



Improving urban tree species classification by deep-learning based fusion of digital aerial images and LiDAR



Matheus Pinheiro Ferreira ^{a,*}, Daniel Rodrigues dos Santos ^a, Felipe Ferrari ^b,
Luiz Carlos Teixeira Coelho Filho ^c, Gabriela Barbosa Martins ^a, Raul Queiroz Feitosa ^b

^a Defense Engineering Department, Military Institute of Engineering (IME), Praça Gen. Tibúrcio 80, 22290-270 Rio de Janeiro, RJ, Brazil

^b Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro (PUC), R. Marquês de São Vicente 225, 22541-041 Rio de Janeiro, RJ, Brazil

^c School of Engineering, Rio de Janeiro State University, R. São Francisco Xavier 524, 20550-011 Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Handling Editor: Raffaele Laforteza

Keywords:

Deep learning
Convolutional neural networks
Optical-LiDAR fusion
Surface normals
Tropical urban forests

ABSTRACT

Accurate information on tree species distribution in urban areas can offer insights into how street trees provide ecosystem services, such as air pollution mitigation and surface cooling. This article presents a method to improve tree species classification in a tropical urban area using LiDAR-derived structural properties of individual tree crowns (ITCs) and digital aerial images. We extracted four LiDAR features, including surface normals of tree leaves, intensity, tree height, and leaf area index (LAI). We conducted two experiments: In the first, we trained encoder-decoder convolutional neural networks using a stack of optical bands and one LiDAR feature at a time. In the second, we developed an optical-LiDAR fusion strategy that combined feature maps from two encoder-decoder networks. One network was trained with optical bands only, while the other was trained with the LiDAR features that improved classification accuracy in the first experiment. Our experiment results demonstrated the usefulness of surface normals and intensity in discriminating among tree species. We found that the optical-LiDAR fusion strategy increased the average F1-score by 12.6 percentage points compared to only optical bands. We also employed the new segment anything (SAM) model to automatically delineate ITCs. SAM outlined ITCs with a boundary F1-score of 98%. The SAM-delineated ITCs were used to improve raw model predictions and produce reliable species maps. This study contributes to mapping and monitoring urban tree species in tropical areas.

1. Introduction

Urban forests and street trees are essential to maintain and improve human well-being in city centers. They provide various valuable ecosystem services, including air pollution mitigation, carbon sequestration, and surface temperature regulation. Species-specific traits significantly influence the ecosystem services provided by urban trees (Cariñanos et al., 2017). For example, the efficacy of street trees in urban cooling is influenced by their morphological and structural canopy traits, including leaf area index (LAI) (Rahman et al., 2020; Smithers et al., 2018). Similarly, the capacity of urban trees to mitigate air pollution is intrinsically linked to species-specific crown features, such as crown geometry and foliage distribution. These characteristics impact the canopy's airflow dynamics, potentially increasing the deposition rates of gaseous pollutants and particulate matter (Grote et al., 2016). Therefore, knowledge of the spatial distribution of tree species in urban

settings is crucial for understanding and maximizing ecosystem service provision across the city landscape.

Ground-based surveys are the most accurate approach to collecting species identity, diversity, and abundance data. However, they are costly, require expert knowledge, and usually only comprise part of the city extent. Integrating ground-based data and remote sensing has been hailed as a promising way to map urban tree species (Shahtahmassebi et al., 2021). Several remote sensing data types acquired by active and passive sensors have been employed (Li et al., 2019). Particularly the use of very high-resolution (ground sampling distance, GSD \leq 30 cm) allows identifying individual tree crowns (ITCs) and provides excellent classification accuracy ($>$ 85%) (e.g., Martins et al., 2021; Zhang and Hu, 2012). Recently, airborne light detection and ranging (LiDAR) data have been employed to improve tree species discrimination (Wang et al., 2018).

LiDAR sensors are characterized by their ability to emit tens of

* Correspondence to: Cartographic Engineering Department, Military Institute of Engineering (IME), Praça Gen. Tibúrcio 80, 22290-270 Rio de Janeiro, RJ, Brazil.
E-mail address: matheus@ime.eb.br (M.P. Ferreira).

thousands of laser pulses, not only in the near-infrared region (with wavelengths approximately 750–1500 nm) but also in the green region (500–550 nm). These sensors record the time delay between the emission of a pulse and its return. LiDAR data is used to construct a three-dimensional (3D) point cloud of the area imaged and accurately measure the height of surface objects. From LiDAR point clouds, one can retrieve species-specific characteristics of ITCs, such as height, crown volume, and shape. LiDAR data alone usually does not provide good results for tree species discrimination. For example, [Alonzo et al. \(2014\)](#) achieved an overall accuracy of 32.9% to classify 29 urban tree species using LiDAR data solely. An overall accuracy value that is significantly lower than the 79.2% achieved using only optical hyperspectral data. Optical-LiDAR fusion is a standard procedure for incorporating LiDAR features to classify tree species. The term ‘fusion’ refers to combining multi-source remote sensing data. Recently, several studies have employed optical-LiDAR fusion to discriminate among species in urban areas (refer to [Wang et al. \(2018\)](#) for a comprehensive review on the topic). Most studies used LiDAR data to incorporate height attributes in the classification process. A few studies explored other LiDAR-derived features, including laser return intensity or surface normals. Intensity showed helpful in reducing the overlap effect between neighboring trees and improving ITC segmentation ([Ke et al., 2010](#)). To our knowledge, no previous study employed surface normals to classify tree species. Surface normals from 3D point clouds describe geometric surface properties and can be used to estimate species-specific leaf angle and orientation. A crucial aspect in this context is the need for high-density LiDAR point clouds, which are essential for accurately computing surface normals. While airborne LiDAR data may not consistently achieve the necessary density for precise leaf normal estimation, this limitation highlights a significant research gap. Despite its challenges, the potential of using airborne LiDAR data to classify tree species based on surface normals represents an unexplored area.

Another critical issue of LiDAR fusion is related to the fusion approach itself. Most studies rely on an object-based approach in which LiDAR is used to segment ITCs and retrieve a limited number of structural features such as height or intensity. These features are combined with optical-derived spectral features to form a single dataset used by a given classifier to discriminate species. Object-based approaches have limitations because they depend on object segmentation and feature engineering. The former requires using a segmentation algorithm that is usually computationally prohibitive and requires empirical parameter tuning. The latter needs expert knowledge to decide among several image feature extraction approaches, such as texture analysis or color transformations. Convolution neural networks (CNNs) overcome the limitations of object-based approaches by performing feature extraction and species classification in an end-to-end fashion ([Kattenborn et al., 2021](#)).

In this study, we fused LiDAR-derived features with high-resolution digital aerial images using CNNs to classify tree species in a highly diverse tropical urban area. Our research tested two hypotheses: (i) that surface normals from 3D LiDAR point clouds are sensitive to species-specific differences in crown structural properties, and (ii) that fusing feature maps from encoder-decoder networks trained with optical and LiDAR metrics separately improves the classification accuracy of tree species.

The results of our study contribute to the accurate mapping of street tree species in three ways: (i) by identifying important LiDAR features that can be used to discriminate species, (ii) by introducing an optical-LiDAR fusion strategy, and (iii) by demonstrating the potential of the new segment anything (SAM, [Kirillov et al. \(2023\)](#)) algorithm to produce species maps at the ITC level.

2. Materials and methods

2.1. Study area

The study area comprises the Grajaú neighborhood in Rio de Janeiro, Brazil. It is a highly diverse urban setting with more than 100 tree species according to local field inventory data ([Gácomo, 2018](#); [Martins et al., 2021](#)). The area receives 1278 mm of rain annually, and the mean annual temperature is $23.2 \pm 5.5^\circ\text{C}$.

