2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

Tree Canopy Segmentation and Characterization Using LiDAR for Machine Learning Models

Jesús A. Monroy-Anieva [1], Alejandro Téllez-Quiñones [1], Rodrigo Lopez-Farías [2] and Hipólito Aguilar-Sierra [3]

Abstract— **This work proposes an alternative solution to address the problem of tree canopy classification, given a digital elevation model (DEM) of the study area obtained using a Light Detection and Ranging (LiDAR) device. This proposal presents a comprehensive methodology that enables the feasible application of machine learning (ML) models for tree canopy classification. Our approach allows to obtain a composite DEM of the tree canopy, which will be characterized using a tree cover selection and extraction strategy based on Euclidean distance and feature point clustering. Subsequently, statistical and geometric measures of the point clouds are calculated for each segmented canopy of the DEM in order to prepare a dataset, which includes some tree canopy features represented by column vectors of 77 predictors. Finally, the generated dataset was trained and validated using the MatLab's classification learner, obtaining considerable performance results in canopy classification, highlighting the improvement in classification performance of models based on Naive Bayes, Neural Networks, and Support Vector Machines. Furthermore, beyond its technical scope, this research has a direct social and environmental impact by promoting urban resilience, biodiversity conservation, and strengthening sustainable development policies. The proposed methodology contributes to global agendas such as the United Nations Sustainable Development Goals (SDGs) and the Mexican National Strategic Programs (PRONACES), positioning remote sensing and artificial intelligence as tools for climate action and social well-being.**

I. INTRODUCTION

In the remote sensing literature, a wide variety of technologies have been employed for geospatial applications, including capture devices at different amplitude scales, such as multispectral systems, radar, and LiDAR for certain spatial resolutions, as well as optical images reaching around 30 m^2 or more per pixel in mapped area analyses. Furthermore, advances in computer vision methods, such as stereo vision, structure-from-motion, and 3D reconstruction, as well as LiDAR systems, have enabled the digital capture of study regions in greater detail through the operation of remote vehicles that have reached areas of interest in aerial, terrestrial, and underwater environments [1], [2], [3].

[1]Jesús A. Monroy-Anieva & Alejandro Téllez-Quiñones are with Remote Sensing of Center for Research in Geospatial Information Sciences, Merida-Yuc., CP. 97302, Mexico jmonroy@centrogeo.edu.mx & atellez@centrogeo.edu.mx

[2]Rodrigo Lopez-Farías is with Artificial Intelligence of Center for Research in Geospatial Information Sciences, Queretaro-Qro., CP. 76703, Mexico rlopez@centrogeo.edu.mx

[3]Hipólito Aguilar-Sierra is with the Engineering Department, Universidad La Salle, Mexico City, CP. 06140, Mexico hipolito.aguilar@lasalle.mx

The availability of satellite data has been useful for Earth monitoring and has driven the development of methods to understand nature's behavior through the occurrence of its critical phenomena [4], [5]. These data are limited in time and capture scale, so the success of the implemented techniques has also been limited. In contrast, close monitoring of local and in situ regions appears to have greater scope, facilitating the generation of data samples provided by the use of various technologies, from traditional cameras to specialized devices such as hyperspectral cameras and LiDAR, subject to the impact and scope of research, but limited in terms of time and space. The fundamental digitized output provided by remote sensing technologies consists of the generation of digital elevation models (DEMs). These products have been necessary to establish a spatial reference, first to delimit the study area and then to define the region of interest (RoI) relative to pattern identification. Furthermore, the production of DEMs depends on the scale resolution offered by the capture capabilities of the system used, as well as the level of detail of the available object detection. These applications range from the acquisition of large-scale cartographic data governed by satellite technologies, with relatively large pixel resolutions of about 30 square meters, to the acquisition of information provided by unmanned aerial vehicles (UAVs) with radar or LiDAR systems, where measurements are on the order of square centimeters per pixel.

Remote sensing applications depend on the analysis of the phenomenon to be addressed, primarily considering spatiotemporal dynamics and leveraging data acquisition using specialized technologies. In line with advances in the design of forestry applications for environmental protection, research ranges from the aerospace industry to terrestrial monitoring. Recent applications include land use land cover (LULC) mapping, deforestation, biomass analysis, and forest carbon estimation. In the case of tree canopy detection in a field or forest, the main limitation lies, again, in the level of detail available. Some recent methods implemented semi-automatically, leveraging digital image processing for canopy detection using satellite images, are found in [6]. Several of them are based on the local maxima neighbor method to locate the central tree and, subsequently, generate the tree insulator and extract it for characterization. On the other hand, other research focuses on tree detection, isolation, and extraction by developing methods based on 3D point clouds [7]. In this way, through the use of data obtained

978-1-6654-5784-2/26/$31.00 ©2026 IEEE

by photogrammetry techniques and LiDAR systems, the third dimension allows obtaining more information about the canopy structure. Each point cloud generation presents advantages and disadvantages in terms of data context, accuracy, and response efficiency. Therefore, once the dataset is generated, the detection, isolation, and extraction of individual trees must be performed by developing methods based on point cloud processing. Several of these methods have been implemented to obtain the best performance in estimating tree centers and, subsequently, in their extraction. Generally, these methods apply the relationship between point distances for clustering and tree isolator, while others are based on strategies designed using likelihood functions for estimating tree centers, as indicated in [8].

