasslands/agriculture; HS, human settlements; OF, old-growth forests; OP, old-growth plantations; R, roads; SF, secondary forest; So, soil, W, water; YP, young plantations. User acc stands for user's accuracy, while Prod acc for producer's accuracy.

	AV	G/A	HS	OF	OP	R	SF	So	W	YP	User Acc	
AV	15	0	0	3	0	0	0	0	0	0	0.83	
G/A	3	100	0	4	2	1	1	5	0	5	0.75	
HS	0	0	22	0	0	0	0	0	0	0	1.00	
OF	4	2	0	113	2	0	2	0	0	0	0.88	
OP	0	1	0	4	18	1	3	0	0	9	0.46	
R	0	0	1	0	1	12	0	0	0	0	0.80	
SF	2	9	0	5	1	1	18	0	1	1	0.44	
So	0	5	1	1	0	10	0	20	0	4	0.42	
W	0	0	0	0	0	0	0	0	23	0	1.00	
YP	1	3	1	0	1	0	0	0	0	6	0.46	
Prod acc	0.60	0.83	0.88	0.86	0.72	0.48	0.72	0.80	0.92	0.24		
											Overall accuracy	0.77
											Batch avgF1-score	-
											Overall avgF1-score	0.68

References

1. Aplin, P. Remote sensing: Land cover. *Prog. Phys. Geogr.* **2004**, *28*, 283–293. [\[CrossRef\]](#)
2. Giri, C.P. Remote Sensing of Land Use and Land Cover. In *Principles and Applications*; CRC Press: Boca Raton, FL, USA, 2020; p. 477.
3. Treitz, P.; Rogan, J. Remote sensing for mapping and monitoring land-cover and land-use change—An introduction. *Prog. Plan.* **2004**, *61*, 269–279. [\[CrossRef\]](#)
4. Congalton, R.G.; Gu, J.; Yadav, K.; Thenkabail, P.; Ozdogan, M. Global land cover mapping: A review and uncertainty analysis. *Remote Sens.* **2014**, *6*, 12070–12093. [\[CrossRef\]](#)
5. Gómez, C.; White, J.C.; Wulder, M.A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS J. Photogramm. Remote Sens.* **2018**, *10*, 55–72. [\[CrossRef\]](#)
6. Rogan, J.; Chen, D.M. Remote sensing technology for mapping and monitoring land-cover and land-use change. *Prog. Plan.* **2004**, *61*, 301–325. [\[CrossRef\]](#)
7. Joshi, N.; Baumann, M.; Ehammer, A.; Fensholt, R.; Grogan, K.; Hostert, P.; Jepsen, M.R.; Kuemmerle, T.; Meyfroidt, P.; Mitchard, E.T.A.; et al. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sens.* **2016**, *8*, 70. [\[CrossRef\]](#)
8. Gutiérrez-Vélez, V.H.; DeFries, R.; Pinedo-Vásquez, M.; Uriarte, M.; Padoch, C.; Baethgen, W.; Fernandes, K.; Lim, Y. High-yield oil palm expansion spares land at the expense of forests in the Peruvian Amazon. *Environ. Res. Lett.* **2011**, *6*, 44029. [\[CrossRef\]](#)
9. Lee, J.S.H.; Wich, S.; Widayati, A.; Koh, L.P. Detecting industrial oil palm plantations on Landsat images with Google Earth Engine. *Remote Sens. Appl. Soc. Environ.* **2016**, *4*, 219–224. [\[CrossRef\]](#)
10. Mercier, A.; Betbeder, J.; Rumiano, F.; Baudry, J.; Gond, V.; Blanc, L.; Bourgoin, C.; Cornu, G.; Ciudad, C.; Marchamalo, M.; et al. Evaluation of Sentinel-1 and 2 Time Series for Land Cover Classification of Forest-Agriculture Mosaics in Temperate and Tropical Landscapes. *Remote Sens.* **2019**, *11*, 979. [\[CrossRef\]](#)
11. Poortinga, A.; Tenneson, K.; Shapiro, A.; Nguyen, Q.; Aung, K.S.; Chishtie, F.; Saah, D. Mapping plantations in Myanmar by fusing Landsat-8, Sentinel-2 and Sentinel-1 data along with systematic error quantification. *Remote Sens.* **2019**, *11*, 831. [\[CrossRef\]](#)

12. Tropek, R.; Sedláček, O.; Beck, J.; Keil, P.; Musilová, Z.; Šímová, I.; Storch, D. Comment on “High-resolution global maps of 21st-century forest cover change”. *Science* **2014**, *344*, 981. [[CrossRef](#)] [[PubMed](#)]