2.2. Airborne images and LiDAR data

Aerial images were taken in October 2019 under clear sky conditions with the digital camera UltraCam-Eagle Prime (Vexcel, Inc.). The camera captures images with four bands (Red, Green, Blue, and Near-infrared - RGBNIR) with a radiometric resolution of 14 bits. The aerial photogrammetric mapping was planned to capture images with a GSD of 15 cm, strip overlap, and sidelap of 80% and 40%, respectively. Overlapping images were ortho-rectified through the method widely known as “true ortho,” i.e., based on a digital surface model (DSM), which includes buildings, tree canopies, and other artifacts beyond the surface, thus minimizing relief displacement due to parallax. For this orthorectification process, a LiDAR DSM was used.

LiDAR data was acquired by the Trimble Harrier 68i (Trimble Germany GmbH) sensor onboard an aircraft at a flight height of 700 m above sea level. The sensor operates with a frequency of 400 kHz (400.000 pulses per second), a wavelength of 1550 nm (near-infrared domain), and a field of view of 60°. The average density of the point cloud was 12 points/m². After post-processing procedures, which included using global navigation satellite systems (GNSS) and inertial measurement unit (IMU) data, the planimetric accuracy of each LiDAR point was 3 cm.

2.3. Individual tree crown dataset

This study used the ITC dataset of [Martins et al. \(2021\)](#). Based on field inventory data, the dataset comprises ITCs that have been visually delineated in aerial RGB images, followed by species-level identification.. The work of [Martins et al. \(2021\)](#) was based on aerial images acquired in 2015. Thus, we manually removed the ITCs not present in the images from 2019. Moreover, we adjusted the polygon boundaries of each ITC because of distortions or displacements between the images from 2015 and 2019. [Table 1](#) summarizes the species and ITCs used in this work. The decision to focus on these species was intentional and pivotal for assessing our deep-learning fusion strategy for tree species classification. Including species with fewer than 20 ITCs could potentially undermine the reliability and validity of our evaluation. Such a small sample size for certain species might not sufficiently represent their characteristics, leading to less accurate or biased classification results. By concentrating on species with a substantial number of ITCs, we can ensure a more robust and representative dataset, which is essential for effectively training and validating our deep-learning model.

Table 1
Tree species and number of individual tree crowns (ITCs).

Species	N° ITCs
<i>Terminalia catappa</i>	130
<i>Pachira aquatica</i>	35
<i>Licania tomentosa</i>	45
<i>Senna siamea</i>	24
<i>Tamarindus indica</i>	24
<i>Caesalpinia pluviosa</i>	25

2.4. LiDAR-derived features

For all LiDAR-derived features, we generated raster data. This process involved taking the individual data points and organizing them into a grid, with each cell representing a square area of 0.5×0.5 m. We then calculated the average value of the point features within each cell, resulting in a continuous raster surface. Finally, we resampled all raster data to 0.15×0.15 cm to match the GSD of the aerial images.

2.4.1. Surface normals in 3-D point clouds

3-D LiDAR point clouds can be used to describe the geometric surface properties of tree crowns. For example, one can compute surface normals of tree leaves and estimate species-specific leaf angle and orientation. However, the accuracy of these computed surface normals is highly dependent on the density of the LiDAR point cloud. Although studies leveraging airborne LiDAR for this purpose are limited—see Montagnoli et al. (2015), Castaño-Díaz et al. (2017), Petras et al. (2023)—this gap in research presents a compelling opportunity. It motivates our work to explore and validate the efficacy of airborne LiDAR in accurately computing surface normals, potentially offering a new dimension in tree species classification.

Our approach is two-fold. First, we define a sphere s around a query point p_q (see Fig. 1), which in turn returns a surface N with the corresponding neighboring points Rusu et al. (2008). Second, we estimate a plane tangent $\rightarrow n$ to the surface N in a least-squares sense. Thus, by assuming a query point p_q into a plane, its coordinates satisfy the following equation:

$$d = \rightarrow n^T p_q \quad (1)$$

where $p_q = [x_q, y_q, z_q]^T$ is a point in 3D space, $\rightarrow n = [n_x, n_y, n_z]^T$ denotes a normal vector of N , d represents the distance from a point $q \in N$ to the plane. The estimation of $\rightarrow n$ can be obtained by minimizing a closed-

form solution based on the residual sum of squares of each point $q \in N$, as follows Pathak et al. (2010):

$$\sum_{i=1}^k r_i^2 = \rightarrow n^T B \rightarrow n^T \quad (2)$$

where k is the number of neighboring points of p_q .

Let $B = \sum_{i=1}^k q_i^T q_i$ be a $m \times m$ symmetric matrix whose eigenvalues in descending order are $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ and their eigenvectors v_1, v_2, v_3, v_4 restricted to $\|v\| = 1$, related to Euclidean inner product of R^m : $\lambda_1 \geq v^T B v \geq \lambda_m$. Notably, $v^T B v = \lambda_1$ if v is an eigenvector of B associated with λ_m , and $v^T B v = \lambda_1$ when v is an eigenvector of B associated with λ_1 . Thus, the surface normal is obtained through the components of eigenvector v_4 , which corresponds to the smallest eigenvalue λ_4 of matrix B . As a result, a normal vector is assigned to p_q . This process repeats until no more query points can be found in the point cloud.

The normal vector can be decomposed by orientation in the n_x , n_y , and n_z directions. In addition, the surface curvature ($curv$) can be estimated by performing an eigenanalysis on the matrix B around p_q , as follows (Rusu et al., 2008):

$$curv = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} \quad (3)$$

2.4.2. Intensity

LiDAR intensity (int) measures the reflectivity of the laser pulse. It is proportional to the strength of the returns. Because the LiDAR sensor operates with pulses of 1550 nm and vegetation reflects this wavelength well (Allen et al., 1969), LiDAR intensity may contain information related to specific characteristics of the species. LiDAR intensity is recorded in the range [0–255], with low values indicating low and high values denoting high reflectivity.

The intensity of the laser returns is affected by several factors, such as

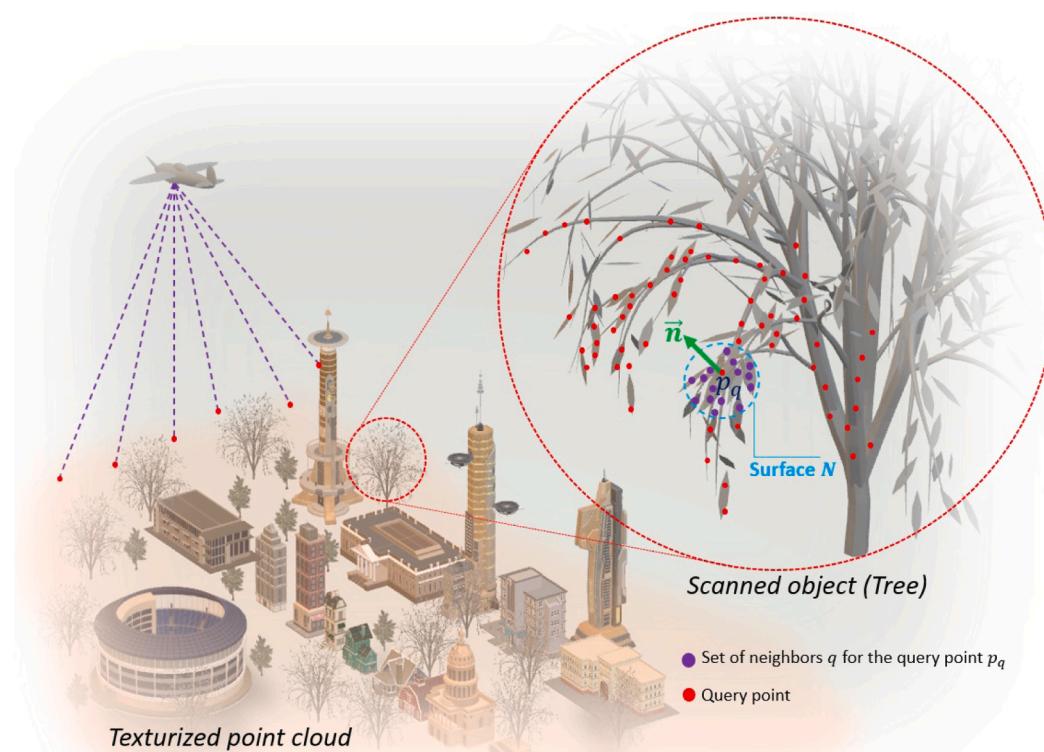


Fig. 1. Illustrative depiction of surface normal computation for tree leaves from 3D LiDAR point clouds. This schematic representation intentionally exaggerates the number of LiDAR points for clearer visualization. It should be noted that airborne LiDAR datasets typically exhibit lower point densities, generally less than 20 points/m², and therefore might not reflect the detail shown in this figure.