For forestry applications, canopy height models (CHMs) generated by LiDAR systems are classified based on footprint size. Small footprints measure a few centimeters, while large ones have an accuracy of tens of meters. This feature allows the definition of 3D objects embedded in CHMs for detection, so the extraction method should be determined based on the availability of the RoI, either for area-based applications or for individual trees. Different efforts have been made to design methodologies for datasets from real study regions in both applications [9], [10], [11]. However, some shortcomings in canopy development persist. On the one hand, the complete processing to obtain accurate CHMs, including height normalization, point cloud classification, digital model shifting, and spatial interpolation, which is usually performed sequentially, could introduce uncertainties such as noise and underestimation of tree heights. On the other hand, the accuracy of canopy isolation and extraction remains a real challenge, as its impact could affect diverse land covers, including forested and urban areas. In the first case, forest density, as well as foliage, tree species variety, and vegetation diversity, primarily display the superposition and dynamics of temporal stationarity to characterize trees by their morphological characteristics. In the second case, the generated digital height models could include several artificial objects, such as buildings, roads, vehicles, overhead cables, and others, which cause disturbances in the canopy extraction process. Furthermore, the problem of pattern identification could be even more complex, especially when tree detection and isolation is addressed by characterization methods based on statistical measures and machine learning techniques, since their performance in terms of accuracy and response time must be monitored and corrected to ensure its robustness. Some relative methodological approaches to address tree characterization using DEMs and 3D point clouds generated by LiDAR systems are presented in [12], [13], most of which are under continuous improvement.

Based on CHMs generated from 3D point clouds using photogrammetry techniques, the structure from motion (SfM) method has been traditionally applied and enhanced to obtain 3D reconstructions and orthophotographs of study regions. This approach presents several advantages in terms of implementation costs, thanks to the image management offered by traditional RGB cameras or specialized high-resolution cameras. A notable technical feature of this method consists of providing spectral information, such as RGB and near-infrared data, for each point of the captured environment. For example, RGB data contribute to the generation of CHMs, adding spectral information of each object included in them. This exceptional feature is useful to complement the characterization process by including vegetation spectral indices, considering the variety of canopies, to strengthen the validation of identification algorithms. According to recent research [14], it has been reported that the best results in the vegetation identification process are obtained using spectral data provided by the application of photogrammetry techniques. Furthermore, recent efforts in photogrammetry enhancement methods have addressed the simultaneous processes of 3D reconstruction and object identification using deep learning methods, currently known as semantic and instance segmentation [15], [16], [17], [18]. These advances have revolutionized the problem of pattern identification, applied to simple images and 3D point clouds during capture. While the cost of optical capture devices seems low, the hardware cost resulting from complex processing is higher, and the processing of categorized class identification could be limited [19], [20], [21], [22].

Due to the higher accuracy of the tree isolator, CHM production must be complemented with samples obtained by terrestrial laser scanning (TLS). This system has allowed to obtain greater knowledge about the structure and morphology of forest species, since not only canopy data are sufficient to characterize tree species, but also data on trunks, branches, and breast. In this way, tree characterization can include different forest metrics, such as allometry and dendrometric parameters, which allow to better distinguish individual tree species. In recent approaches, the most popular processing method for diameter estimation consists of extracting cross-sectional data from the original point cloud and obtaining a stem model using point cloud clustering or circle search [23]. This approach assumes that all trees have a clear stem at the same cross-sectional height. However, some limitations of this approach should be considered when paired or intertwined trees exist. The progress of research on this topic are reported in [24], [25]. For example, the tree isolator and 3DFin algorithms have been useful to estimate the inventory of some study regions given a specific dataset [26], [27].

Addressing the pattern identification problem by the 3D point cloud management aimed in the isolated tree species classification, still presents a challenge due to the delineted defects in the tree segmentations by the overlaping and linked branches as well as spectral variation over the species diversity among others uncertainties. Significative efforts have been done to the development of fusion strategies using spectral and structure information on CHMs [28]. In one hand, spectral data provided by advanced systems such as multi-spectral and hyper-espectral can capture spectral firms of canopies better than traditional RGB imagery can provide. However, these systems are limited to the wavelenght operation and spatial resolution arise upon the study environments,

since regarding to structural capture, specially for canopies detection, the data provided by these optical systems can powered only for the surface and spectral content on CHMs produced by traditional photogrammetry as well as the individual tree characterization using the 3D point cloud segmentation [29]. Furthermore, leveraging the application of fusion strategies, for the production of CHMs, adding at least two types of data given by photogrammetry and LiDAR, this last to provide better accuracy in the canopies capture, overall upon the tree structures can help significativelly the fidelity data content as well as the CHM production and individual tree isolators regarding to the spatial information. Also, point clouds generated by LiDAR system provide good accuracy over the georeferencing of point clouds, such that improve the data coregistration process between both nature of data obtained in simultenously way. Some considerable results using this approach are reported in [30], [31], [32].