13. Gibson, L.; Lee, T.M.; Koh, L.P.; Brook, B.W.; Gardner, T.A.; Barlow, J.; Peres, C.A.; Bradshaw, C.J.A.; Laurance, W.F.; Lovejoy, T.E.; et al. Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature* **2011**, *478*, 378–381. [[CrossRef](#)] [[PubMed](#)]
14. Singh, D.; Slik, J.W.F.; Jeon, Y.S.; Tomlinson, K.W.; Yang, X.; Wang, J.; Kerfahi, D.; Porazinska, D.L.; Adams, J.M. Tropical forest conversion to rubber plantation affects soil micro- & mesofaunal community & diversity. *Sci. Rep.* **2019**, *9*, 1–13. [[CrossRef](#)]
15. Wright, S.J. Tropical forests in a changing environment. *Trends Ecol. Evol.* **2005**, *20*, 553–560. [[CrossRef](#)] [[PubMed](#)]
16. Zhang, D.; Stanturf, J. Forest plantations. In *Encyclopedia of Ecology; Ecosystems*; Jørgensen, S.E., Fath, B.D., Eds.; Academic Press: Oxford, UK, 2008; pp. 1673–1680.
17. Carlson, K.M.; Curran, L.M.; Asner, G.P.; Pittman, A.M.D.; Trigg, S.N.; Marion Adeney, J. Carbon emissions from forest conversion by Kalimantan oil palm plantations. *Nat. Clim. Chang.* **2013**, *3*, 283–287. [[CrossRef](#)]
18. Guo, L.B.; Gifford, R.M. Soil carbon stocks and land use change: A meta analysis. *Glob. Chang. Biol.* **2002**, *8*, 345–360. [[CrossRef](#)]
19. Datcu, M.; Schwarz, G.; Dumitru, C.O. Deep Learning Training and Benchmarks for Earth Observation Images: Data Sets, Features, and Procedures. In *Recent Trends in Artificial Neural Networks. From Training to Prediction*; Sadollah, A., Ed.; InTech Open: London, UK, 2020.
20. Hoese, T.; Kuenzer, C. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. *Remote Sens.* **2020**, *12*, 1667. [[CrossRef](#)]
21. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine learning in geosciences and remote sensing. *Geosci. Front.* **2016**, *7*, 3–10. [[CrossRef](#)]
22. Melesse, A.M.; Weng, Q.; Thenkabail, P.S.; Senay, G.B. Remote Sensing Sensors and Applications in Environmental Resources Mapping and Modelling. *Sensors* **2007**, *7*, 3209–3241. [[CrossRef](#)]
23. Tang, X.; Bullock, E.L.; Olofsson, P.; Estel, S.; Woodcock, C.E. Near real-time monitoring of tropical forest disturbance: New algorithms and assessment framework. *Remote Sens. Environ.* **2019**, *224*, 202–218. [[CrossRef](#)]
24. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [[CrossRef](#)]
25. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [[CrossRef](#)]
26. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [[CrossRef](#)]
27. Buduma, N. *Fundamentals of Deep Learning*; O’Reilly: Boston, MA, USA, 2017; p. 283.
28. Chollet, F.; Allaire, J.J.; Planet Team. *Deep Learning with R*; Manning Publications Co.: Shelter Island, NY, USA, 2018; p. 335.
29. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; p. 787.
30. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
31. Hoese, T.; Bachofer, F.; Kuenzer, C. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review—Part II: Applications. *Remote Sens.* **2020**, *12*, 3053. [[CrossRef](#)]
32. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in Vegetation Remote Sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [[CrossRef](#)]
33. Liu, X.; Han, F.; Ghazali, K.; Mohamed, I.; Zhao, Y. A review of Convolutional Neural Networks in Remote Sensing Image. In Proceedings of the ICSCA 2019 8th International Conference on Software and Computer Applications, Penang, Malaysia, 19–21 February 2019; pp. 263–267.