the range to the target, angle of incidence, and atmospheric dispersion. The reflection intensity of the laser beam decreases with the distance from the sensor. Nadir beams have more energy (higher intensity) than those at the edge of the field of view. Similarly, shorter ranges have higher intensity readings. We performed a dynamic range-based normalization of intensity to correct for range effects following the approach of [Gatziolis \(2011\)](#) implemented in the *lidR* ([Roussel et al., 2020](#)) R package.

2.4.3. Canopy height

Canopy height models (CHM) derived from LiDAR represent the actual height of objects. It varies according to the species, tree age, and environmental conditions, such as proximity to buildings and other artificial structures. To compute tree height, we first identified ground points using a cloth simulation filter ([Zhang et al., 2016](#)) and normalized the point cloud with the ground at 0.

2.4.4. Leaf area index

LAI is defined as the area of leaves per unit ground surface area ($m^2 \cdot m^{-2}$). It is considered a critical canopy structural parameter that drives ecophysiological processes such as radiation interception, water, and carbon exchange (Bréda, 2003). Field-based approaches to measure LAI are costly and labor-intensive because they require leaf harvesting and measurement. Indirect approaches based on discrete airborne LiDAR have been widely used to estimate LAI (Tian et al., 2021). LAI retrieval from LiDAR is based on computing laser pulses' penetration or transmittance ratio and then using the Beer-Lambert Law to convert transmittance to LAI (Nilson, 1971). We estimated LAI with the LiDAR point cloud following the approach proposed by Almeida et al. (2019). First, we binned the point cloud into canopy volume units (voxels) of 1 m^3 and estimated their leaf area density (LAD). LAD was calculated using the first returns within each voxel. Then, we estimated LAI by the sum of LAD's vertical profiles (voxel columns). Finally, we obtained a raster in which each cell contains the LAI information. We used the *leafR* (Almeida et al., 2019) R package.

2.5. CNN architecture and fusion strategy

We used the ResUnet architecture proposed by [Zhang et al. \(2018\)](#). We inserted an additional residual block before the last convolutional layer with *softmax*. The network comprises three modules ([Fig. 2](#)). The encoder performs feature extraction of the input image. The decoder upsamples the feature maps from the encoder, and the classifier module performs semantic segmentation. The residual blocks comprise 3×3 convolutional layers, batch normalization (BN), and rectified linear unit (ReLU) operations. The identity mapping in the residual blocks consists

of 1×1 convolutions.

To fuse RGBNIR images and LiDAR-derived features, we trained two encoder-decoder networks, one for each data type (Fig. 3). The decoder's outputs (feature maps) of each network were concatenated along the third dimension and served as input to the classifier module (red dotted box in Fig. 2).

2.6. Experimental setup

First, we randomly selected 60% of the ITCs for model training and 40% to assess the classification accuracy. We extracted patches of size 128×128 pixels with an overlap of 90% from the training ITCs to feed the network. We split these patches into training/validation sets in a proportion of 70%/30%. We used the cross-entropy loss function with the hyper-parameter β with a value of 0.9 and the Adam optimizer (Kingma and Ba, 2014). We trained the model with seven classes (six tree species + urban background) and employed random rotations and horizontal and vertical flips for data augmentation. We used a batch size of 32 and trained the model for 500 epochs with an early stop if no improvements greater than 0.00009 were observed in the average F1-score. We performed patch-wise predictions using sliding windows with overlaps of 10%, 30%, 50%, and 70%.

The baseline model was trained using the RGBNIR bands only. To assess the utility of each LiDAR-derived feature (see section 2.4 for a detailed explanation of each LiDAR-derived feature) to discriminate the species, we stacked the RGBNIR bands and one LiDAR metric along the spectral dimension. We trained seven models: *RGBNIR_nx*, *RGBNIR_ny*, *RGBNIR_nz*, *RGBNIR_curvature*, *RGBNIR_intensity*, *RGBNIR_CHM* and *RGBNIR_LAI*. We trained each of these models five times with different training and testing ITCs, which allowed us to assess model generalization and the variability in classification accuracy. The data fusion model used the LiDAR metrics that did not reduce the Kappa index concerning the baseline model.

We used a desktop workstation with an Intel Core i9-12900 F 2.4 GHz CPU, 128 GB of main memory, and an NVIDIA GeForce Titan V GPU with 12 GB of dedicated memory for training and inference. All image processing procedures were implemented in *Python* with the *PyTorch* framework. The codes used in this work are freely available in the GitHub repository: https://github.com/felferrari/tree_fusion.

2.7. Post-processing procedure

The post-processing procedure aims to produce tree species maps at the ITC level. We performed ITC delineation in the aerial image and assigned a class label to each ITC based on the output of the classification model.

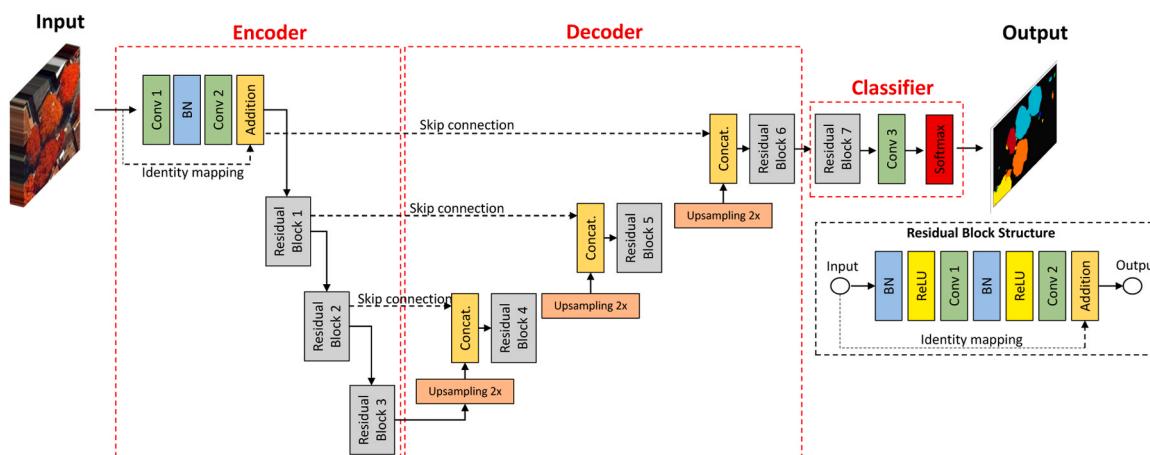


Fig. 2. Residual U-net (ResUNet) architecture. The network receives an input image patch that passes through the encoder module for feature extraction. The decoder performs upsampling to recover the original dimension of the image patch. Then, the softmax classifier is applied to assign a label to each image pixel.