A forest can be characterized by its attributes, such as parameters, characteristics, and variables [33]. Some basic tree attributes include height, the diameter breast height (DBH), crown diameter at different heights, basal area, and estimated surface and volume. Therefore, these attributes can be measured directly or calculated using statistical methods or physical models [34]. Regarding deep learning techniques, there are studies that analyze the suitability of models such as the Mask Region-Based Convolutional Network [35], as well as interesting novel proposals that introduce Transformer-based mechanisms for canopy tree semantic segmentation and comprehensive comparison of Transformer models against classical machine learning techniques [36], [37]. In the next section we describe an alternative methodology to primarily address the problem of tree canopy segmentation and characterization as a preprocessing step in DEMs to provide key data for the machine learning classifiers and improve their performance applied to tree canopy identification in a defined study area.

Tree canopy detection and characterization are central to sustainable development because they enable improved monitoring of deforestation, biomass, forest carbon, and urban green infrastructure. In this sense, this research is not only a technical contribution but also a response to social and environmental demands framed within the SDG 11 (Sustainable Cities and Communities), SDG 13 (Climate Action), and SDG 15 (Life on Land). Moreover, in the Mexican context, it aligns with PRONACES in Climate Change, Polluting Agents and Toxic processes, and Food Sovereighty, strengthening the link between advanced research and the needs of society.

II. Methodology

A. Airbone LiDAR dataset

The study area is the University of Hawaii at *Mānoa* project (UHM, see Fig. 2), whose dataset is referenced in [34]. This LiDAR point cloud contains buildings and vegetation available for our approach's segmentation process, as this application focuses on the characterization and classification of tree canopies. Besides, the study area was selected considering the diversity of objects present in the 3D scene of a real environment, as well as the high variability of the data captured by the LiDAR system (see Fig. 3). Considering the diagram of Fig. 1, which describes the workflow methodology of our approach, the processing starts with the generation of DEMs, such as the digital terrain model (DTM) and the digital height model (DHM), see Fig. 4 and Fig. 5. These models have been essential for subsequent procedures in handling 3D point clouds, especially when working with point cloud applications in urban environments, such as object detection and extraction. Likewise, filtering, classification, and interpolation of 3D point clouds are essential operations in the generation of DEMs, as well as in the detection and extraction of tree cover, as shown in the generation of the canopy height model (CHM) in this experiment, which is explained in the next subsection (see Fig. 6 and Fig. 7). Although the focus is on developing a robust technical pipeline, its design also contemplates scalability and applicability in decision-making contexts. Urban and rural authorities could adopt the proposed workflow to maintain updated inventories of tree cover, which are fundamental for ecosystem services management, health promotion, and mitigation of environmental risks.

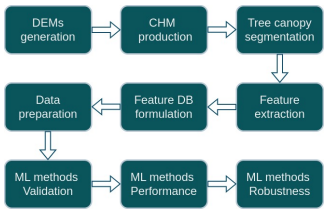

Fig. 1. Workflow of the methodology proposed.

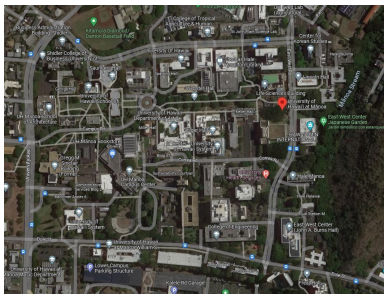

Fig. 2. Study area (UHM) captured by google maps.

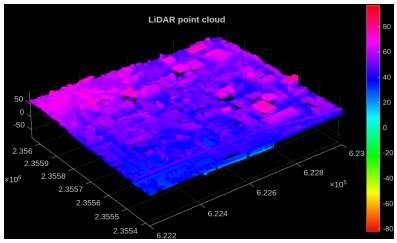

Fig. 3. LiDAR point cloud of the study area (UHM).

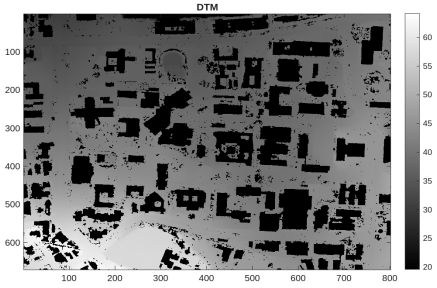

Fig. 4. Digital terrain model of the study area.

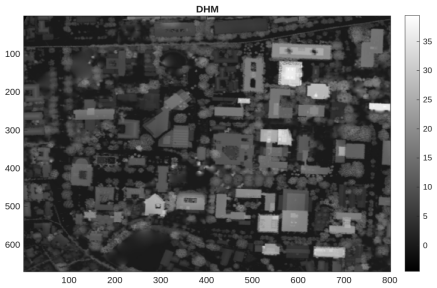

Fig. 5. Digital height model of the study area.