34. Luo, C.; Huang, H.; Wang, Y.; Wang, S. Utilization of Deep Convolutional Neural Networks for Remote Sensing Scenes Classification. In *Advanced Remote Sensing Technology for Synthetic Aperture Radar Applications, Tsunami Disasters, and Infrastructure*; Marghany, M., Ed.; IntechOpen: London, UK, 2019; pp. 1–18.
35. Ball, J.E.; Anderson, D.T.; Chan, C.S. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [[CrossRef](#)]
36. Pelletier, C.; Webb, G.I.; Petitjean, F. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sens.* **2019**, *11*, 523. [[CrossRef](#)]
37. Yuan, X.; Shi, J.; Gu, L. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Syst. Appl.* **2021**, *169*, 114417. [[CrossRef](#)]
38. Zhang, C.; Sargent, I.; Pan, X.; Li, H.; Gardiner, A.; Hare, J.; Atkinson, P.M. Joint Deep Learning for land cover and land use classification. *Remote Sens. Environ.* **2019**, *221*, 173–187. [[CrossRef](#)]
39. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [[CrossRef](#)]
40. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science In Medical Image Computing and Computer-Assisted Intervention, International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [[CrossRef](#)]
41. Clark, A.; McKechnie, J. Detecting banana plantations in the wet tropics, Australia, using aerial photography and U-net. *Appl. Sci.* **2020**, *10*, 1–15. [[CrossRef](#)]

42. Du, L.; McCarty, G.W.; Zhang, X.; Lang, M.W.; Vanderhoof, M.K.; Li, X.; Huang, C.; Lee, S.; Zou, Z. Mapping Forested Wetland Inundation in the Neural Networks. *Remote Sens.* **2020**, *12*, 644. [CrossRef]
43. Flood, N.; Watson, F.; Collett, L. Using a U-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across Queensland, Australia. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *82*, 101897. [CrossRef]
44. Isaienkov, K.; Yushchuk, M.; Khramtsov, V.; Seliverstov, O. Deep Learning for Regular Change Detection in Ukrainian Forest Ecosystem with Sentinel-2. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 1–15. [CrossRef]
45. Neves, A.K.; Körting, T.S.; Fonseca, L.M.G.; Neto, C.D.G.; Wittich, D.; Costa, G.A.O.P.; Heipke, C. Semantic segmentation of Brazilian savanna vegetation using high spatial resolution satellite data and U-net. In Proceedings of the 2020 XXIV ISPRS Congress (2020 Edition), Nice, France, 31 August–2 September 2020; pp. 505–511.
46. Wagner, F.H.; Sanchez, A.; Aidar, M.P.M.; Rochelle, A.L.C.; Tarabalka, Y.; Fonseca, M.G.; Phillips, O.L.; Gloor, E.; Aragão, L.E.O.C. Mapping Atlantic rainforest degradation and regeneration history with indicator species using convolutional network. *PLoS ONE* **2020**, *15*, e0229448. [CrossRef] [PubMed]
47. Wagner, F.H.; Sanchez, A.; Tarabalka, Y.; Lotte, R.G.; Ferreira, M.P.; Aidar, M.P.M.; Gloor, E.; Phillips, O.L.; Aragão, L.E.O.C. Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. *Remote Sens. Ecol. Conserv.* **2019**, *5*, 360–375. [CrossRef]
48. Yi, Y.; Zhang, Z.; Zhang, W.; Zhang, C.; Li, W.; Zhao, T. Semantic Segmentation of Urban Buildings from VHR Remote Sensing Imagery Using a Deep Convolutional Neural Network. *Remote Sens.* **2019**, *11*, 1774. [CrossRef]
49. Ulmas, P.; Liiv, I. Segmentation of satellite imagery using U-Net models for land cover classification. *arXiv* **2020**, arXiv:2003.02899.