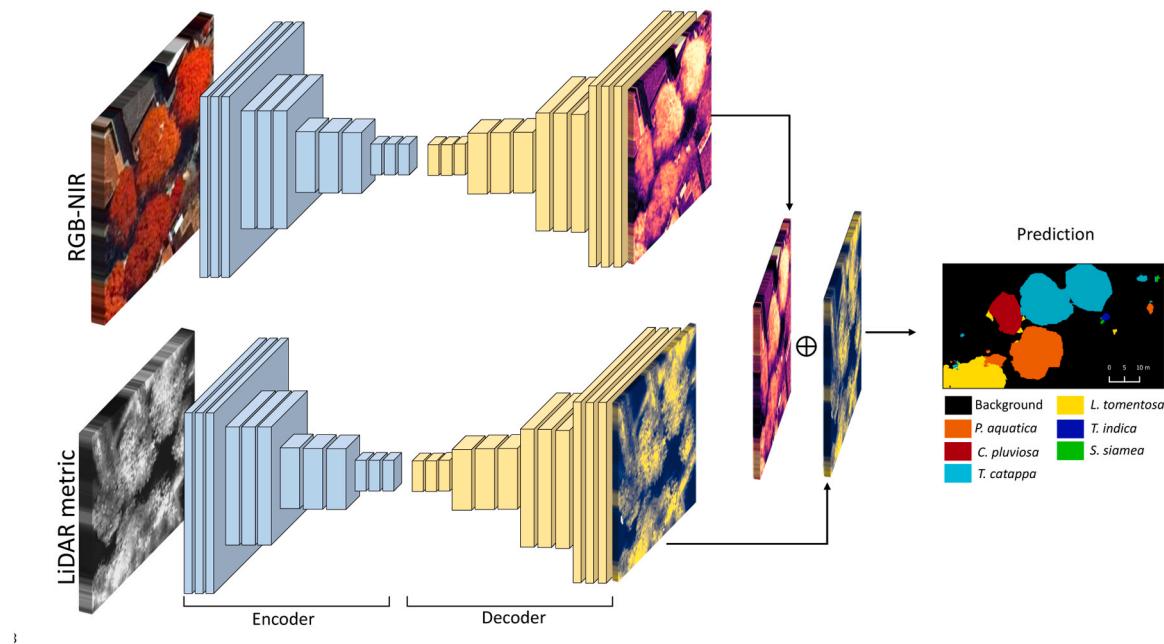


Fig. 3. Data fusion strategy used in this work. Two ResNet encoder-decoder architectures (Fig. 2) are trained with the RGBNIR image and LiDAR-derived metrics separately. Each network decoder output (feature maps) was concatenated and input in the classifier module for semantic segmentation.

2.7.1. ITC delineation

We first used the Segment Anything (SAM) (Kirillov et al., 2023) to delineate objects in an RGB image composition. SAM, a creation of Meta AI, represents a significant advancement in image segmentation. It showcases remarkable generalization abilities across various image datasets. Uniquely, SAM requires no extra training to identify and segment unfamiliar objects accurately, distinguishing itself as a highly adaptable and efficient tool in remote sensing image analysis (e.g., Osco et al. (2023)). We used the Segment-Geospatial (*samgeo*) Python package (Wu and Osco, 2023) to employ the SAM algorithm in the digital aerial images to delineate ITCs. This package is designed to adapt the SAM algorithm for remote sensing imagery. It includes tools that apply SAM to overlapping patches of an image, a technique that effectively eliminates the patchy appearance often encountered in segmentation tasks using CNNs. This approach ensures a smoother, more continuous output, enhancing the accuracy and visual quality of the delineated ITCs in the aerial imagery. We used the ViT-H SAM model (see Kirillov et al. (2023) for more details) with default parameters.

After segmenting the aerial imagery, we computed each segment's average normalized difference vegetation index (NDVI). This process was essential in identifying non-tree segments. By employing an NDVI threshold of 0.5, we could filter out segments not representative of trees. This particular threshold was carefully chosen to optimize identifying a substantial number of tree-representative polygons while distinguishing between arboreal and non-arboreal surface types within the image. A lower threshold might include non-tree elements, whereas a higher threshold could omit trees with lower leaf density.

2.7.2. ITC classification

After segmenting the ITCs, we classified them by assigning a species label based on the predictions from our network model (refer to Fig. 3 for an illustration of the raw model output). Each SAM-delineated ITC received a label corresponding to the most frequent species class within its boundaries. Based on this approach, we produced a reliable species map that shows the spatial distribution of the target species at the ITC level over the study area.

2.8. Accuracy assessment

We assessed the classification accuracy by computing the F1-score (harmonic mean of precision and recall values) with the ITCs not used for model training. To determine the ITC segmentation accuracy, we computed the boundary F1 (BF, Csurka et al. (2013)) contour matching score. The BF score measures the alignment between the predicted and manually delineated boundaries of the ITCs. It extends the F1-score to semantic segmentation. The F1 score is computed with a distance error tolerance threshold to decide whether a pixel in the predicted boundary matches the manually delineated boundary. The BF score varies in the range of 0–100%, in which 100% means a perfect match between the predicted and manually outlined boundaries of the ITCs.

3. Results

The F1-scores and Kappa indexes for each species obtained by training the ResNet (Fig. 2) model with the RGBNIR bands alone and combined with the LiDAR-derived features are shown in Table 2. The *nz*, *curv*, and *intensity* features improved the Kappa index. In general, the F1-scores of all species increased with the addition of LiDAR metrics in the classification process. The highest increase (10.5 percentage points) was observed for *C. pluviosa* using *curvature*. Conversely, CHM and LAI decreased the classification accuracy for most species with a reduction in the Kappa index compared with RGBNIR of 8.8 and 4.9 percentage points, respectively.

It is worth noting the variability in F1-score among the species. *P. aquatica* showed the highest standard deviations among realizations with the RGBNIR bands, which shows high variability in structural characteristics among ITCs. The F1-score of this species was also the lowest on average (56.3%, Table 2). Conversely, *T. catappa* showed the highest F1-scores and the lowest variability in classification accuracy.

Visual examples of the LiDAR-derived metrics are shown in Fig. 4. Again, differences among the species are notable. For example, low *curvature* values are observed in the *C. pluviosa* tree if compared with *T. catappa* and *L. tomentosa* (Fig. 4e). It is also visible within-species differences, particularly in the CHM image in which one ITC of *T. catappa* is higher than the other.

In Fig. 5, we show the variability in the LiDAR features. The *nx* and

Table 2

F1-scores and Kappa indexes for each tree species were obtained using the RGBNIR bands alone and combined with LiDAR-derived features. The LiDAR features that improved Kappa if compared to RGBNIR are highlighted in bold.

Species	RGBNIR	RGBNIR_nx	RGBNIR_ny	RGBNIR_nz	RGBNIR_curvature	RGBNIR_intensity	RGBNIR_CHM	RGBNIR_LAI
<i>T. catappa</i>	87.0 ± 4.8	90.5 ± 1.4	88.4 ± 1.9	90.9 ± 2.2	90.5 ± 1.5	87.8 ± 4.1	87.0 ± 5.5	87.1 ± 3.3
<i>P. aquatica</i>	56.3 ± 13.7	59.3 ± 16.9	52.2 ± 18.3	54.5 ± 14.8	56.0 ± 19.3	60.9 ± 20.0	40.0 ± 7.0	47.6 ± 9.6
<i>L. tomentosa</i>	76.9 ± 4.4	66.7 ± 7.0	69.0 ± 5.8	75.9 ± 24.8	75.0 ± 4.6	78.6 ± 3.2	65.1 ± 12.6	71.1 ± 6.2
<i>S. siamea</i>	72.7 ± 9.6	66.7 ± 15.9	76.9 ± 20.2	72.7 ± 8.5	66.7 ± 8.6	72.7 ± 10.7	50.0 ± 27.2	58.8 ± 14.1
<i>T. indica</i>	72.7 ± 8.0	76.2 ± 4.5	66.7 ± 7.8	66.7 ± 8.1	75.0 ± 3.3	63.6 ± 8.0	58.8 ± 6.6	63.2 ± 7.2
<i>C. pluviosa</i>	63.2 ± 10.6	70.0 ± 15.1	70.0 ± 12.4	63.2 ± 10.9	73.7 ± 12.1	57.1 ± 16.9	63.6 ± 22.2	44.4 ± 23.0
Kappa	68.7 ± 8.7	67.5 ± 5.7	67.7 ± 5.7	72.9 ± 9.5	71.5 ± 6.5	69.9 ± 9.2	59.9 ± 10.6	63.7 ± 6.4

