Revista Brasileira de Cartografia
ISSN 1808-0936 | https://doi.org/10.14393/revbrascartogr
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto

Deep Learning para Classificação Supervisionada de Imagens CBERS-4A da Área Urbana de Rio Claro (SP)

Deep Learning for Supervised Classification of CBERS-4A Imagery in Rio Claro (SP) Urban Area

Danilo Marques de Magalhães[1], Julya Paes de Souza[2], Edgar Auler Galvão de França[3]

1 Universidade Estadual Paulista "Júlio de Mesquita Filho", Departamento de Geografia e Planejamento Ambiental, Rio Claro (SP), Brasil. danilo.magalhaes@unesp.br
ORCID: https://orcid.org/0000-0001-9306-4326
2 Universidade Estadual Paulista "Júlio de Mesquita Filho", Departamento de Engenharia Ambiental, Rio Claro (SP), Brasil. julya.paes@unesp.br
ORCID: https://orcid.org/0009-0005-4162-0024
3 Universidade Estadual Paulista "Júlio de Mesquita Filho", Departamento de Estatística Matemática Aplicada e Computação, Rio Claro (SP), Brasil. Edgar.galvao@unesp.br
ORCID: https://orcid.org/0009-0002-5625-8011

Recebido: 09.2024 | Aceito: 04.2025

Resumo: O presente artigo tem por objetivo avaliar a acurácia do mapeamento do uso e cobertura da terra, realizado em um trecho da área urbana de Rio Claro (SP), a partir de técnicas de *Deep Learning* e utilizando uma imagem CBERS-4A (WPM) com 2 m de resolução espacial. Foi estruturada uma rede neural convolucional U-Net a partir de script em Python, utilizando as bibliotecas *Keras* e *Tensor Flow*. A verdade terrestre, utilizada para treinamento e verificação da acurácia do modelo, foi elaborada por meio de classificação supervisionada da mesma imagem no software ArcGIS Pro, utilizando o algoritmo *Support Vector Machine* (SVM) e procedimentos de pós-classificação, incluindo aplicação de filtro majoritário e edição manual de pixels. O resultado obtido pela U-Net foi comparado ao resultado obtido pelo SVM (sem pós-classificação), visando compreender se há ganhos de acurácia, tendo em vista o maior esforço humano, para a criação do *ground truth,* e computacional, para processamento dos dados, inerente às técnicas de *Deep Learning*. Para isso, ambos os resultados foram submetidos a avaliação de acurácia utilizando as métricas *Overall Accuracy, Precision, Recall, F1 Score e Kappa.* Constatou-se que o modelo U-Net apresenta melhor acurácia em todas elas, destacando-se o aumento da *Precision* de 0,48 (SVM) para 0,78 (U-Net). Tais resultados indicam o potencial das técnicas de *Deep Learning* para o mapeamento do uso e cobertura da terra em áreas urbanas a partir de imagens de alta resolução, o que pode contribuir, de modo significativo, para ações de planejamento e gestão territorial nos municípios brasileiros.
Palavras-chave: Deep Learning. CBERS-4A. Uso e Cobertura da Terra. U-Net. Classificação Supervisionada.

Abstract: This study evaluates the accuracy of land use and land cover (LULC) mapping in an urban area of Rio Claro (SP) using Deep Learning techniques and a CBERS-4A (WPM) image with 2 m spatial resolution. A U-Net convolutional neural network was developed using Python and the Keras and TensorFlow libraries. Ground truth data for training and accuracy assessment were generated through supervised classification of the same image in ArcGIS Pro, employing the Support Vector Machine (SVM) algorithm, followed by post-classification procedures, including majority filtering and manual pixel editing. U-Net results were compared with SVM results (pre-refinement) to evaluate potential accuracy improvements associated with the greater computational and human effort required by Deep Learning techniques. Both approaches were assessed using Overall Accuracy, Precision, Recall, F1 Score, and Kappa metrics. The U-Net model demonstrated superior performance across all metrics, with a notable increase in Precision from 0.48 (SVM) to 0.78 (U-Net). These findings highlight the potential of Deep Learning methods for high-resolution urban LULC mapping, providing valuable tools for urban planning and territorial management in Brazilian municipalities.
Keywords: Deep Learning. CBERS-4A. Land Use and Land Cover. U-Net. Supervised Classification.

1 INTRODUCTION

Land use and land cover (LULC) are widely recognized as essential variables for conducting environmental assessments and supporting territorial planning at various scales (Pabi, 2007; Yu et al., 2022). Since the launch of the first Landsat satellite in 1972, computational resources have been employed to classify identifiable features in satellite imagery to map LULC patterns. These innovations have enabled large-scale mapping with increased efficiency and satisfactory accuracy, while also opening possibilities for multitemporal monitoring of landscape changes (Odenyo & Pettry, 1977).

Currently, the advanced technological and scientific state of Remote Sensing allows for the identification and mapping of terrestrial features and phenomena using orbital imagery with very high spatial resolution (greater than 1 meter). These images are typically captured by multispectral bands in the visible and near-infrared portions of the electromagnetic spectrum, have revisit times of a few days, and record surface radiance values with up to 16-bit depth (65,536 gray levels). This vast data availability poses challenges for practical use, requiring greater computational processing capacity and enhanced class separation performance from image classification and segmentation algorithms (Amani et al., 2020; Chi et al., 2016; Parente et al., 2019; Picoli et al., 2020).

To address these challenges, Deep Learning (DL) methods have been increasingly applied to extract information from aerial and orbital remote sensing imagery, showing promising results in solving classification problems in complex scenes. For example, Wagner et al. (2020a) used a U-Net convolutional neural network to map buildings from WorldView-3 imagery, while Braga et al. (2020) employed a CNN to identify vegetation species in tropical forests using WorldView-2 imagery. Both studies reported satisfactory outcomes, suggesting that DL techniques can achieve high accuracy even with high to very high-resolution imagery.

This recent application of DL algorithms appears to be shifting the paradigm of remote sensing image processing, as evidenced by the growing volume of related publications (Li et al., 2022). Recent studies have demonstrated the potential of these methods for LULC mapping using medium, high, and very high-resolution images. In general, DL methods have shown performance improvements over traditional Machine Learning (ML) algorithms in both pixel-based and region-based classification approaches (Carranza-García et al., 2019; Ge et al., 2019; Karimian et al., 2024; Nigar et al., 2024; Vali et al., 2020; Yu et al., 2022; Zhang et al., 2023; Zhao et al., 2023). These advances enable the production of critical information for environmental diagnostics and land-use planning with higher accuracy and reduced human intervention during post-classification.

Although the literature highlights the potential of DL techniques for high-resolution image classification, no studies were found at the time of this writing that applied these methods to CBERS-4A imagery, particularly within major scientific databases such as Scopus, Web of Science, and SciELO. Additionally, it is worth emphasizing that CBERS-4A is the result of a Sino-Brazilian partnership and its data are freely distributed by Brazil's National Institute for Space Research (INPE). Given these observations, this study aims to evaluate the accuracy of LULC mapping in a section of the urban area of Rio Claro (São Paulo, Brazil), using the U-Net architecture and CBERS-4A (WPM) imagery with 2-meter spatial resolution. The following sections present the theoretical background, methodology, study area characteristics, classifier accuracy assessments (for both ML and DL approaches), comparative results, and concluding considerations on the research.

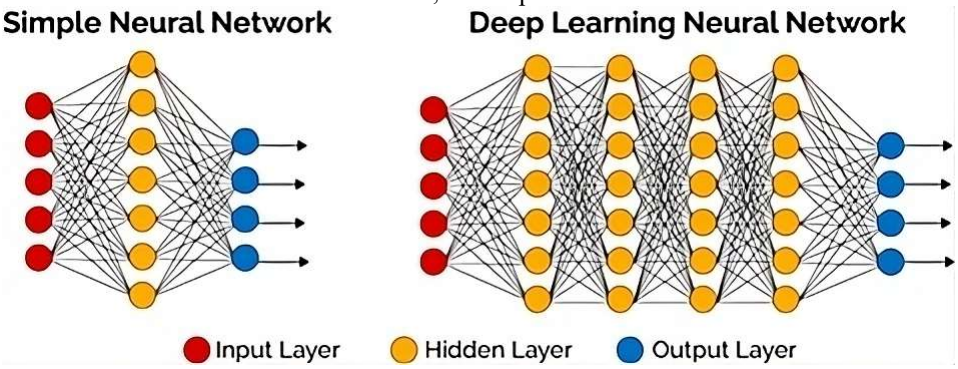
2 DEEP LEARNING

According to Chollet (2021), Deep Learning is a subfield of Machine Learning (ML), which in turn belongs to the broader domain of Artificial Intelligence (AI). The development of ML techniques, particularly in the 1980s, marked a shift in computational programming paradigms by enabling software systems to be trained to perform tasks. In contrast to rule-based systems, where input data and predefined rules produce logical outcomes, ML systems are trained using input data paired with expected outputs, allowing the algorithm to infer the underlying rules.