B. Forestry segmentation

In [38], a methodology for creating CHMs has been applied, leveraging the DEM generation base as a starting point. Primarily, the classification of the Laser LiDAR file format (LAS) point cloud, at least the ground and non-ground point cloud labels, is required to carry out the filtering process. In this way, the normalized digital surface model (nDSM) can be generated. This model should be the essential reference for applying filtering criteria based on the relationship of distances between points for the candidate 3D points belonging to the centers of the trees, as shown in the flowchart as an alternative proposal (see Fig. 8). Our methodological proposal focuses on identifying the predominant shapes of the different tree canopies contained in the nDSM, allowing the segmentation process to concentrate the different samples obtained, including data with uncertainty (fractions of contiguous objects, incomplete canopies, and overlaps), as well as their different shapes and sizes. Furthermore, the proposed segmentation algorithm can separate overlaps between tree canopies with intertwined branches (see Fig. 6 and Fig. 7).

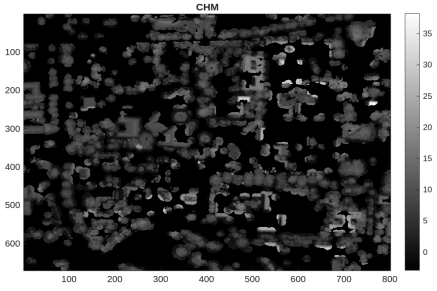

Fig. 6. Canopy height model of the study area.

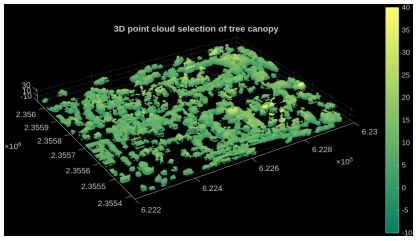

Fig. 7. Point cloud of CHM.

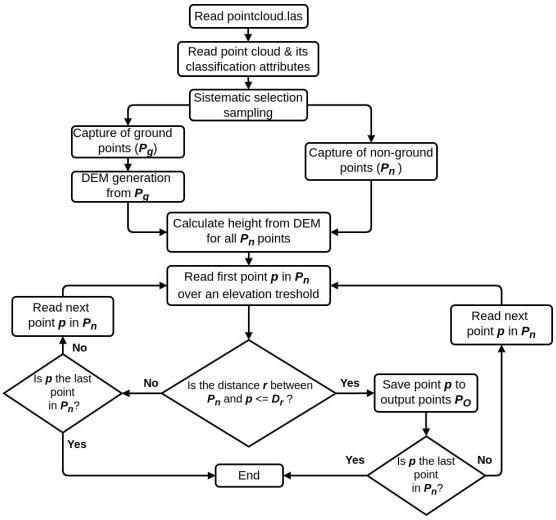

Fig. 8. Flowchart for creating new point clouds P_O.

C. Forestry Characterization

To achieve the canopy characterization, we consider tree attributes based on the shape of the extracted object point cloud, such as the ratio metrics of the produced solid and some basis statistical metrics of the individual point cloud distribution [39], [40], [41]. The tree canopy feature extraction vector is performed by parameters obtaining such as the number of 3D points and their cube-bounded measurements, and then including some robust geometry estimates based on the spatial distribution as shown in Fig. 9. These estimations allow to produce areas, volumes and surfaces of the solid generated by each segmented point cloud, as well as the compute of some geometry ratio metrics. Table I include all the metrics needed in this approach for the canopy characterization, we describe some of them in the following.

Absolute spatial refuge:

$$R = volHull - volSolid \tag{1}$$

Surface area to volume ratio:

$$surfVol = \frac{areaSurf}{volSolid} \tag{2}$$

Shelter size factor:

$$S = \frac{R}{areaSurf} \tag{3}$$

Proportion occupied:

$$prOcc = \frac{volSolid}{volHull} \tag{4}$$

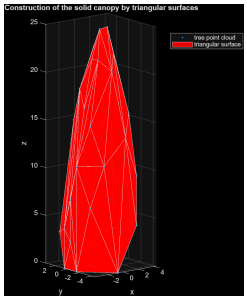

Fig. 9. Triangular surface applied in the canopy point cloud.

The features vector includes some basis statistical metric, such as arithmetic mean, median, standart deviation and variance. Some equations of them are defined following.

Arithmetic mean:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \tag{5}$$

Skewness:

$$skew = \frac{\sqrt{n}\sum_{i=1}^{n}(x_i-\mu)^3}{(\sum_{i=1}^{n}(x_i-\mu)^2)^{\frac{3}{2}}} \tag{6}$$

Kurtosis:

$$kts = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2)^2} - 3 \tag{7}$$

Energy function:

$$E = \sum_{i=1}^{n} p(i)^2 \tag{8}$$

Shanon entropy:

$$H = -\sum_{i=1}^{n} p(i)\log_2 p(i) \tag{9}$$

Table I contains 77 features of segmented point clouds that are organized as follows: the first are features related to the dimensions of the point cloud (22 metrics), the next features are based on the geometry of the point cloud and its dimension relationship (21 metrics), and the last features are obtained through measures of central tendency (44 metrics).