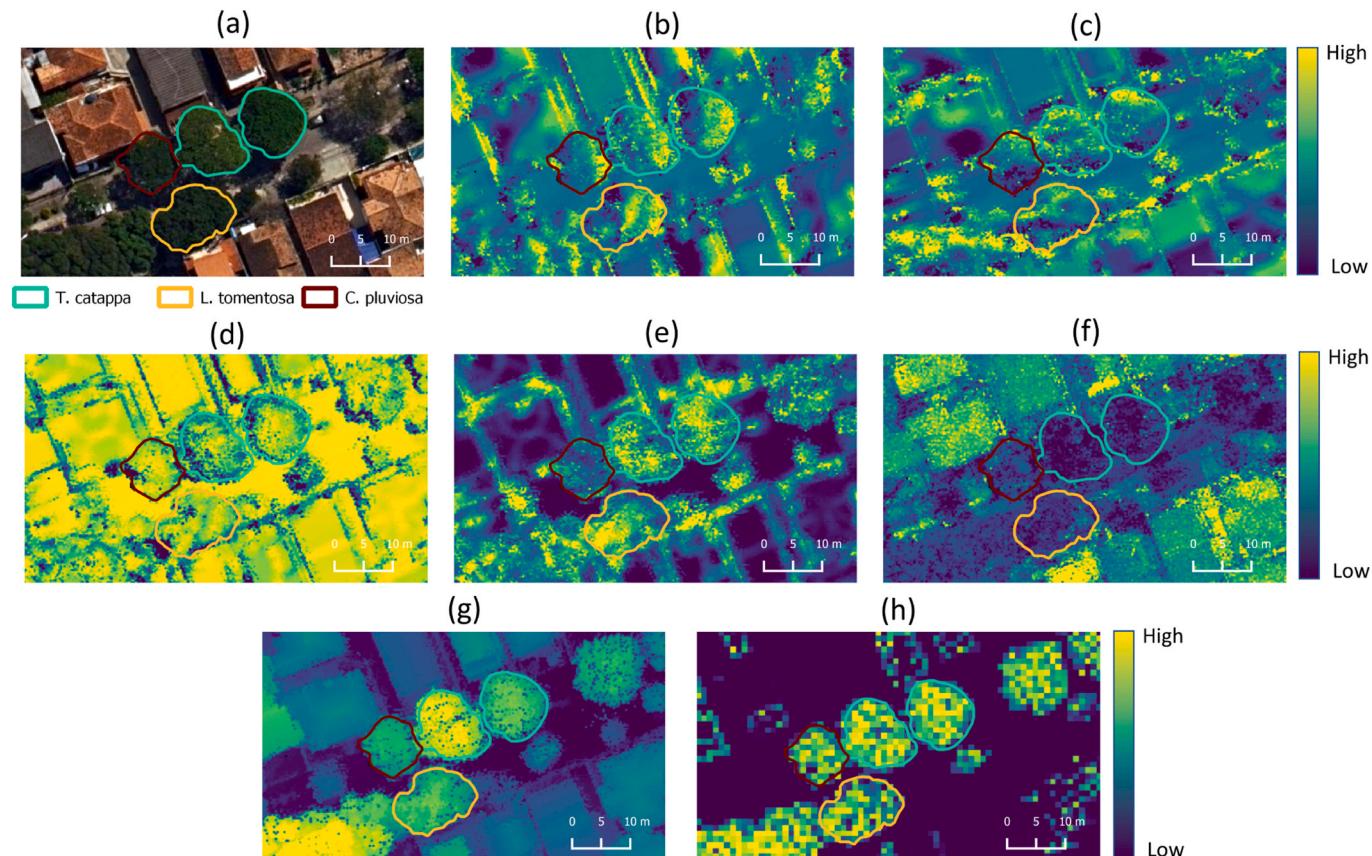


Fig. 4. Examples of the LiDAR-derived features used in this study. (a) RGB composition of the digital aerial image. Surface normals of tree leaves: (b) nx, (c) ny, (d) nz, (e) curvature; (f) intensity image; (g) canopy height model and (h) leaf area index image. Individual tree crowns of *T. catappa*, *L. tomentosa* and *C. pluviosa* are overlaid on each image to help visualize species-specific differences in the LiDAR metrics.

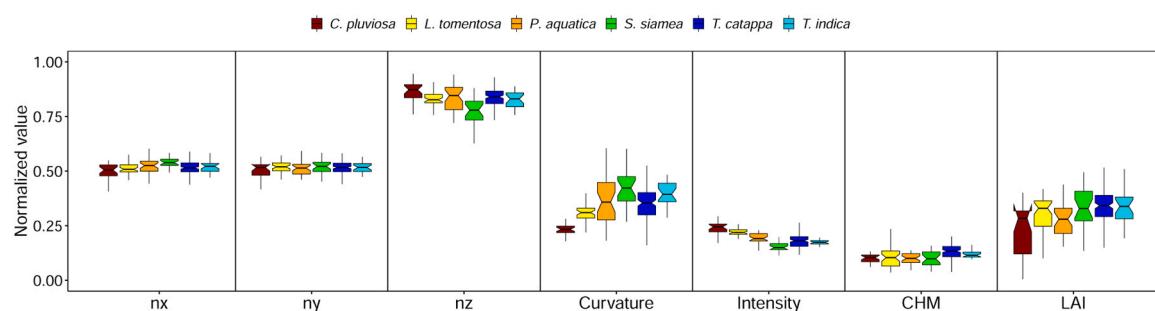


Fig. 5. Boxplots showing the variability in the LiDAR-derived features per species. The central line within each box represents the median, while the boxes' edges represent the upper and lower quartiles. LiDAR metrics were normalized to the [0,1] range for clarity.

ny did not change significantly among species, while nz and curvature and intensity showed variable patterns. The canopy height was similar among the species, except *T. catappa* that was the species with the highest ITCs. The leaf area index among the species did not change significantly on average but showed high within-species variability.

The strategy to fuse RGNIR bands with the LiDAR-derived metrics (Section 2.5, Fig. 3) improved the F1-score of all species (Fig. 6). The average of the F1-score was 71.5% and 84.1% using RGNIR and the fusion approach, respectively. The highest F1-score increase was observed for *T. indica* and reached 19.54 percentage points. The lowest increases were for *T. catappa* and *S. siamea* with 3.34 and 6.05 percentage points, respectively.

The results of post-processing (Section 2.7) are shown in Fig. 7. One can note that ITCs were accurately delineated by the SAM algorithm (Fig. 7a). The BF score reached 98%. The original model prediction (Fig. 7c) labeled spurious pixels within ITCs. For example, pixels of *P. aquatica* were labeled within an ITC of *T. catappa* (see the white arrow in Fig. 7c). After applying the majority rule within SAM-detected ITCs, each ITC was assigned to a given species, improving the classification accuracy.

4. Discussion

4.1. LiDAR-derived metrics for tree species classification

This is the first study to use surface normals of 3D LiDAR point clouds to classify tropical tree species. We showed that ITCs from the investigated species have significantly different values of nz and curvature, which confirms our first hypothesis. Such metrics provide information about the slope and local shape of the better-fitted surface by influencing the quality of tree species classification. Furthermore, surface curvature helps determine LiDAR points with high curvature values, representing the leaves' geometric features. We show that the curvature of ITCs is related to species-specific crown architectural patterns. Crown architecture refers to the arrangements of leaves and branches and depends, among other factors, on species identity. Trees of *C. pluviosa* showed the lowest curvature values, which suggests that its crowns are more planar than the other species. Conversely, *S. siamea* showed high and more variable curvature values, indicating a particular architectural pattern.