What differentiates DL from traditional ML is how learning occurs. In ML, one or more mathematical functions process input data to generate results, and these models may or may not consist of layered structures.

When such models are neural in nature, they are often referred to as “shallow learning” models, since they typically involve only one or two layers. Each layer applies a mathematical function to process incoming data and produce intermediate outputs. In contrast, DL models are composed of many successive layers—referred to as hidden layers—each performing increasingly complex transformations. Modern DL models can include hundreds of layers, all of which are trained using input data (Chollet, 2021; Johnson & Khoshgoftaar, 2019; Ponti & Costa, 2017; Zhao et al., 2019). Figure 1 illustrates a comparison between a shallow neural network and a deep neural network.

Figure 1 – Comparison between a shallow neural network (left) and a deep neural network (right). Circles represent the network’s neurons, organized in layers. Red circles represent the inputs; orange, the neurons in the hidden layers; and blue, the outputs.



Source: Vázquez (2017).

Artificial neural networks are inspired by the functioning of biological neurons. The computational process unfolds sequentially through layers (or neurons), each applying an activation function to determine whether a given signal is propagated. This decision is based on a weighted sum of input signals and a defined threshold. If the weighted input exceeds the threshold, the neuron "fires", outputting a value of 1; otherwise, it outputs 0. This binary decision-making mechanism mirrors human cognitive processes, wherein predefined criteria and priorities guide task execution. Sigmoid functions are commonly used to ensure positive-only outputs, enabling forward propagation of signals (Chollet, 2021; Ponti & Costa, 2017; Zhang et al., 2021). ReLU (Rectified Linear Unit) has become a widely adopted activation function in DL due to its simplicity and non-saturating behavior compared to sigmoid functions.

Among DL models, Convolutional Neural Networks (CNNs) are particularly prominent for their effectiveness in semantic image segmentation (Ponti & Costa, 2017).

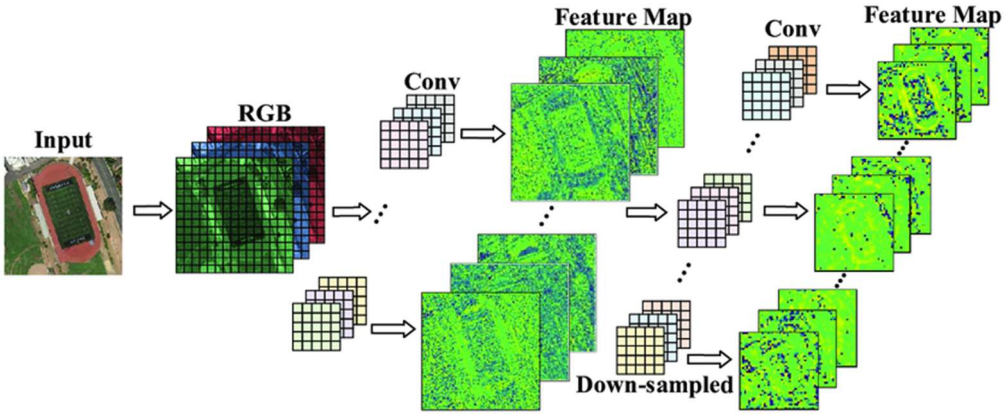
2.1 Convolutional Neural Networks

As noted earlier, CNNs are inspired by the human brain, particularly in terms of experiential learning and decision-making processes. A neural network analyzes input data through a sequence of mathematical functions and decision criteria that guide learning and, consequently, influence the classification of that data.

CNNs are structured in convolutional layers, which apply linear spatial filters—known as masks or kernels—to input images. These filters generate new outputs called feature maps. These feature maps are subsequently passed through deeper layers, enabling the progressive extraction of increasingly abstract spatial and spectral information (Goodfellow et al., 2016; Prakash & Rao, 2017). Pooling layers are often included to reduce the spatial resolution of feature maps, thereby lowering computational costs and limiting the volume of data propagated through the network (Brownlee, 2019; Chollet, 2021; Ponti & Costa, 2017).

Figure 2 schematically illustrates the application of convolutional filters, pooling operations, and the resulting feature maps within a CNN.

Figure 2 – Convolutional filters, pooling operations, and feature map generation in a CNN.



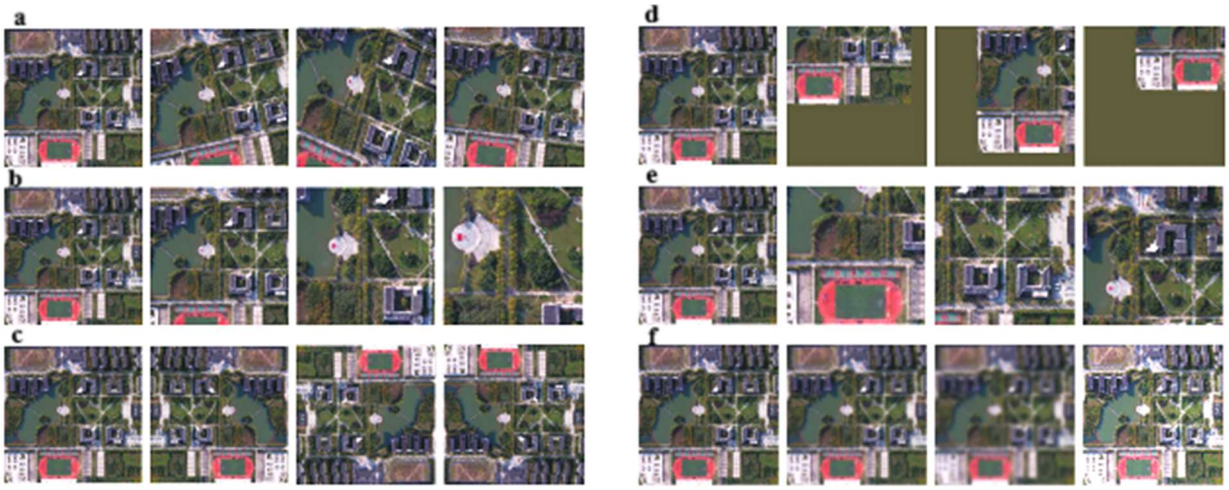
Source: Deng, Liu & Mao (2022).

Among various CNN architectures, U-Net has proven especially effective for remote sensing image classification tasks, including the segmentation and extraction of individual objects such as buildings (Wagner et al., 2020a) and tree species (Wagner et al., 2020b), as well as for broader LULC mapping efforts (Solórzano et al., 2021; Yu et al., 2022; Zhang et al., 2021).

The U-Net model is named for its U-shaped architecture, composed of a dual-path workflow. The first path functions as a conventional CNN encoder, performing convolution and pooling operations that reduce image resolution and extract features. The second path operates as a decoder, employing upsampling (or deconvolution) to restore the original resolution. A key feature of U-Net is the use of skip connections, which concatenate feature maps from the encoding path with those in the decoding path. These connections preserve spatial information from the original image and enable the model to accurately relocate each classified pixel, thus producing output with the same resolution as the input image. This enhances classification performance by integrating both spectral and spatial context, resulting in highly precise segmentation (Malik et al., 2021; Vali et al., 2020; Wagner et al., 2020a; Yan et al., 2022; Yu et al., 2022).

For remote sensing applications, the literature emphasizes the importance of data augmentation during training. Augmentation introduces variations in the training images—such as rotation, flipping, zooming, and distortion—to simulate diverse real-world conditions and increase the size of the training dataset. This technique improves the model's robustness and generalization by enabling it to correctly classify target features even when they appear in altered orientations or contexts (Ma et al., 2019; Vali et al., 2020; Yu et al., 2017). Figure 3 provides an example of augmented imagery transformations. In remote sensing, objects such as buildings or roads frequently vary in shape and size, despite belonging to the same class.

Figure 3 – Example of augmented image generation. Each horizontal sequence shows transformations applied to the original image (first in each row), including: a. rotation; b. zoom; c. flipping; d. random cropping; e. distortion; and f. noise.



Source: Adapted from Hao et al. (2023).

This complex image processing architecture has consistently produced high-performing results in remote sensing applications, as discussed in the next section.