The prepared database contains 77 metrics for each segmented point cloud. In this experiment, matrix data must be considered for the classifier's learning and validation process. The following sections explains the required conditions to carry out the classification processing and the performance results obtained from three different robust learning models.

D. Classification processing

In this section is established the definition of the canopy morphology categories to be classified, relative to the shape incident on the study region, that is, after carrying out the segmentation process, the 3D point distributions of the tree crowns defined as RoI were considered and the following canopy morphologies were determined: irregular, peduncular, palmiform, parasol and longitudinal (see Fig. 10-A). The latter group contains all those canopy morphologies that present columnar, conical and fusiform shape tendencies, as

TABLE I
TABLE OF CANOPY FEATURES

Metric	Description
n	*Number of 3D points*
h	*Maximum height of the point cloud (z distance)*
h_{dz}	*Maximum long of the point cloud (z distance)*
z_{den}	*Point cloud density (over z)*
w	*Maximum wide of the point cloud (x distance)*
l	*Maximum long of the point cloud (y distance)*
$C_{xyz} \in \mathbb{R}^{3\times1}$	*Geometric center of the point cloud*
$D_{xyz} \in \mathbb{R}^{3\times1}$	*Inner product of the point cloud components*
$Ncros_{obj}$	*Vectorial product normal of the point cloud*
$SumD_{obj}$	*Inner product sum respect to geometric center*
$SumZ$	*Sum of vertical magnitudes of the point cloud*
ρ	*Point cloud density (cilindrical bounded)*
$perimPS$	*Perimeter of the basis 4-edged polygon cover*
$perimPC$	*Perimeter of the complet planar polygon*
$area$	*Maximum area of the cuboid's cover*
$volumen$	*Maximum volume of the cuboid's cover*
$areaPS$	*Coverage area of the basis 4-edged polygon*
$areaPC$	*Planar area of the 3D point cloud*
$areaSurf$	*3D surface area of the point cloud*
$volSolid$	*Total volume of the estimated solid*
$volHull$	*Convex hull volume of the estimated solid*
$normCt$	*Estimated centroid norm*
$normK$	*Mean convexity norm of the point cloud hull*
$normRad$	*Curvature radio norm of the surface mesh*
$facesSurf$	*Faces number of the point cloud surface mesh*
R	*Absolute spatial refuge*
$surfVol$	*Surface area to volume ratio*
S	*Absolute spatial refuge to 3D surface area ratio*
$prOcc$	*Ratio of colony volume to convex hull volume*
$meanDn \in \mathbb{R}^{3\times1}$	*Mean of normalized distances (μ)*
$medianDn \in \mathbb{R}^{3\times1}$	*Median of normalized distances*
$varDn \in \mathbb{R}^{3\times1}$	*Variance of normalized distances (σ^2)*
$stdDn \in \mathbb{R}^{3\times1}$	*Standart deviation of normalized distances (σ)*
$varCoef \in \mathbb{R}^{3\times1}$	*Coefficient of variation $(\frac{\sigma}{\mu})*100$ in 3D*
$perctsDn \in \mathbb{R}^{3\times6}$	*Percentiles: 5th, 10th, 25th, 50th, 75th, & 95th*
$skewDn \in \mathbb{R}^{3\times1}$	*Skewness*
$ktsDn \in \mathbb{R}^{3\times1}$	*Kurtosis*
$m3Dn \in \mathbb{R}^{3\times1}$	*3rd moment*
E	*Energy function applied in z*
H	*Shanon entropy applied in z*

well as the tree crowns that are formed by the dispersed branching along their trunk (see Fig. 10-B).

Due to the complexity of the study area, the segmentation of submerged tree crowns from the DEM is another canopy morphology defined in our approach, called small crowns. This category includes canopy morphologies with a very irregular shape, expressed in variable size, both vertically and horizontally. For this reason, we categorized this type of vegetation, since after the segmentation process, we may find multiple observations of them in one or more sections, as well as some overlap between crowns. Therefore, we contemplated providing this category to the classifier to expand the identification of unknown observations. Fig. 11 shows examples of small crowns captured from the study area. Note the variation in size and shape that generally appears across the study area; we can include other vegetation subgroups such as plants, shrubs, dense thickets, and others in this category.

On the other hand, based on the definition of the canopy

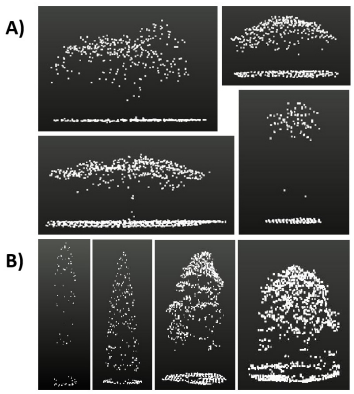

Fig. 10. Canopy morphologies included in the study area: A) Irregular, peduncular, palmiform & parasol (clockwise order, starting from the top left figure), B) Logitudinal tendency.