The effectiveness of LiDAR-derived surface normals for species discrimination in temperate forests has been debated. While some studies have demonstrated promising results, indicating a positive impact on species identification, others have found no significant improvements. For example, Barilotti et al. (2009) used LiDAR-derived curvature to classify coniferous and broad-leaf trees. The authors found that curvature is highly affected by point cloud density and achieved classification accuracies between 85–95% with laser point densities higher than four points/m². The study of Briechele et al. (2020) explored integrating LiDAR geometry features and surface normals for classifying pine, birch, alder, and standing dead trees in Ukraine.

Incorporating surface normals alongside geometry features led to a nuanced impact on classification efficacy, as reflected in the F1-scores. Specifically, there was a slight decline in the F1-score for pine by 3.3%. Conversely, their approach yielded improvements for the other species, with birch, alder, and dead trees witnessing an increase in their F1-scores by 2.7, 1.2, and 5.5 percentage points, respectively. Another study performed in European forests by Hell et al. (2022) estimated normals from airborne LiDAR data and used them to distinguish between coniferous and deciduous trees in European forests. The authors found that the normals do not significantly contribute to the classification results.

The intensity of LiDAR returns is related to the reflectance properties of leaves and branches and has been explored in tree species discrimination studies. Kim et al. (2009), in a study performed in an urban green space in Seattle, Washington, USA, found that the arrangement of foliage and branches within ITCs of different species produced variation in intensity values. Similarly, Alonso et al. (2014) found that intensity metrics varied significantly among species of urban trees in Santa Barbara, California, USA. Our results corroborate previous studies showing that intensity values varied among species. Moreover, the F1-score improved after including the intensity image in the classification process.

Tree height is the most common LiDAR-derived metric for classifying tree species (Li et al., 2019). In our study, tree height negatively impacted the classification accuracy as most species did not show differences in the CHM. Urban trees in the study area were planted more than 50 years ago, and dead or fallen trees were not replaced. The LAI, estimated for each ITC, also reduced the classification accuracy according to our experiments. Although the species showed variable LAI patterns, the high within-species variability hampered accurate discrimination.

4.2. Optical-LiDAR fusion

The object-based approach is the predominant methodology for fusing optical imagery and LiDAR data to classify tree species. This approach unfolds in three distinct steps: (i) object segmentation, (ii) feature extraction, and (iii) species classification. First, segmentation algorithms delineate ITCs from optical or LiDAR data. Then, structural and spectral features are extracted from the delineated ITCs. Finally, the extracted features are used by a classifier to assign each ITC to its respective tree species. Object-based approaches show significant limitations, including empirical parameter tuning of segmentation algorithms and manual feature extraction. CNNs overcome these limitations of object-based approaches by integrating feature extraction and species classification in a single approach (Kattenborn et al., 2021). Because CNNs rely on local linear operations, known as convolutions, they are powerful feature extractors (Zhang et al., 2020).

Using CNNs to fuse optical and LiDAR data usually involves training a single encoder-decoder architecture by stacking optical image bands

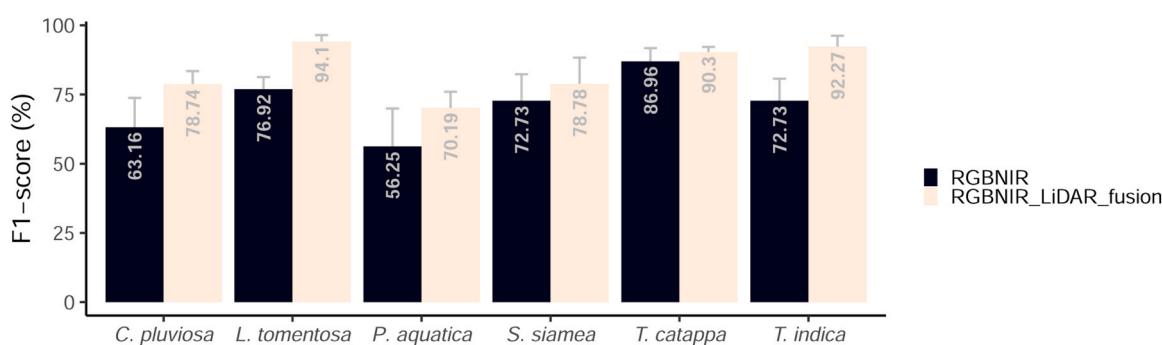


Fig. 6. F1-scores for each tree species obtained by the ResNet (Fig. 2) model trained with RGNIR bands (black color bars) and RGNIR bands fused with LiDAR-derived features (beige color bars). The fusion model strategy is shown in Fig. 3.

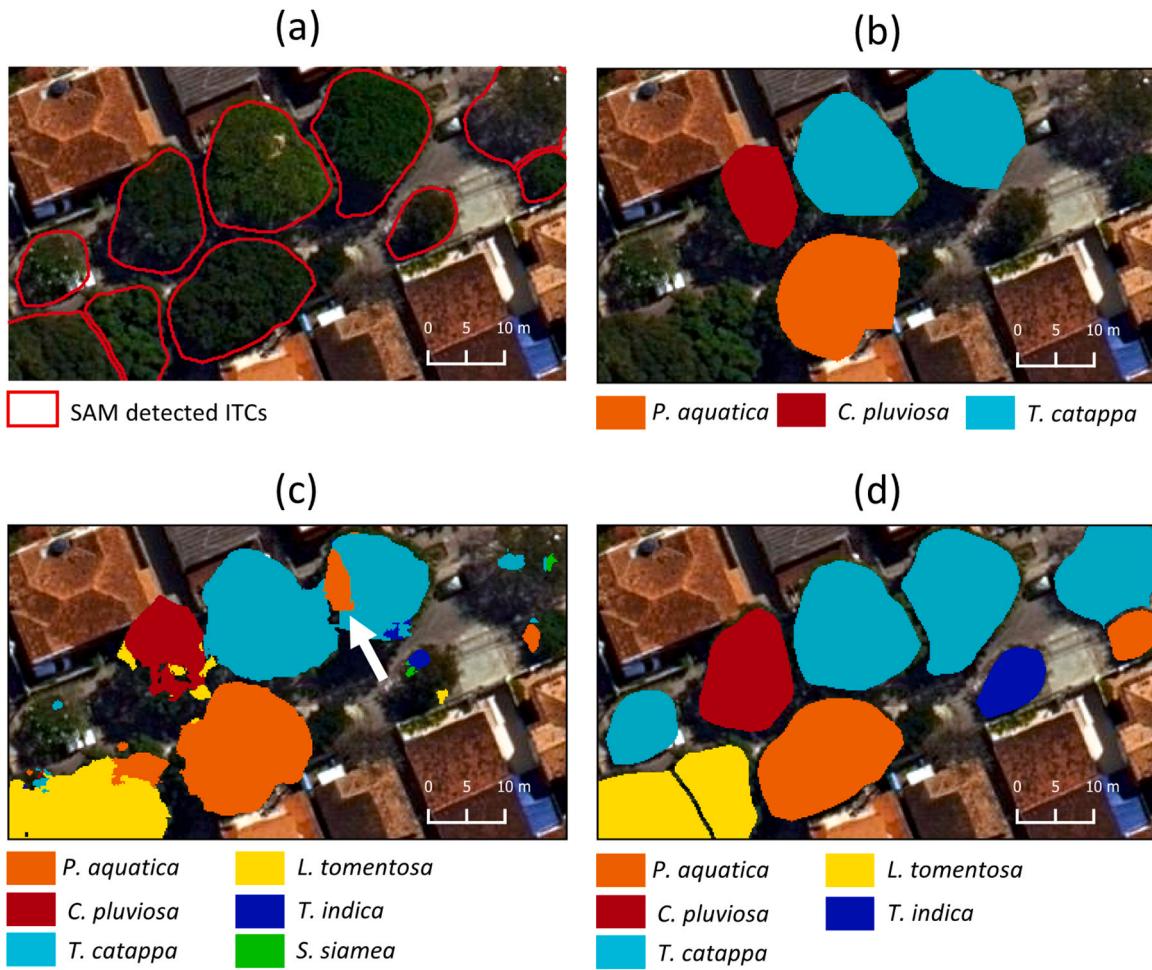


Fig. 7. Results of post-processing. (a) ITCs are automatically delineated by the Segment Anything (SAM) (Kirillov et al., 2023) model (red polygons). (b) reference ITCs. (c) original network prediction of the data fusion model (Section 3), before post-processing (Section 2.7) (d) ITCs classified to the species level using the original network prediction (c) and SAM detected ITCs (a), i.e., after post-processing. The white arrow in (c) points to pixels of *P. aquatica* labeled within an ITC of *T. catappa* before post-processing.