2.2 Application of Deep Learning for Land Use and Land Cover Mapping

In recent years, the application of DL algorithms in remote sensing has grown significantly—particularly for image segmentation, feature extraction, and LULC classification (Wang et al., 2024). This expansion is driven not only by improvements in classification accuracy compared to traditional methods but also by the adaptability of DL algorithms to diverse user needs, which encourages experimentation and novel applications (Kuras et al., 2021; Ma et al., 2019; Vali et al., 2020).

One of the primary challenges in applying DL to remote sensing is the development of accurate and representative ground truth datasets, which serve as training and validation references (Vali et al., 2020). Some researchers generate this data through manual, high-precision vectorization (Wagner et al., 2020b), while others experiment with semi-automated approaches, such as autoencoders for image segmentation—though these still require manual validation and class labeling (Jozdani et al., 2019).

Ma et al. (2019) observed that most published studies on LULC mapping using DL employed CNNs for images with spatial resolutions equal to or better than 10 meters. Their meta-analysis illustrates the strong potential of CNNs for supervised classification of complex scenes, particularly in urban environments with medium to high spatial resolution imagery.

Yu et al. (2022) also discussed the challenges of urban LULC classification, emphasizing that poor-quality ground truth data can compromise DL model performance. To address this, they generated a robust reference dataset using GIS tools and proposed an adaptation of the U-Net architecture, which achieved nearly 90% class-wise accuracy when compared to ground truth.

Abdollahi et al. (2021) also modified U-Net to enhance segmentation of roads and buildings in aerial imagery. By incorporating a specialized Boundary-Aware Loss (BAL) function, they improved the model's focus on complex object boundaries such as small and irregularly shaped features—resulting in a ~2% accuracy gain over the standard U-Net.

Carranza-García et al. (2019) compared the performance of traditional ML classifiers—including Random Forest, Support Vector Machines (SVM), and K-Nearest Neighbors—against a CNN developed by the authors. The CNN consistently outperformed the ML algorithms in classifying hyperspectral and radar imagery across all validation scenarios.

These findings align with the general consensus in the literature: the U-Net model can deliver accurate LULC classifications across various data sources. Certain adaptations may yield marginal improvements, as demonstrated by Lv et al. (2023), who developed an Enhanced U-Net (E-UNet) incorporating adversarial training, multiscale convolutions, and a polarized self-attention module to detect changes in bitemporal remote sensing imagery. Their model achieved a 4% increase in overall accuracy on tested datasets.

3 MATERIALS AND METHODS

This section presents the materials and methods employed in this study, which involved the supervised classification of a CBERS-4A image over an area of approximately 9 km² located in the municipality of Rio Claro (São Paulo, Brazil), with the goal of mapping land use and land cover (LULC).

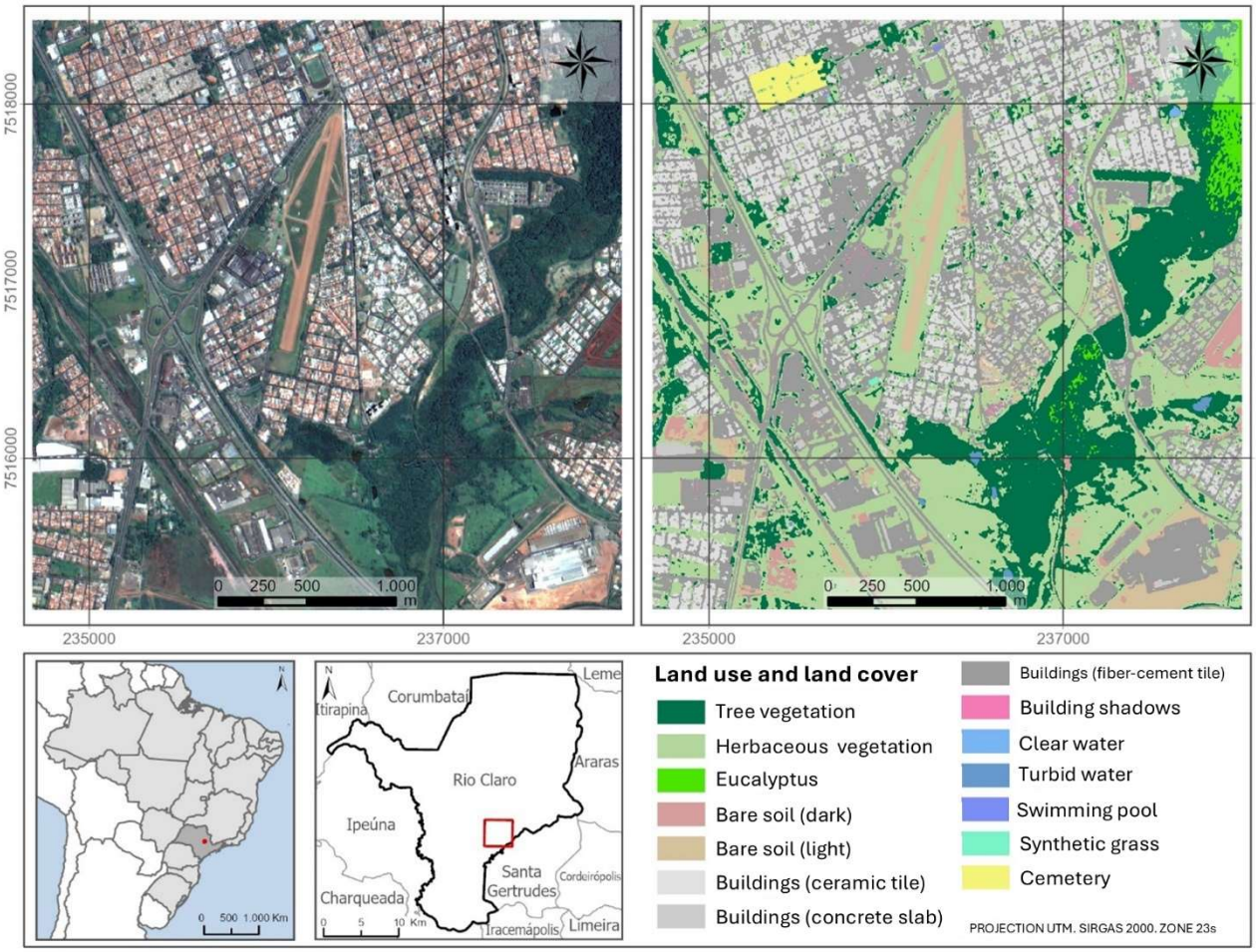
The study area is in the southern portion of the municipality and was selected due to its diverse LULC typologies, which pose classification challenges due to the spectral similarity among certain classes. Additionally, the authors reside in the region, which facilitated the construction of the ground truth through frequent in situ observations.

Initially, a CBERS-4A WPM sensor image from May 2023 was selected, as it was the most recent available image with no cloud cover. The image was processed to enhance its spatial resolution using the pansharpening technique, which merges high-frequency spatial detail from the panchromatic band with the spectral information from multispectral bands. This process was performed using the CBERS4A Downloader plugin in QGIS, which supports RGB+NIR band composition and Brovey pansharpening transformation

(Klippel, 2021).

LULC classes were defined through visual interpretation of the imagery, comparison with higher-resolution satellite imagery available freely online, and targeted field visits. The LULC typologies identified in the CBERS-4A image include: Tree vegetation, Herbaceous vegetation, Eucalyptus, Bare soil (dark), Bare soil (light), Buildings (ceramic tile), Buildings (concrete slab), Buildings (fiber-cement tile) / Asphalt (grouped in one class), Building shadows, Clear water, Turbid water, Swimming pool, Synthetic grass, and Cemetery. These typologies are shown in Figure 4, which also presents the study area location.

Figure 4 – Location and characteristics of the study area. Left: true color composite of the CBERS-4A (WPM) image with 2-meter spatial resolution. Right: ground truth derived from classification and post-processing.



Produced by: The authors (2025).

The ground truth dataset used for image classification and accuracy assessment (Figure 4) was generated via supervised classification of the same CBERS-4A image, followed by post-classification procedures including majority filtering (3×3 pixel mask) and manual pixel editing. This process drew on the findings of Magalhães (2024), who reported superior performance of the SVM algorithm over Random Trees, both available in ArcGIS Pro. Based on these findings, classification tests were conducted using SVM with three sample size levels: low (10 sample polygons per class), medium (50), and high (100). For each sampling level, the following values were tested for the "Maximum Number of Samples per Class" (MNSC) parameter: 125, 250, 500 (default), and 1,000.

The MSC parameter helps balance class representation during training by limiting the influence of majority classes. This was essential in the present study due to class imbalance, where classes such as Cemetery, Swimming pool, and Synthetic grass covered minimal areas compared to dominant classes like Tree vegetation and Herbaceous vegetation. In total, 12 supervised classifications were performed (3 sampling levels × 4 MSC values).