50. Baek, J.; Kim, J.W.; Lim, G.J.; Lee, D.-C. Electromagnetic land surface classification through integration of optical and radar remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1214–1222. [CrossRef]
51. Gargiulo, M.; Dell’aglio, D.A.G.; Iodice, A.; Riccio, D.; Ruello, G. Integration of sentinel-1 and sentinel-2 data for land cover mapping using w-net. *Sensors* **2020**, *20*, 2969. [CrossRef] [PubMed]
52. Heckel, K.; Urban, M.; Schratz, P.; Mahecha, M.D.; Schullius, C. Predicting Forest Cover in Distinct Ecosystems: The Potential of Multi-Source Sentinel-1 and -2 Data Fusion. *Remote Sens.* **2020**, *12*, 302. [CrossRef]
53. Zhang, Y.; Ling, F.; Foody, G.M.; Ge, Y.; Boyd, D.S.; Li, X.; Du, Y.; Atkinson, P.M. Mapping annual forest cover by fusing PALSAR/PALSAR-2 and MODIS NDVI during 2007–2016. *Remote Sens. Environ.* **2019**, *224*, 74–91. [CrossRef]
54. Morel, A.; Saatchi, S.; Malhi, Y.; Berry, N.; Banin, L.; Burslem, D.; Nilus, R.; Ong, R. Estimating aboveground biomass in forest and oil palm plantation in Sabah, Malaysian Borneo using ALOS PALSAR data. *For. Ecol. Manag.* **2011**, *262*, 1786–1798. [CrossRef]
55. Malenovsky, Z.; Rott, H.; Cihlar, J.; Schaepman, M.E.; García-Santos, G.; Fernandes, R.; Berger, M. Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sens. Environ.* **2012**, *120*, 91–101. [CrossRef]
56. Carabias, J.; De la Maza, J.; Cadena, R. El escenario natural y social. In *Conservación y Desarrollo Sustentable en la Selva Lacandonia. 25 Años de Actividades y Experiencias*; Carabias, J., De la Maza, J., Cadena, R., Eds.; Natura y Ecosistemas Mexicanos A.C: Mexico City, Mexico, 2015; pp. 16–18.
57. Mendoza, E.; Dirzo, R. Deforestation in Lacandonia (Southeast Mexico): Evidence for the declaration of the northernmost tropical hot-spot. *Biodivers. Conserv.* **1999**, *8*, 1621–1641. [CrossRef]
58. Castillo-Santiago, M.A.; Hellier, A.; Tipper, R.; De Jong, B.H.J. Carbon emissions from land-use change: An analysis of causal factors in Chiapas, Mexico. *Mitig. Adapt. Strateg. Glob. Chang.* **2007**, *12*, 1213–1235. [CrossRef]
59. Fernández-Montes de Oca, A.I.; Gallardo-Cruz, A.; Ghilardi, A.; Kauffer, E.; Solórzano, J.V.; Sánchez-Cordero, V. An integrated framework for harmonizing definitions of deforestation. *Environ. Sci. Policy* **2021**, *115*, 71–78. [CrossRef]
60. Vaca, R.A.; Golicher, D.J.; Cayuela, L.; Hewson, J.; Steininger, M. Evidence of incipient forest transition in Southern Mexico. *PLoS ONE* **2012**, *7*, e42309. [CrossRef] [PubMed]
61. Cassol, H.L.; Shimabukuro, Y.E.; Beuchle, R.; Aragão, L.E.O.C. Sentinel-1 Time-Series Analysis for Detection of Forest Degradation By Selective Logging. In Proceedings of the Anais do XIX Simpósio Brasileiro de Sensoriamento Remoto, São José dos Campos, São José dos Campos, Brazil, 14–17 April 2019; pp. 1–4.
62. Small, D. Flattening gamma: Radiometric terrain correction for SAR imagery. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3081–3093. [CrossRef]
63. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
64. Planet Team. *Planet Application Program Interface: In Space for Life on Earth*; Planet Team: San Francisco, CA, USA, 2017; Available online: <https://api.planet.com> (accessed on 2 September 2021).
65. Google. Google Satellite Images. 2021. Available online: <http://www.google.cn/maps/vt?lyrs=s@189&gl=cn&x=\{x\}&y=\{y\}&z=\{z\}> (accessed on 7 September 2021).
66. Yandex. Yandex Satellite Images. 2021. Available online: https://core-sat.maps.yandex.net/tiles?l=sat&v=3.564.0&x=\{x\}&y=\{y\}&z=\{z\}&scale=1&lang=ru_RU (accessed on 7 September 2021).
67. Bing. Bing Satellite Images. 2021. Available online: http://ecn.t3.tiles.virtualearth.net/tiles/a\{q\}.jpeg?g=0&dir=dir_n\T1\textquoteright (accessed on 7 September 2021).
68. QGIS Development Team. QGIS Geographic Information System 3.16; Open Source Geospatial Foundation. 2021. Available online: https://docs.qgis.org/3.16/en/docs/user_manual/ (accessed on 5 June 2021).

69. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
70. Hijmans, R.J. Raster: Geographic Data Analysis and Modeling. 2020. Available online: <https://cran.r-project.org/web/packages/raster/index.html> (accessed on 1 September 2021).
71. Vaughan, D. Rray: Simple Arrays. 2020. Available online: <https://github.com/r-lib/rRAY> (accessed on 1 September 2021).
72. Ushey, K.; Allaire, J.J.; Tang, Y. Reticulate: Interface to 'Python'. 2020. Available online: <https://cran.r-project.org/web/packages/reticulate/index.html> (accessed on 1 September 2021).
73. Hamdi, Z.M.; Brandmeier, M.; Straub, C. Forest Damage Assessment Using Deep Learning on High Resolution Remote Sensing Data. *Remote Sens.* **2019**, *11*, 1976. [[CrossRef](#)]
74. Allaire, J.; Chollet, F. Keras: R Interface to 'Keras'. 2018. Available online: <https://cran.r-project.org/web/packages/keras/index.html> (accessed on 1 September 2021).
75. Falbel, D.; Zak, K. Unet: U-Net: Convolutional Networks for Biomedical Image Segmentation. 2020. Available online: <https://github.com/r-tensorflow/unet> (accessed on 1 September 2021).
76. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A System for Large-Scale Machine Learning. In Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
77. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958. [[CrossRef](#)]
78. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015; pp. 1–13, arXiv:1412.6980.
79. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1026–1034.
80. Goutte, C.; Gaussier, E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. *Lect. Notes Comput. Sci.* **2005**, *3408*, 345–359. [[CrossRef](#)]
81. Allaire, J.J. Tfruns: Training Run Tools for 'TensorFlow'. 2018. Available online: <https://cran.r-project.org/web/packages/tfruns/index.html> (accessed on 1 September 2021).
82. Talukdar, S.; Singha, P.; Mahato, S.; Shahfahad, P.S.; Liou, Y.A.; Rahman, A. Land-use land-cover classification by machine learning classifiers for satellite observations. A review. *Remote Sens.* **2020**, *12*, 1135. [[CrossRef](#)]
83. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
84. Sawangarereak, S.; Thanathamthee, P. Random forest with sampling techniques for handling imbalanced prediction of university student depression. *Information* **2020**, *11*, 519. [[CrossRef](#)]
85. Liaw, A.; Wiener, M. Classification and Regression by randomForest. 2002. Available online: <https://cran.r-project.org/web/packages/randomForest/index.html> (accessed on 1 September 2021).
86. Cochran, W.G. *Sampling Techniques*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 1977.
87. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [[CrossRef](#)]
88. Card, D.H. Using Known Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy. *Photogramm. Eng. Remote Sens.* **1982**, *48*, 431–439.
89. FAO (Food and Agriculture Organization). Openforis Accuracy Assessment Tool. 2017. Available online: <https://github.com/openforis/accuracy-assessment> (accessed on 1 September 2021).
90. De Bem, P.P.; de Carvalho, O.A.; Guimarães, R.F.; Gomes, R.A.T. Change detection of deforestation in the Brazilian Amazon using Landsat data and convolutional neural networks. *Remote Sens.* **2020**, *12*, 901. [[CrossRef](#)]
91. Giang, T.L.; Dang, K.B.; Le, Q.T.; Nguyen, V.G.; Tong, S.S.; Pham, V.-M. U-Net Convolutional Networks for Mining Land Cover Classification Based on High-Resolution UAV Imagery. *IEEE Access* **2020**, *8*, 186257–186273. [[CrossRef](#)]
92. Ienco, D.; Interdonato, R.; Gaetano, R.; Ho Tong Minh, D. Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 11–22. [[CrossRef](#)]
93. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [[CrossRef](#)]
94. Robinson, C.; Hou, L.; Malkin, K.; Soobitsky, R.; Czawlytko, J.; Dilkina, B.; Jojic, N. Large scale high-resolution land cover mapping with multi-resolution data. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2019**, *2019*, 12718–12727. [[CrossRef](#)]