and LiDAR-derived images. For example, Hartling et al. (2019) stacked WorldView-2, WorldView-3, and a LiDAR return intensity image to classify urban tree species in St. Louis, MO, USA. Utilizing a CNN model designed for scene classification, they showed an improvement of 0.05 in the Kappa index when LiDAR data was combined with optical bands to discriminate among eight tree species. In our case, stacking LiDAR features to the RGBNIR bands decreased the classification accuracy in some cases if compared to using only RGBNIR bands. However, fusing the feature maps derived from two encoder-decoder networks trained with optical and LiDAR data confirms our second hypothesis. New approaches based on transformers have provided promising results for multi-source remote sensing classification (e.g., Zhang et al. (2023b)) and will be explored in future studies.

4.3. The species mapping

Our study shows an innovative approach integrating the novel SAM algorithm for producing species maps at the ITC level. The use of SAM to delineate tree crowns was successful in reducing the raw, noisy predictions of our CNN model (Fig. 7c) and producing a reliable species map (Fig. 7d). The application of SAM in remote sensing is a burgeoning field, currently marked by a limited but growing number of studies that have utilized it. Osco et al. (2023) tested SAM across multi-scale datasets, including training-free personalization approaches proposed by Zhang et al. (2023a), which enhanced SAM's capabilities, rendering it a more

adaptable model.

4.4. Limitations and future directions

One of the primary limitations of this study is the reliance on low-density airborne LiDAR data for computing surface normals. While our analysis indicates significant differences among tree species based on this data, there is an inherent limitation in the point density of our LiDAR dataset. The density might need to be increased to capture the full range of variation in surface normals, which is crucial for accurate species identification and differentiation. Future research should focus on evaluating point density's impact in retrieving surface normals. This would involve using high-density LiDAR data to determine whether increased point density can lead to more nuanced and accurate surface normal variations, thereby enhancing species classification accuracy.

The current application of the SAM algorithm in our research presents another avenue for improvement. Following the findings of Osco et al. (2023), integrating the personalization approaches proposed by Zhang et al. (2023a) could significantly enhance the performance of the SAM model. These personalization techniques, tailored to the specific characteristics of the dataset, could refine the model's application, leading to more accurate and reliable results in remote sensing and species classification tasks. Future studies should explore incorporating these personalization methods into the SAM model, assessing potential model flexibility and accuracy improvements.

Lastly, the burgeoning field of transformers in neural networks offers a promising direction for future research. Transformers have shown exceptional performance in various domains, including natural language processing and image recognition (Han et al., 2022), and their application in remote sensing and species classification represents an untapped potential. Future studies may explore the implementation of transformer networks to understand their efficacy in handling the complex spatial and spectral relationships inherent in remote sensing data. This could include comparing the performance of transformer networks with traditional CNN architectures and assessing whether transformers can better capture the intricate patterns and variations crucial for accurate species classification in remote sensing.

5. Conclusions

This study marks a significant advancement in using airborne LiDAR point clouds for tree species classification. Our findings demonstrate that surface normals from these point clouds effectively capture species-specific differences in crown structure, thereby enhancing tree species classification. Additionally, we observed notable variations in the intensity of LiDAR returns among different species, which further contributed to classification accuracy. Our innovative optical-LiDAR fusion strategy, which combines feature maps from encoder-decoder networks trained separately with optical and LiDAR-derived metrics, substantially improved species classification. This approach yielded a 12.6 percentage points increase in the average F1-score compared to classifications using only RGBNIR bands. Moreover, the successful application of the SAM algorithm in delineating ITCs in an RGB composition of the digital aerial image produced accurate species maps, highlighting the algorithm's effectiveness in reducing the noise in predictions from our CNN model. Our research also encountered limitations, primarily the low point density of our LiDAR data and the potential for further enhancement of the SAM algorithm, as indicated by recent studies. Exploring advanced deep learning techniques, such as transformer networks, presents a future direction for improving species classification and mapping.

Funding

This study was funded by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) grants SEI-260003/004764/2021 and SEI-260003/002763/2022. MPF and DRS were supported by the Brazilian National Council for Scientific and Technological Development (CNPq) grants 306345/2020-0 and 305147/2022-6, respectively.

CRediT authorship contribution statement

Coelho Filho Luiz Carlos Teixeira: Data curation, Resources, Writing – original draft, Writing – review & editing. **Martins Gabriela Barbosa:** Data curation, Validation, Visualization. **dos Santos Daniel Rodrigues:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. **Ferrari Felipe:** Data curation, Methodology, Software, Writing – original draft, Writing – review & editing. **Feitosa Raul Queiroz:** Supervision, Writing – original draft, Writing – review & editing. **Ferreira Matheus Pinheiro:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research and for the NVIDIA Applied Research Accelerator Program.