Initial evaluation of the classification outputs was conducted visually, comparing the results to the original CBERS-4A image, and quantitatively through accuracy assessments using randomly sampled points.

Since the ground truth was still under construction, this intermediate evaluation aimed to select the best-performing SVM result for comparison with the U-Net classification and to guide the ground truth refinement.

Accuracy was assessed using the Accuracy Assessment Points tool with stratified random sampling, allocating sample points to each class proportionally to its area. A minimum of 323 points was calculated based on Foody's (2009) formula (Equation 1), using a 95% confidence level (Z = 1.96). Due to the number of classes and the total area analyzed, the tool automatically generated 394 points distributed randomly over the classified image.

$$N = \left(\frac{Z^2.P.(1-P)}{E^2}\right) = \left(\frac{1{,}96^2.0{,}70.(1-0{,}70)}{0{,}05^2}\right) \approx 323 \quad (1) \qquad (1)$$

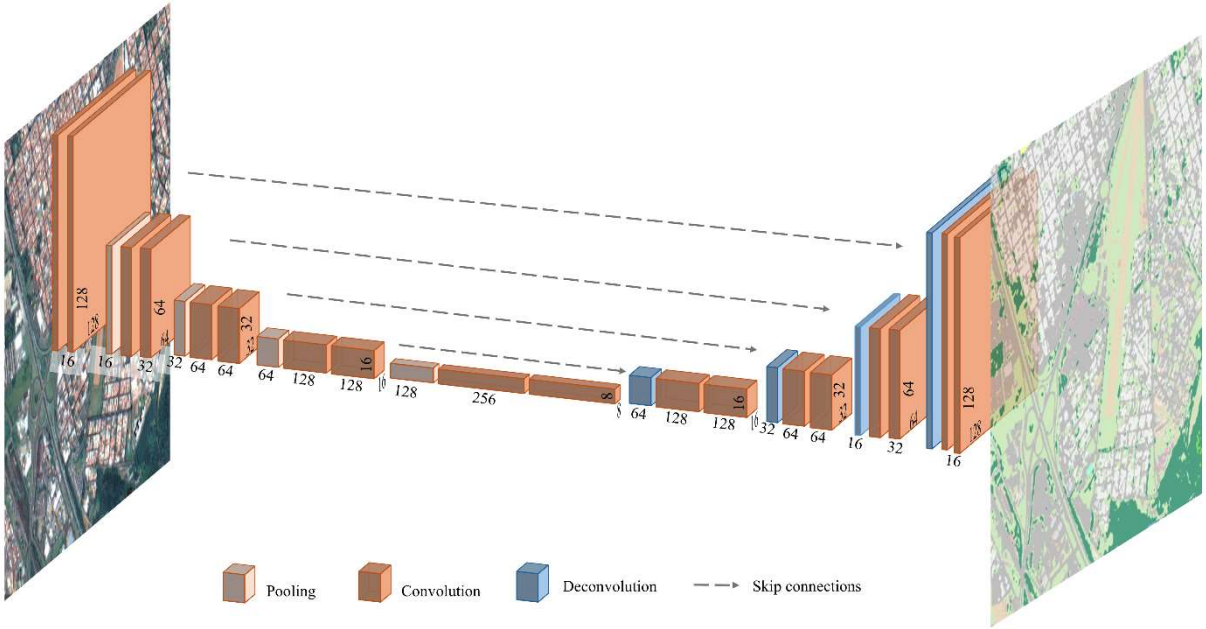
Where N is the required sample size, Z is the standard normal value for the desired confidence level, P is the expected proportion of correctly classified samples, and E is the maximum acceptable error.

The best result was obtained with medium sampling (50 samples per class) and an MNSC of 500, yielding an overall accuracy of 0.68 and a Kappa coefficient of 0.64. Based on this output, the ground truth was refined. First, the Majority Filter was applied to reduce the salt-and-pepper effect caused by isolated misclassified pixels. Then, a detailed manual editing process was conducted to match the classified map to observed typologies in the WPM image as closely as possible.

The CBERS-4A image served as the primary reference for this refinement, with support from higher-resolution imagery, web-based map services, street view applications, and in situ observations by the authors. Once the ground truth was finalized, classification was performed using the U-Net architecture.

The U-Net model was implemented in Python using the Keras and TensorFlow libraries. The original 3-channel (RGB) image was subdivided into 128 × 128 pixel tiles to enable manageable processing. During the encoding phase, 3 × 3 convolutional filters and pooling operations were applied, progressively reducing the image size to 64^2, 32^2, 16^2, and 8^2 pixels. The decoding phase restored the resolution via upsampling (deconvolution) and concatenated the results with feature maps from earlier layers—ensuring correct pixel alignment and spatial coherence. The architecture is shown in Figure 5.

Figure 5 – U-Net architecture used for classifying the CBERS-4A image.

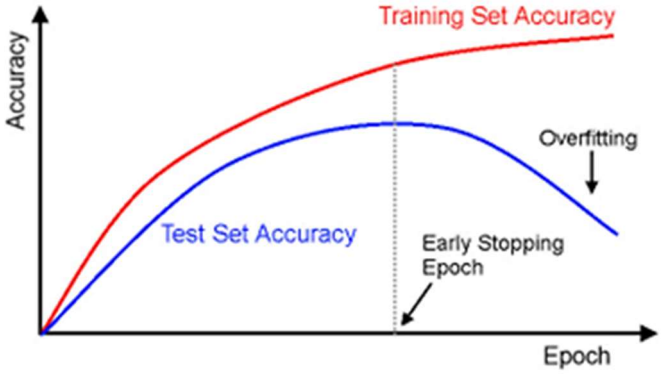


Produced by: The authors (2025).

Model training occurred over multiple epochs, defined as full passes through the training dataset. Few epochs may result in underfitting, while excessive epochs can cause overfitting—where the model starts to memorize noise rather than learning generalizable patterns (DSA, 2022). The ideal number of epochs was determined by observing accuracy versus epoch count and applying early stopping—terminating training at

peak accuracy before overfitting begins (Figure 6). For this study, the optimal value was determined to be 200 epochs.

Figure 6 – Determining the Early Stopping Epoch.



Source: DSA (2022).

The image processing performed in the Python environment does not require georeferencing information, as used in GIS software like ArcGIS Pro, since images are processed individually. Therefore, in order to reconstruct the georeferencing of the classified image and ensure its correct alignment over the original CBERS-4A image, the auxiliary georeferencing files automatically generated by ArcGIS Pro—namely, the .pgw, .xml, and .ovr extensions—were preserved. The association of the classified images with these auxiliary files was achieved by maintaining consistent file naming conventions, which allowed the classification results to be correctly georeferenced.

The classification output, generated in 128 × 128 pixel tiles, was spatialized in ArcGIS Pro to create a mosaic of the classification with georeferencing corresponding to the original image. This procedure enabled a direct comparison between the classification produced by the U-Net model and that generated by the SVM algorithm (excluding post-classification edits), with the aim of determining whether the U-Net model yields more accurate results than the traditional ML approach. The accuracy of the classifications was evaluated using the metrics Overall Accuracy (OA), Precision, Recall, F1 Score, and Intersection over Union (IoU), which were derived from the comparison between the ground truth and the classified image using a confusion matrix. These metrics are widely adopted in the literature for assessing the accuracy of remote sensing image classification and are detailed in studies such as those by Maxwell, Warner, and Guillén (2021a, 2021b), among others.

In simplified terms, from the confusion matrix, the classified image pixels are identified as True Positives (TP), which are those correctly predicted; True Negatives (TN), which are those incorrectly predicted; False Positives (FP), which were incorrectly predicted as belonging to a class; and False Negatives (FN), which were incorrectly omitted. Based on these categories, mathematical formulations are applied to measure the classifier's reliability in relation to a reference dataset.

Overall Accuracy, for instance, calculates the number of correct predictions made by the model in relation to the total sample size, through the ratio of all pixels classified as true (TP + TN) to the total number of pixels. This metric provides a general indication of the model's correctness. However, it presents limitations in scenarios involving class imbalance, as it may mask performance issues when one or more classes dominate the dataset (Morales-Barquero et al., 2019). To address this, other metrics may be used, such as Precision, which assesses the proportion of correct predictions in relation to the total number of positive predictions, calculated as the ratio between TP and the sum of TP and FP. This measure indicates how much confidence we can place in a model when it predicts that a pixel belongs to a given class. Recall, on the other hand, is the ratio between the total number of correctly classified pixels (TP) for a given class and the total number of pixels in that class within the ground truth (TP + FN). It provides an indication of the model's detection rate—its ability to correctly identify a class with respect to the real distribution. F1 Score combines the Precision and Recall metrics into a single measure using their harmonic mean, and is particularly useful in cases of class imbalance, as it balances omission and commission errors. Lastly, the IoU metric evaluates the overlap between the prediction and the ground truth by calculating the ratio between the area of intersection (TP) and the union

of the predicted and reference areas (TP + FN + FP) (Maxwell, Warner & Guillén, 2021a, 2021b; Morales-Barquero et al., 2019).