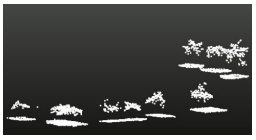

Fig. 11. Small canopy morphologies of the study area.

morphology in question, a category of 3D point clouds was also defined in which the objects that the segmentation algorithm considered as 3D points included in the treetops prevailed, since the selection algorithm based on the Euclidean distance between points can hardly discriminate these points due to their uniformity, which are located between the cylindrical shapes projected on the horizontal plane of three-dimensional space, as can be seen in the Fig. 12. The main 3D points that were added to the treetops generally coincided with the building roofs provided by the nDSM; i.e., in the segmentation process, some roofs or buildings were added as treetops.

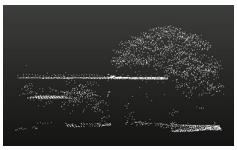

Fig. 12. Overlap between roofs and canopies.

To train the supervised classification process, 3D point clouds will first be obtained through manual and semi-automatic segmentation, determining the polygons projected onto the horizontal plane of the original point cloud to identify the desired canopy samples, as well as the observations obtained through the segmentation processing of our approach. These observations will then undergo a learning process using the selected classifier to subsequently identify unknown observations, including different tree canopies contained in the DEM (see Fig. 13).

The databases have been prepared for training and validation of different supervised classifiers, considering the following conditions:

- Observations: 130 canopy point clouds
- Features: 77 predictors
- Classes: 7 (6 canopy morphologies & 1 roofs category)
- Validation: 5-fold cross-validation
- Unknown observations: 50 canopy point clouds

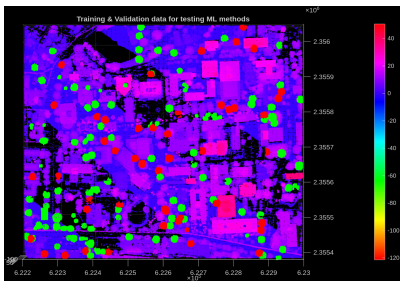

Fig. 13. Training & validation data for the classification process (130 trained observations - green, 50 unknown observations - red).

III. Validation Results

For the implementation and validation of the methodological proposal presented in this work, MatLab™ version R2025a software was used, specifically for the LiDAR, Statistics, and Machine Learning toolbox. To apply all the procedures of our methodological proposal described in the diagram of Fig. 1, as well as the segmentation strategy established in the decision diagram presented in Fig. 8 for the generation of CHM using the LAS file of the study area defined in Fig. 3, this point cloud is well delimited, and it is feasible to test and validate our approach perspective. It is important to note that the LiDAR dataset for the study area uses a single capture stage and does not include captures from different stations throughout the year. In case of testing these algorithms using other 3D LiDAR point clouds and even point clouds provided by photogrammetry techniques, it will be necessary to properly define the spatial constraints, especially for the generation of DEMs and the segmentation and extraction strategy of the 3D objects included in the DEMs that represent the study area.

In order to search the best performance of machine learning (ML) models using Matlab's classification learner for training and validation data, as shown in Fig. 13, this experiment tested the training of three ML models: the Naive Bayes (NB), Neural Network (NN) and Support Vector Machine (SVM) obtaining validation accuracies of 77.7%, 78.5% and 80.0%, respectively. Likewise, the matrix confusion validation for these ML models is presented in the left side of figures 14, 15 and 16. The positive predicted values (PPV) and false discovery rates (FDR) are depicted in these images. Furthermore, the relative operating characteristic (ROC) curve validation for each training model applied as well as some training results are shown in the right side of figures 14, 15 and 16. Note that the weighted average of the area under the curve (AUC) for each of these training models resulted in 0.9095, 0.9304 and 0.9277, respectively.

Following are the main defined hyperparameters for each optimizable ML model that performed best during the train-

ing and validation processing. Only some of the NN model's hyperparameters were manually adjusted to improve its performance, by adding a layer and increasing the number of neurons.

Naive Bayes

- Preset: Optimizable Naive Bayes
- Support: Unbounded
- Optimized Hyperparameters
  - + Distribution names: Gaussian
  - + Kernel type: Epanechnikov
  - + Standardize data: No

Neural Network

- Preset: Optimizable Neural Network
- Number of fully connected layers: 2
- Iteration limit: 1000
- First layer size: 1000
- Second layer size: 100
- Optimized Hyperparameters
  - + Activation: ReLU (Rectified Linear Unit)
  - + Regularization strength (Lambda): 2.626e-07
  - + Standardize data: Yes

Support Vector Machine

- Preset: Optimizable SVM
- Kernel scale: 1
- Optimized Hyperparameters
  - + Kernel function: Quadratic
  - + Box constraint level: 0.0010002
  - + Multiclass coding: One-vs-All
  - + Standardize data: Yes

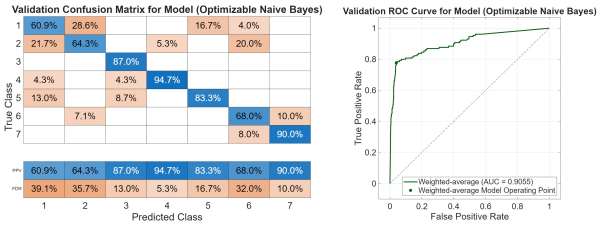

Fig. 14. Confusion matrix validation for the Naive Bayes model.