References

- Allen, W.A., Gausman, H.W., Richardson, A.J., Thomas, J.R., 1969. Interaction of isotropic light with a compact plant leaf. *Josa* 59, 1376–1379.
- Almeida, D.R.A.d., Stark, S.C., Shao, G., Schiatti, J., Nelson, B.W., Silva, C.A., Gorgens, E.B., Valbuena, R., Papa, D.D.A., Brancalion, P.H.S., 2019. Optimizing the remote detection of tropical rainforest structure with airborne lidar: leaf area profile sensitivity to pulse density and spatial sampling. *Remote Sens.* 11, 92.
- Alonso, M., Bookhagen, B., Roberts, D.A., 2014. Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sens. Environ.* 148, 70–83. <https://doi.org/10.1016/j.rse.2014.03.018>.
- Bariotti, A., Crosilla, F., Sepic, F., 2009. Curvature analysis of lidar data for single tree species classification in alpine latitude forests. *Laser Scanning* 1–2. <https://doi.org/10.5194/isprs-archives-XLII-2-W13-1077-2019>.
- Bréda, N.J., 2003. Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. *J. Exp. Bot.* 54, 2403–2417.
- Briechle, S., Krzystek, P., Vosselman, G., 2020. Classification of tree species and standing dead trees by fusing uav-based lidar data and multispectral imagery in the 3d deep neural network pointnet++. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* 2, 203–210. <https://doi.org/10.5194/isprs-annals-V-2-2020-203-2020>.
- Cariñanos, P., Calaza-Martínez, P., O'Brien, L., Calfapietra, C., 2017. The Cost of Greening: Disservices of Urban Trees. Springer International Publishing, Cham, pp. 79–87. https://doi.org/10.1007/978-3-319-50280-9_9.
- Castaño-Díaz, M., Álvarez-Álvarez, P., Tobin, B., Nieuwenhuis, M., Afif-Khoury, E., Cámaras-Obregón, A., 2017. Evaluation of the use of low-density lidar data to estimate structural attributes and biomass yield in a short-rotation willow coppice: an example in a field trial. *Ann. For. Sci.* 74, 1–16.
- Csurka, G., Larlus, D., Perronnin, F., Meylan, F., 2013. What is a good evaluation measure for semantic segmentation? In: Bmve, Bristol. 10–5244.
- Gatziolis, D., 2011. Dynamic range-based intensity normalization for airborne, discrete return lidar data of forest canopies. *Photogramm. Eng. Remote Sens.* 77, 251–259.
- Giácomo, R.G., 2018. Arborio: Sistema de gestão da arborização urbana. (https://apps.data.rio/datiaroresources/frames/palestrantes.Siurb/ppt/dia3/TB17_RomuloGuima%r%C3%A3esGiácomo.pptx).in: II Seminário sobre o Sistema Municipal de Informações Urbanas.
- Grote, R., Samson, R., Alonso, R., Amorim, J.H., Cariñanos, P., Churkina, G., Fares, S., Thiec, D.L., Niinemets, I., Mikkelsen, T.N., Paoletti, E., Tiwary, A., Calfapietra, C., 2016. Functional traits of urban trees: air pollution mitigation potential. *Front. Ecol. Environ.* 14, 543–550. <https://doi.org/10.1002/fee.1426>.
- Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., et al., 2022. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 87–110. <https://doi.org/10.1109/TPAMI.2022.3152247>.
- Hartling, S., Sagan, V., Sidiķe, P., Maimaitijiang, M., Carron, J., 2019. Urban tree species classification using a worldview-2/3 lidar and lidar data fusion approach and deep learning. *Sensors* 19, 1284. <https://doi.org/10.3390/s19061284>.
- Hell, M., Brandmeier, M., Briechle, S., Krzystek, P., 2022. Classification of tree species and standing dead trees with lidar point clouds using two deep neural networks: pointnet and 3dmfv-net. *PGF-J. Photogramm. Remote Sens. Geoinf. Sci.* 90, 103–121.
- Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on convolutional neural networks (cnn) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* 173, 24–49. <https://doi.org/10.1016/j.isprsjprs.2020.12.010>.
- Ke, Y., Quackenbush, L.J., Im, J., 2010. Synergistic use of quickbird multispectral imagery and lidar data for object-based forest species classification. *Remote Sens. Environ.* 114, 1141–1154. <https://doi.org/10.1016/j.rse.2010.01.002>.
- Kim, S., McGaughey, R.J., Andersen, H.E., Schreuder, G., 2009. Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sens. Environ.* 113, 1575–1586. <https://doi.org/10.1016/j.rse.2009.03.017>.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R., 2023. Segment anything. *arXiv*: 2304.02643.
- Li, X., Chen, W.Y., Sanesi, G., Laforteza, R., 2019. Remote sensing in urban forestry: recent applications and future directions. *Remote Sens.* 11. <https://doi.org/10.3390/rs11101144>.
- Martins, G.B., La Rosa, L.E.C., Happ, P.N., Coelho Filho, L.C.T., Santos, C.J.F., Feitosa, R.Q., Ferreira, M.P., 2021. Deep learning-based tree species mapping in a highly diverse tropical urban setting. *Urban For. Urban Green.* 64, 127241. <https://doi.org/10.1016/j.ufug.2021.127241>.
- Montagnoli, A., Fusco, S., Terzaghi, M., Kirschbaum, A., Pflugmacher, D., Cohen, W.B., Scippa, G.S., Chiattante, D., 2015. Estimating forest aboveground biomass by low density lidar data in mixed broad-leaved forests in the italian pre-alps. *For. Ecosyst.* 2, 1–9.
- Nilson, T., 1971. A theoretical analysis of the frequency of gaps in plant stands. *Agric. Meteorol.* 8, 25–38.

- Oscó, L.P., Wu, Q., de Lemos, E.L., Gonçalves, W.N., Ramos, A.P.M., Li, J., Marcato, J., 2023. The segment anything model (sam) for remote sensing applications: From zero to one shot. *Int. J. Appl. Earth Obs. Geoinf.* 124, 103540 <https://doi.org/10.1016/j.jag.2023.103540>.
- Pathak, K., Birk, A., Vaškevičius, N., Poppinga, J., 2010. Fast registration based on noisy planes with unknown correspondences for 3-d mapping. *IEEE Trans. Robot.* 26, 424–441.
- Petras, V., Petrasova, A., McCarter, J.B., Mitasova, H., Meentemeyer, R.K., 2023. Point density variations in airborne lidar point clouds. *Sensors* 23, 1593.
- Rahman, M.A., Stratopoulos, L.M., Moser-Reischl, A., Zöchl, T., Häberle, K.H., Rötzer, T., Pretzsch, H., Pauleit, S., 2020. Traits of trees for cooling urban heat islands: a meta-analysis. *Build. Environ.* 170, 106606.
- Roussel, J.R., Auty, D., Coops, N.C., Tompalski, P., Goodbody, T.R., Meador, A.S., Bourdon, J.F., de Boissieu, F., Achim, A., 2020. lidr: an r package for analysis of airborne laser scanning (als) data. *Remote Sens. Environ.* 251, 112061 <https://doi.org/10.1016/j.rse.2020.112061>. (<https://www.sciencedirect.com/science/article/pii/S0034425720304314>).
- Rusu, R.B., Marton, Z.C., Blodow, N., Dolha, M., Beetz, M., 2008. Towards 3d point cloud based object maps for household environments. *Robot. Auton. Syst.* 56, 927–941.
- Shah Tahmassebi, A.R., Li, C., Fan, Y., Wu, Y., lin, Y., Gan, M., Wang, K., Malik, A., Blackburn, G.A., 2021. Remote sensing of urban green spaces: a review. *Urban For. Urban Green.* 57, 126946 <https://doi.org/10.1016/j.ufug.2020.126946>.
- Smithers, R.J., Doick, K.J., Burton, A., Sible, R., Steinbach, D., Harris, R., Groves, L., Blicharska, M., 2018. Comparing the relative abilities of tree species to cool the urban environment. *Urban Ecosyst.* 21, 851–862.
- Tian, L., Qu, Y., Qi, J., 2021. Estimation of forest lai using discrete airborne lidar: a review. *Remote Sens.* 13. <https://doi.org/10.3390/rs13122408>. (<https://www.mdpi.com/2072-4292/13/12/2408>).
- Wang, K., Wang, T., Liu, X., 2018. A review: Individual tree species classification using integrated airborne lidar and optical imagery with a focus on the urban environment. *Forests* 10, 1. <https://doi.org/10.3390/f10010001>.
- Wu, Q., Oscó, L.P., 2023. samgeo: a python package for segmenting geospatial data with the segment anything model (sam). *J. Open Source Softw.* 8, 5663. <https://doi.org/10.21105/joss.05663>.
- Zhang, K., Hu, B., 2012. Individual urban tree species classification using very high spatial resolution airborne multi-spectral imagery using longitudinal profiles. *Remote Sens.* 4, 1741–1757.
- Zhang, R., Jiang, Z., Guo, Z., Yan, S., Pan, J., Dong, H., Gao, P., Li, H., 2023a. Personalize segment anything model with one shot. *arXiv:2305.03048*. 10.48550/arXiv.2305.03048.
- Zhang, S., Zhang, S., Wang, B., Habetler, T.G., 2020. Deep learning algorithms for bearing fault diagnostics—a comprehensive review. *IEEE Access* 8, 29857–29881. <https://doi.org/10.1109/ACCESS.2020.2972859>.
- Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X., Yan, G., 2016. An easy-to-use airborne lidar data filtering method based on cloth simulation. *Remote Sens.* 8, 501.
- Zhang, Y., Peng, Y., Tu, B., Liu, Y., 2023b. Local information interaction transformer for hyperspectral and lidar data classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* <https://doi.org/10.1109/JSTARS.2022.3232995>.
- Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual U-Net. *IEEE Geosci. Remote Sens. Lett.* 15, 749–753. <https://doi.org/10.1109/LGRS.2018.2802944>. (<http://ieeexplore.ieee.org/document/8309343/>).