4 RESULTS AND DISCUSSIONS

Figure 7 presents the confusion matrices obtained for the classifications performed with the SVM (7a) and U-Net (7b) models, respectively, in comparison with the ground truth. The Y-axis indicates the reference classes from the ground truth, while the X-axis shows the classes predicted by each classifier. The values inside the matrices correspond to the number of pixels assigned to each class. Higher values appear in darker tones, indicating the most dominant classes in the study area. When these values are located along the diagonal—where the true class and predicted class intersect—they represent correct classifications. Values in other matrix positions indicate misclassifications and help identify which classes generated the most confusion.

The comparison between the matrices reveals a slightly superior performance by U-Net relative to SVM for the classification of the tested image. Notably, the Deep Learning model demonstrated the ability to detect highly specific classes, such as Cemetery, which the SVM was unable to identify. This result can be attributed to the U-Net's capacity to incorporate contextual and textural information, which enhances its ability to distinguish between classes that are spectrally similar—such as the Cemetery and Buildings with concrete slab roofs. In the same vein, the class Bare soil (light) was more accurately predicted by U-Net, whereas in the SVM classification, it was frequently confused with Buildings with ceramic tile roofs. Again, these are elements with similar spectral characteristics but different shapes, textures, and spatial patterns.

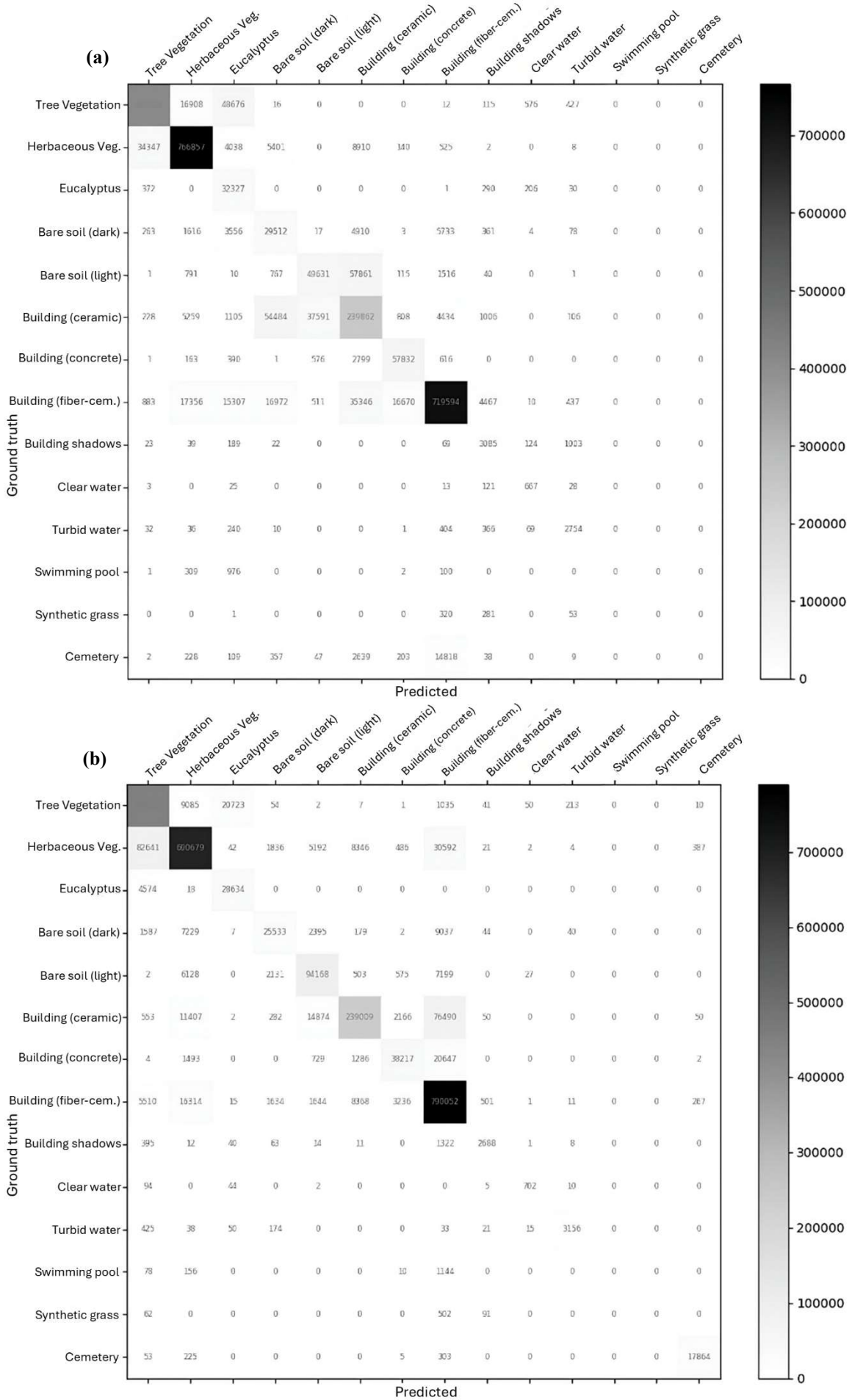
It is worth noting that the classes Synthetic grass and Swimming pool were not identified by either classifier, most likely because they occupy very small areas within the study region.

For classes with more distinct spectral signatures—such as Tree vegetation, Herbaceous vegetation, and Buildings with fiber-cement roofs (grouped with Asphalt)—both classifiers performed similarly. In some instances, SVM even demonstrated slightly better accuracy than U-Net. This is further evidenced by the class-by-class metrics for Precision, Recall, F1 Score, and IoU, presented in Figure 8.

Among the twelve classes identified by at least one of the classifiers, SVM outperformed or matched U-Net in four. This reinforces the reliability of SVM as a classification model for LULC mapping, as highlighted by Sheykhmousa et al. (2020). On the other hand, it also exposes its limitations in distinguishing classes with highly similar spectral characteristics, which hinders its applicability for land use mapping in urban areas using high-resolution imagery. Analysis of the chart (Figure 8) further shows that SVM yielded irregular, unbalanced results across the class set, whereas U-Net delivered more consistent and homogeneous performance. This reinforces U-Net's potential for classifying complex scenes composed of high-resolution images with high class heterogeneity—typical of urbanized areas, as also noted by Ma et al. (2019). The instability observed in the SVM classification impacts the classifier's overall performance metrics, as shown in Table 1. Among the evaluated metrics, Precision stands out, with scores of 0.48 for SVM and 0.72 for U-Net, indicating a notable difference in the classifiers' ability to make accurate predictions.

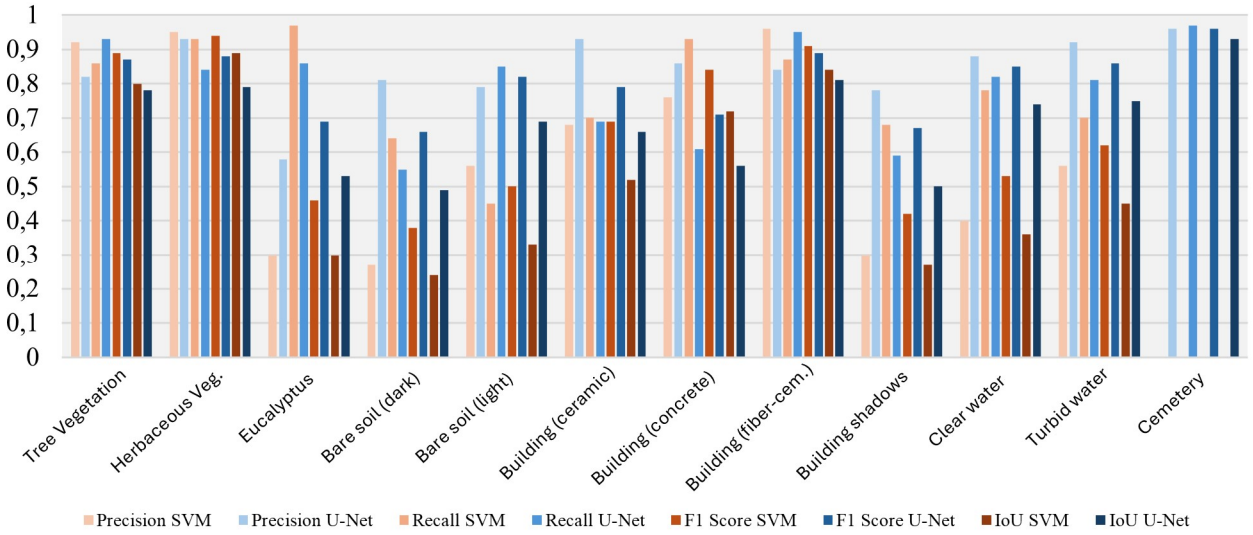
 DOI: http://dx.doi.org/10.14393/rbcv77n0a-75495

Figure 7 – Confusion matrices for the supervised classifications of the CBERS-4A image: (a) presents the results from the SVM, and (B), those from U-Net.