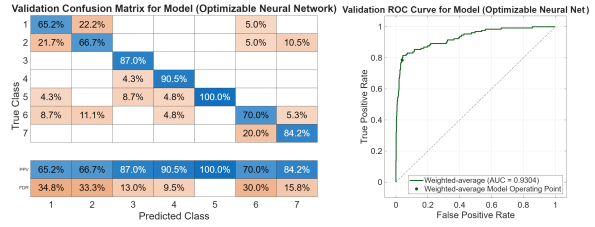

Fig. 15. Confusion matrix validation for the Neural Network model.

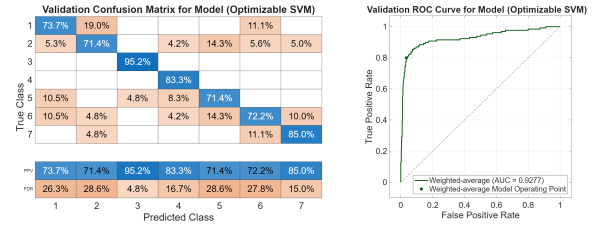

Fig. 16. Confusion matrix validation for the cubic SVM model.

Furthermore, to complement the analysis of each machine learning model's performance, Fig. 17 compares the models' response to the provided feature set, considering the use of the Kruskal Wallis method for strategic feature selection. An increasing trend in the accuracy of the learning process is observed as the number of features in each validation increases.

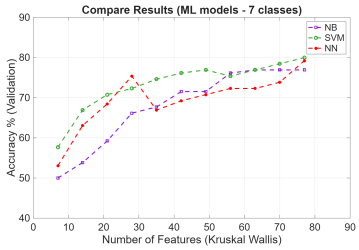

Fig. 17. Accuracy of validation models vs feature set.

According to the validation of unknown data the experiment has obtained considerable results, since the unknown observations are chosen disturbed and very variational with respect to shape and size, these samples are some fractions or sections of the different canopies after segmentation processing, obtaining a good precision for canopies with irregular and peduncular tendency as well as a better identification for small canopies and roofs due to the homogeneity of their shape tendency. As shown in the previous results, classification processing performance could be improved by adding more observations and predictors. The experiment has limitations in terms of the DEM's extent and, consequently, the number of observations and the fidelity of the training data. However, as a starting point for this learning process, its impact is promising with large spatial and homogeneity data. This work offers a practical, replicable, and data-driven approach for LiDAR data processing in environmental applications, establishing a robust baseline accuracy that can be further improved, also using other open-source tools and programming languages, such as Python and R. While these results highlight the technical feasibility of the method, they also reinforce its potential as a tool for social transformation. A reliable classification of urban tree canopies provides municipalities with data for heat island mitigation, pollution reduction, and equitable urban planning, directly impacting public health (SDG 3 – Health and Well-Being).

IV. CONCLUSIONS

In this work, we propose an alternative solution to address the problem of pattern identification using LiDAR point clouds, focusing on the tree canopies classification. Our main contribution lies in proposing a framework based on a preprocessing stage that includes object segmentation using a decision-making approach and the substantial extraction of features inherent in the point cloud morphologies. To further strengthen the feature extraction process, we will complement our proposal by considering the spectral information of the point clouds, such as RGB and reflectivity data, adding their statistical metrics and spectral indices, thus achieving

better results in tree canopy classification using traditional machine learning models. Furthermore, our next research challenge is to significantly expand the study area to obtain a sufficient amount of tree canopy data to feed the classifiers using our proposed preprocessing method, as well as to perform a systematic analysis of the performance of each classifier based on the provided canopy samples, in order to characterize the robustness of the machine learning models employed. Subsequently, we will focus on the temporal analysis of LiDAR captures from the study area to further strengthen the classification process. Beyond technical validation, the true contribution of this work lies in bridging scientific innovation and societal needs. By aligning with the SDGs and PRONACES, this research positions canopy segmentation and classification as tools that can foster urban sustainability, environmental resilience, and equitable social development.

REFERENCES

[1] P. K. Garg, Remote sensing: theory and applications, Ed. Mercury Learning and Information, 2024.

[2] H. Yao et al., Unmanned Aerial Vehicle for Remote Sensing Applications – A Review, Remote Sensing, vol. 11, no. 12. 2019.

[3] Z. Yang et al., UAV remote sensing applications in marine monitoring: Knowledge visualization and review, Science of The Total Environment, vol. 838, pp. 155939, Sept. 2022.

[4] J. Pišl et al., Mapping drivers of tropical forest loss with satellite image time series and machine learning, Environmental Research Letters, vol. 19, no. 6, pp. 064053, May 2024.

[5] N. van Tiel et al., Regional uniqueness of tree species composition and response to forest loss and climate change, Nature Communications, vol. 15, no. 1, pp. 4375, May 2024.

[6] E. Beech, M. Rivers, S. Oldfield, and P. P. Smith, GlobalTreeSearch: The first complete global database of tree species and country distributions, Journal of Sustainable Forestry, March 2017.