95. Stoian, A.; Poulain, V.; Inglada, J.; Poughon, V.; Derksen, D. Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. *Remote Sens.* **2019**, *11*, 1986. [[CrossRef](#)]
96. Hirschmugl, M.; Deutscher, J.; Sobe, C.; Bouvet, A.; Mermoz, S.; Schardt, M. Use of SAR and Optical Time Series for Tropical Forest Disturbance Mapping. *Remote Sens.* **2020**, *12*, 727. [[CrossRef](#)]
97. Khan, A.; Govil, H.; Kumar, G.; Dave, R. Synergistic use of Sentinel-1 and Sentinel-2 for improved LULC mapping with special reference to bad land class: A case study for Yamuna River floodplain, India. *Spat. Inf. Res.* **2020**, *28*, 669–681. [[CrossRef](#)]

98. Tavares, P.A.; Beltrão, N.E.S.; Guimarães, U.S.; Teodoro, A.C. Integration of sentinel-1 and sentinel-2 for classification and LULC mapping in the urban area of Belém, eastern Brazilian Amazon. *Sensors* **2019**, *19*, 1140. [[CrossRef](#)]
99. Van Tricht, K.; Gobin, A.; Gilliams, S.; Piccard, I. Synergistic use of radar sentinel-1 and optical sentinel-2 imagery for crop mapping: A case study for Belgium. *Remote Sens.* **2018**, *10*, 1642. [[CrossRef](#)]
100. Flores-Anderson, A.I.; Herndon, K.E.; Thapa, R.B.; Cherrington, E. *The SAR Handbook. Comprehensive Methodologies for Forest Monitoring and Biomass Estimation*; NASA: Huntsville, AL, USA, 2019. [[CrossRef](#)]
101. Li, Y.; Zhang, H.; Xue, X.; Jiang, Y.; Shen, Q. Deep learning for remote sensing image classification: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, 1–17. [[CrossRef](#)]
102. Cheng, G.; Wang, Y.; Xu, S.; Wang, H.; Xiang, S.; Pan, C. Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3322–3337. [[CrossRef](#)]
103. McRoberts, R.E.; Stehman, S.V.; Liknes, G.C.; Næsset, E.; Sannier, C.; Walters, B.F. The effects of imperfect reference data on remote sensing-assisted estimators of land cover class proportions. *ISPRS J. Photogramm. Remote Sens.* **2018**, *142*, 292–300. [[CrossRef](#)]
104. He, H.; Yang, D.; Wang, S.; Wang, S.; Li, Y. Road Extraction by Using Atrous Spatial Pyramid Pooling Integrated Encoder-Decoder Network and Structural Similarity Loss. *Remote Sens.* **2019**, *11*, 1015. [[CrossRef](#)]
105. Huang, B.; Lu, K.; Audebert, N.; Khaleel, A.; Tarabalka, Y.; Malof, J.; Boulch, A.; Saux, B.L.; Collins, L.; Bradbury, K.; et al. Large-scale semantic classification: Outcome of the first year of inria aerial image labeling benchmark. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)—IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1–4.
106. Wagner, F.H.; Dalagnol, R.; Casapia, X.T.; Streher, A.S.; Phillips, O.L.; Gloor, E.; Aragão, L.E.O.C. Regional Mapping and Spatial Distribution Analysis of Canopy Palms in an Amazon Forest Using Deep Learning and VHR Images. *Remote Sens.* **2020**, *23*, 2225. [[CrossRef](#)]
107. Zhang, P.; Ke, Y.; Zhang, Z.; Wang, M.; Li, P.; Zhang, S. Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery. *Sensors* **2018**, *18*, 3717. [[CrossRef](#)]
108. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; Hasan, M.; Van Essen, B.C.; Awwal, A.A.S.; Asari, V.K. A state-of-the-art survey on deep learning theory and architectures. *Electronics* **2019**, *8*, 292. [[CrossRef](#)]
109. Zhao, H.; Liu, F.; Zhang, H.; Liang, Z. Convolutional neural network based heterogeneous transfer learning for remote-sensing scene classification. *Int. J. Remote Sens.* **2019**, *40*, 8506–8527. [[CrossRef](#)]
110. Praticò, S.; Solano, F.; Di Fazio, S.; Modica, G. Machine learning classification of mediterranean forest habitats in google earth engine based on seasonal sentinel-2 time-series and input image composition optimisation. *Remote Sens.* **2021**, *13*, 586. [[CrossRef](#)]
111. Chuvieco, E.; Ventura, G.; Martín, M.P. AVHRR multitemporal compositing techniques for burned land mapping. *Int. J. Remote Sens.* **2005**, *26*, 1013–1018. [[CrossRef](#)]