Produced by: The authors (2025).

Figure 8 – Results for the Precision, Recall, F1 Score, and IoU metrics for each mapped land use and land cover class. Red tones indicate the performance of SVM, while blue tones represent that of U-Net.



Produced by: The authors (2025).

Table 1 – Summary of the classification accuracy for U-Net and SVM.

Metric	OA	IoU	Precision	Recall	F1 Score
U-Net	0,86	0,59	0,72	0,68	0,69
SVM	0,84	0,41	0,48	0,61	0,51

Produced by: The authors (2025).

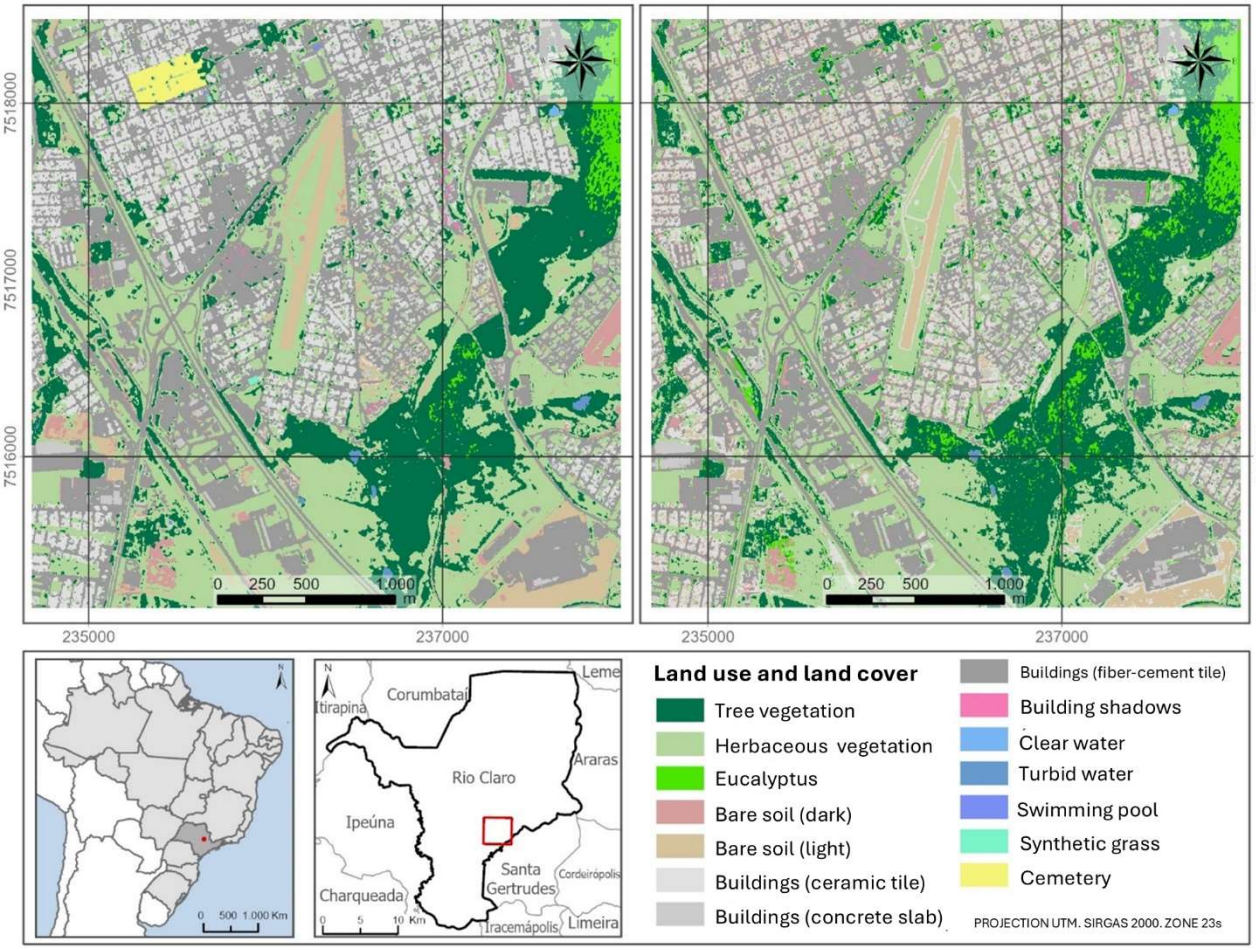
A visual assessment of the classification produced by the SVM (Figure 9) shows that typologies occupying large and continuous areas, such as Tree and Herbaceous vegetation, were well represented. In contrast, smaller elements with spectral characteristics similar to other materials—such as ceramic roof tiles and bare soils—were misclassified, compromising their usefulness for urban planning and land management purposes.

This limitation is likely due to the significant amount of spatial information available in the image, stemming from its 2-meter spatial resolution, combined with the limited spectral information provided by the sensor, which includes only visible and near-infrared bands. The inclusion of additional spectral bands—such as the mid-infrared bands available on the Landsat satellites—could potentially improve classification performance for this type of imagery.

By visually analyzing the classification results produced by the U-Net model (Figure 10), the classifier's superior accuracy for the proposed task becomes evident. The image corroborates the performance metrics, as there is strong visual agreement between the predicted LULC classes and the ground truth. Particularly noteworthy is the accurate separation of classes that exhibited the greatest confusion in the SVM classification, such as Buildings with ceramic tile roofs and the Cemetery. Despite occupying small areas and presenting spectral characteristics similar to those of other materials, these typologies were mapped with remarkable precision by the U-Net model.

 DOI: http://dx.doi.org/10.14393/rbcv77n0a-75495

Figure 9 – Land use and land cover classification using the SVM algorithm. Left: ground truth; right: classification result by SVM.



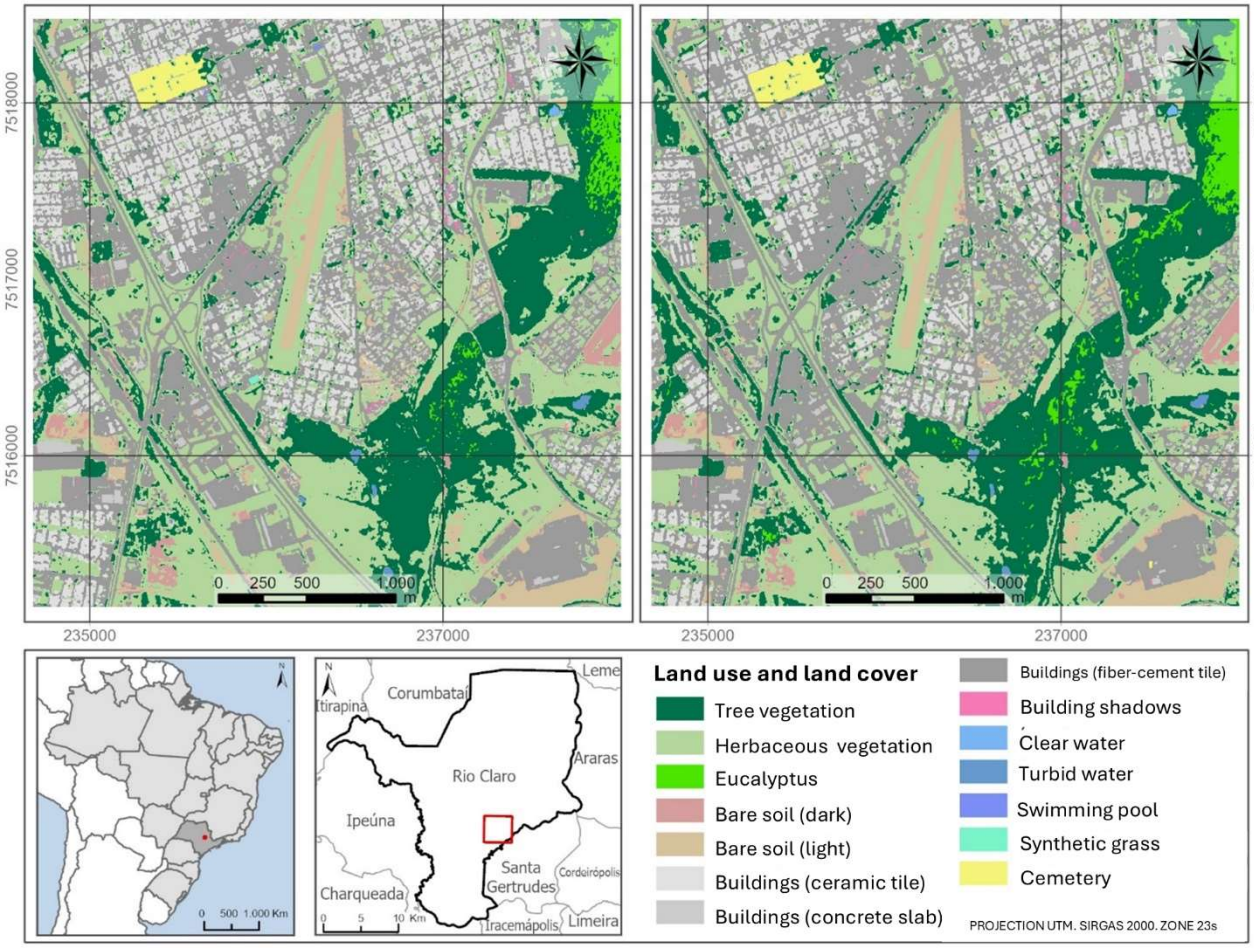
Produced by: The authors (2025).

The results presented in this study are preliminary and based on tests conducted in a limited urban section of Rio Claro (SP). Nevertheless, it is expected that the model's accuracy could be further improved using larger training datasets. Still, the findings are promising and demonstrate the potential of this technique for urban LULC mapping using high-resolution, freely available images produced by the CBERS-4A WPM sensor. These findings are consistent with those reported in the specialized literature. For instance, Carranza-García et al. (2019), in comparing ML and DL algorithms for supervised image classification, concluded that CNNs outperformed other techniques across all tested datasets. Similarly, Jozdani et al. (2019) compared the performance of various DL and ML models for urban LULC mapping using very high-resolution imagery (30–50 cm) and found that the Multilayer Perceptron architecture achieved the highest classification accuracy. They also emphasized that SVM yielded highly accurate results, reinforcing the versatility of ML algorithms.

Similarly, Solórzano et al. (2021) reported that the use of CNNs in Earth observation applications has led to improvements in detailed LULC classification, noting that even when trained with small datasets, U-Net outperformed the Random Trees algorithm.

Finally, Zhang et al. (2021) argue that LULC mapping is a complex task in the context of remote sensing, especially as spatial resolution increases. Faced with the challenge of pixel-based methods (which are prone to noise) and object-based approaches (which require substantial manual intervention), the authors proposed an end-to-end CNN model. Their results showed that the CNN outperformed traditional pixel-based methods in classifying high-resolution imagery.

Figure 10 – Land use and land cover classification by the U-Net model. Left: ground truth; right: classification result by U-Net.



Produced by: The authors (2025).

5 CONCLUSIONS

This article presented the initial findings of a research project investigating the application of Deep Learning (DL) algorithms for the supervised classification of remote sensing images, with a specific focus on imagery from the CBERS-4A satellite's WPM sensor, which offers a spatial resolution of 2 meters.

The results point to the potential of this technique for mapping a significant range of land use and land cover typologies using high-resolution images, as it achieved higher accuracy than the SVM algorithm available in ArcGIS Pro for the tested dataset. These findings are promising, as the model was capable of distinguishing between classes with highly similar spectral characteristics—such as Buildings with concrete slab roofs and Cemetery, or Bare soil and Buildings with ceramic tile roofs. Although preliminary, the results suggest that further testing with larger training datasets could enhance the model's performance. At this stage, however, it is already important to highlight the potential of the employed technique for mapping land use and land cover in Brazilian municipalities using freely available high-resolution imagery, thereby contributing valuable data to support urban and regional planning and management efforts. For instance, such classification could be used to monitor vegetation cover within urban areas, assess soil permeability, study urban heat island formation, or inform land subdivision regulations, among other applications.

The accuracy analysis of the results aligns with international literature in the field, which demonstrates the high performance of DL classifiers in processing remote sensing images. The study also highlights the capability of DL models to extract information from high-resolution imagery in complex urban environments—an inherently challenging task in remote sensing image analysis.

The research is ongoing, and the next steps will involve replicating the classification procedure across the entire municipal area of Rio Claro (SP). This expansion will require significant human effort to generate ground truth data, as well as substantial computational resources to process such a large area. These challenges

are common in DL applications, as evidenced in related works such as those by Vali et al. (2020) and Yu et al. (2022). Nevertheless, the completion of this task may enable the creation of a training dataset for DL classifiers specific to CBERS-4A imagery, which could in turn be applied to images from other dates and locations with similar characteristics. This is possible because the algorithm does not rely on the spatial location of training samples to operate, allowing the method to be replicated elsewhere and contributing to the generation of LULC data based on free, high-resolution imagery.

References

Abdollahi, A., Pradhan, B., Shukla, N., Chakraborty, S., & Alamri, A. (2021). Multi-object segmentation in complex urban scenes from high-resolution remote sensing data. *Remote Sensing*, 13(18), 3710. https://doi.org/10.3390/rs13183710

Amani, M., Ghorbanian, A., Ahmadi, S. A., Kakooei, M., Moghimi, A., Mirmazloumi, S. M., Alizadeh Moghaddam, S. H., Mahdavi, S., Ghahremanloo, M., Parsian, S., Wu, Q., & Brisco, B. (2020). Google Earth Engine cloud computing platform for remote sensing big data applications: A comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5326–5350. https://doi.org/10.1109/JSTARS.2020.3011956

Braga, J. R. G., Peripato, V., Dalagnol, R., Ferreira, M. P., Tarabalka, Y., Aragão, L. E. O. C., Campos Velho, H. F. de, Shiguemori, E. H., & Wagner, F. H. (2020). Tree crown delineation algorithm based on a convolutional neural network. *Remote Sensing*, 12(8), 1288. https://doi.org/10.3390/rs12081288

Brownlee, J. (2019). *Deep learning for computer vision: Image classification, object detection and face recognition in Python*. Machine Learning Mastery.

Carranza-García, M., García-Gutiérrez, J., & Riquelme, J. (2019). A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sensing*, 11(3), 274. https://doi.org/10.3390/rs11030274

Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., & Zhu, Y. (2016). Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, 104(11), 2207. https://doi.org/10.1109/JPROC.2016.2598228

Chollet, F. (2021). *Deep learning with Python* (2nd ed.). Manning Publications.

Data Science Academy (DSA). (2022). *Deep learning book*. https://www.deeplearningbook.com.br/

Deng, T., Liu, X. & Mao, G. (2022). Improved YOLOv5 Based on Hybrid Domain Attention for Small Object Detection in Optical Remote Sensing Images. *Eletronics*, 11(17), 2657. https://doi.org/10.3390/electronics11172657

Foody, G. M. (2009). Sample size determination for image classification accuracy assessment and comparison. *International Journal of Remote Sensing*, 30(20), 5273-5291. https://doi.org/10.1080/01431160802582899

Ge, P., He, J., Zhang, S., Zhang, L., & She, J. F. (2019). An integrated framework combining multiple human activity features for land use classification. *ISPRS International Journal of Geo-Information*, 8(2), 90. https://doi.org/10.3390/ijgi8020090

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Hao, X., Liu, L., Yang, R., Yin, L., Zhang, L. & Li, X. (2023). A Review of Data Augmentation Methods of Remote Sensing Image Target Recognition. *Remote Sensing*, 15(3), 827. https://doi.org/10.3390/rs15030827

Johnson, J. M., & Khoshgoftaaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1), 27. https://doi.org/10.1186/s40537-019-0192-0

Jozdani, S. E., Johnson, B. A., & Chen, D. (2019). Comparing deep neural networks, ensemble classifiers, and support vector machine algorithms for object-based urban land use/land cover classification. *Remote Sensing*, 11(14), 1713. https://doi.org/10.3390/rs11141713

Karimian, R., Rangzan, K., Karimi, D. & Einali, G. (2024). Spatiotemporal Monitoring of Land Use-Land Cover and Its Relationship with Land Surface Temperature Changes Based on Remote Sensing, GIS, and Deep Learning. *J Indian Soc Remote Sens*, 52, 2461. https://doi.org/10.1007/s12524-024-01958-3

Klippel, S. (2022). *CBERS4A downloader QGIS plugin*. https://github.com/sandroklippel/cbers4a

Kuras, A., et al. (2021). Hyperspectral and Lidar data applied to the urban land cover machine learning and neural-network-based classification: A review. *Remote Sensing*, 13(17), 3393. https://doi.org/10.3390/rs13173393

Li, Z., Wang, Y., Zhang, N., Zhang, Y., Zhao, Z., Xu, D., Ben, G., & Gao, Y. (2022). Deep Learning-Based Object Detection Techniques for Remote Sensing Images: A Survey. *Remote sensing*, 14(10), 2385. https://doi.org/10.3390/rs14102385

Lv, Z., Huang, H., Sun, W., Lei, T., Benediktsson, J. A. & Li, J. (2023). Novel Enhanced UNet for Change Detection Using Multimodal Remote Sensing Image. *IEEE Geoscience and Remote Sensing Letters*, 20, 1. 10.1109/LGRS.2023.3325439

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.05.019

Magalhães, D. M. (2024). Avaliação da acurácia da classificação supervisionada de imagens de sensoriamento remoto utilizando o software ArcGIS Pro. In N. I. Ladwig, T. Sutil, C. H. R. da Silva, & B. Giaccom (Eds.), *Planejamento e gestão territorial* (1st ed., pp. 141-164). Pedro e João. http://dx.doi.org/10.51795/9786526514276

Malik, K., Robertson, C., Braun, D., & Greig, C. (2021). U-Net convolutional neural network models for detecting and quantifying placer mining disturbances at watershed scales. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102510. https://doi.org/10.1016/j.jag.2021.102510

Maxwell, A. E., Warner, T. A. & Guillén, L. A. (2021a). Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review. *Remote Sensing*, 13(13), 2450. https://doi.org/10.3390/rs13132450

Maxwell, A. E., Warner, T. A. & Guillén, L. A. (2021b). Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 2: Recommendations and Best Practices. *Remote Sensing*, 13(13), 2591. https://doi.org/10.3390/rs13132591

Morales-Barquero, L., Lyons, M. B., Phinn, S. R., & Roelfsema, C. M. (2019). Trends in Remote Sensing Accuracy Assessment Approaches in the Context of Natural Resources. *Remote Sensing*, 11(19), 2305. https://doi.org/10.3390/rs11192305

Nigar, A., Li, Y., Jat Baloch, M. Y., Alrefaei, A. F. & Almutairi, M. H. (2024). Comparison of machine and deep learning algorithms using Google Earth Engine and Python for land classifications. *Frontiers in Environmental Science*, 12, 01. 10.3389/fenvs.2024.1378443

Odenyo, V. A. O., & Pettry, D. E. (1977). Land-use mapping by machine processing of LANDSAT-1 data. *Photogrammetric Engineering and Remote Sensing*, 43(4), 515–523.

Pabi, O. (2007). Understanding land-use/cover change process for land and environmental resources use management policy in Ghana. *GeoJournal*, 68(4), 369–383. https://doi.org/10.1007/s10708-007-9120-7

Parente, L., Taquary, E., Silva, A. P., Souza, C., & Ferreira, L. (2019). Next generation mapping: Combining deep learning, cloud computing, and big remote sensing data. *Remote Sensing*, 11(23), 2881. https://doi.org/10.3390/rs11232881

Picoli, M. C. A., Simoes, R., Chaves, M., Santos, L. A., Sanchez, A., Soares, A., Sanches, I. D., Ferreira, K. R., & Queiroz, G. R. (2020). CBERS data cube: A powerful technology for mapping and monitoring Brazilian biomes. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-3–2020, 533–539. https://doi.org/10.5194/isprs-annals-V-3-2020-533-2020

Ponti, M. A., & Costa, G. B. P. (2017). Como funciona o deep learning. In: *Tópicos em gerenciamento de dados e informações* (1st ed., p. 31). SBC.

Prakash, D. P., & Rao, A. S. K. R. (2017). *Deep learning cookbook: Solve complex neural net problems with TensorFlow, H2O, and MXNET*. Packt.

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P. & Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 6308. 10.1109/JSTARS.2020.3026724.

Solórzano, J. V., Mas, J. F., Gao, Y., & Gallardo-Cruz, J. A. (2021). Land use land cover classification with U-Net: Advantages of combining Sentinel-1 and Sentinel-2 imagery. *Remote Sensing*, 13(18), 3600. https://doi.org/10.3390/rs13183600

Vali, A., Comai, S., & Matteucci, M. (2020). Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. *Remote Sensing*, 12(15), 2495. https://doi.org/10.3390/rs12152495

Vázquez, F. (2017, December 21). *Deep learning made easy with deep cognition*. Becoming Human: Artificial Intelligence Magazine. https://becominghuman.ai/deep-learning-made-easy-with-deep-cognition-403fbe445351

Wagner, F. H., Dalagnol, R., Tarabalka, Y., Segantine, T. Y. F., Thomé, R., & Hirye, M. C. M. (2020a). U-Net-Id, an instance segmentation model for building extraction from satellite images — Case study in the Joanópolis City, Brazil. *Remote Sensing*, 12(10), 1544. https://doi.org/10.3390/rs12101544

Wagner, F. H., Dalagnol, R., Tagle Casapia, X., Streher, A. S., Phillips, O. L., Gloor, E., & Aragão, L. E. O. C. (2020b). Regional mapping and spatial distribution analysis of canopy palms in an Amazon forest using deep learning and VHR images. *Remote Sensing*, 12(14), 2225. https://doi.org/10.3390/rs12142225

Wang, L., Zhang, M., Gao, X., & Shi, W. (2024). Advances and challenges in deep learning-based change detection for remote sensing images: A review through various learning paradigms. *Remote Sensing*, 16, 804. https://doi.org/10.3390/rs16050804

Yan, C., Fan, X., Fan, J., & Wang, N. (2022). Improved U-Net remote sensing classification algorithm based on multi-feature fusion perception. *Remote Sensing*, 14(5), 1118. https://doi.org/10.3390/rs14051118

Yang, Y., Wan, W., Huang, S., Lin, P., & Que, Y. (2017). A novel pan-sharpening framework based on matting model and multiscale transform. *Remote Sensing*, 9(4), 391. https://doi.org/10.3390/rs9040391

Yu, J., Zeng, P., Yu, Y., Yu, H., Huang, L., & Zhou, D. (2022). A combined convolutional neural network for urban land-use classification with GIS data. *Remote Sensing*, 14(5), 1128. https://doi.org/10.3390/rs14051128

Zhang, X., Du, L., Tan, S., Wu, F., Zhu, L., Zeng, Y., & Wu, B. (2021). Land use and land cover mapping using RapidEye imagery based on a novel band attention deep learning method in the Three Gorges Reservoir area. *Remote Sensing*, 13(6), 1225. https://doi.org/10.3390/rs13061225

Zhang, P., Wu, Y., Li, C., Li, R., Yao, H., Zhang, Y., Zhang, G. & Li, D. (2023). National-Standards- and Deep-Learning-Oriented Raster and Vector Benchmark Dataset (RVBD) for Land-Use/Land-Cover Mapping in the Yangtze River Basin. *Remote Sensing*, 14(15), 3907. https://doi.org/10.3390/rs15153907

Zhao, Z.-Q., Zheng, P., Xu, S.-T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*. https://doi.org/10.1109/TNNLS.2019.2897602

Zhao, S., Tu, K., Ye, S., Tang, H., Hu, Y. & Xie, C. (2023). Land Use and Land Cover Classification Meets Deep Learning: A Review. *Remote Sensing*, 23(21), 8966. https://doi.org/10.3390/s23218966

Author Contributions

D. M. Magalhães was responsible for Conceptualization, Project Administration, Supervision, Writing – Original Draft, and Writing – Review & Editing. J. P. Souza contributed to Data Curation and Formal Analysis (GIS). E. A. G. França contributed to Data Curation and Formal Analysis (Python).

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Biography of the Main Author

Danilo Marques de Magalhães (Belo Horizonte, MG, 1983) holds a Bachelor's (2010), Master's (2013), and Ph.D. (2021) in Geography from UFMG, including a research internship at the University of Bologna (Italy). He is an Assistant Professor at the Department of Geography and Environmental Planning at UNESP, Rio Claro, where he also teaches in the Graduate Program in Geography. He coordinates the MAPEAR Research and Extension Group, advising work in Remote Sensing and Geographic Information Systems. He currently leads the research project "Deep Learning for Land Use and Land Cover Mapping Using CBERS-4A Imagery."