[7] S. Wang, J. Ji, L. Zhao, J. Li, M. Zhang, and S. Li, Canopy Segmentation of Overlapping Fruit Trees Based on Unmanned Aerial Vehicle LiDAR, Agriculture, vol. 15, no. 3, 2025.

[8] C. Vega et al., PTrees: A point-based approach to forest tree extraction from lidar data, International Journal of Applied Earth Observation and Geoinformation, vol. 33, pp. 98-108, Dec. 2014.

[9] B. Brede, A. Lau, H. M. Bartholomeus, and L. Kooistra, Comparing RIEGL RiCOPTER UAV LiDAR Derived Canopy Height and DBH with Terrestrial LiDAR, Sensors, vol. 17, no. 10, 2017.

[10] B. C. Budei, B. St-Onge, C. Hopkinson, and F.-A. Audet, Identifying the genus or species of individual trees using a three-wavelength airborne lidar system, Remote Sensing of Environment, vol. 204, Oct. 2017.

[11] Q. Chen, P. Gong, D. Baldocchi, and Y. Tian, Estimating Basal Area and Stem Volume for Individual Trees from Lidar Data, Photogrammetric Engineering and Remote Sensing, vol. 73, pp. 1355-1365, Dec. 2007.

[12] W. Li, Q. Guo, M. Jakubowski, and M. Kelly, A New Method for Segmenting Individual Trees from the Lidar Point Cloud, Photogrammetric Engineering and Remote Sensing, vol. 78, pp. 75-84, Jan. 2012.

[13] C. Qian, C. Yao, H. Ma, J. Xu, and J. Wang, Tree Species Classification Using Airborne LiDAR Data Based on Individual Tree Segmentation and Shape Fitting, Remote Sensing, vol. 15, no. 2, 2023.

[14] F. I. Matias, M. V. Caraza-Harter, and J. B. Endelman, FIELDimageR: An R package to analyze orthomosaic images from agricultural field trials, vol. 3, no. 1, pp. e20005, 2020.

[15] Q. Hu et al., RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds, in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11105-11114, 2020.

[16] Y. Huang et al., Tree Species Classification from UAV Canopy Images with Deep Learning Models, Remote Sensing, vol. 16, no. 20, 2024.

[17] N. Varney, V. K. Asari, and Q. Graehling, DALES: A Large-scale Aerial LiDAR Data Set for Semantic Segmentation, in Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 717-726, 2020

[18] J. Mo et al., Deep Learning-Based Instance Segmentation Method of Litchi Canopy from UAV-Acquired Images, Remote Sensing, vol. 13, no. 19, 2021.

[19] Z. Li et al., Fruit tree canopy segmentation from UAV orthophoto maps based on a lightweight improved U-Net, Computers and Electronics in Agriculture, vol. 217, pp. 108538, Feb. 2024.

[20] P. Marques et al., UAV-Based Automatic Detection and Monitoring of Chestnut Trees, Remote Sensing, vol. 11, no. 7, 2019.

[21] T.-A. Nguyen, B. Kellenberger, and D. Tuia, Mapping forest in the Swiss Alps treeline ecotone with explainable deep learning, Remote Sensing of Environment, vol. 281, pp. 113217, Nov. 2022.

[22] T.-A. Nguyen, M. Ruwurm, G. Lenczner, and D. Tuia, Multi-temporal forest monitoring in the Swiss Alps with knowledge-guided deep learning, Remote Sensing of Environment, vol. 305, pp. 114109, May 2024.

[23] G. Liu, J. Wang, P. Dong, Y. Chen, and Z. Liu, Estimating Individual Tree Height and Diameter at Breast Height (DBH) from Terrestrial Laser Scanning (TLS) Data at Plot Level, Forests, vol. 9, no. 7, 2018.

[24] A. Burt, M. Disney, and K. Calders, Extracting individual trees from lidar point clouds using treeseg, vol. 10, no. 3, pp. 438-445, 2019.

[25] C. Cabo, C. Ordóñez, C. A. López-Sánchez, and J. Armesto, Automatic dendrometry: Tree detection, tree height and diameter estimation using terrestrial laser scanning, International Journal of Applied Earth Observation and Geoinformation, vol. 69, pp. 164-174, July 2018.

[26] D. Laino et al., 3DFin: a software for automated 3D forest inventories from terrestrial point clouds, Forestry: An International Journal of Forest Research, vol. 97, no. 4, pp. 479-496, 2024.

[27] Z. Xi and C. Hopkinson, 3D Graph-Based Individual-Tree Isolation (Treeiso) from Terrestrial Laser Scanning Point Clouds, Remote Sensing, vol. 14, no. 23, 2022.

[28] L. Wallace, A. Lucieer, Z. Malenovský, D. Turner, and P. Vopěnka, Assessment of Forest Structure Using Two UAV Techniques: A Comparison of Airborne Laser Scanning and Structure from Motion (SfM) Point Clouds, Forests, vol. 7, no. 3, 2016.

[29] Z. Wang et al., Crown-level tree species classification using integrated airbone hyperspectral and lidar remote sensing data, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., vol. XLII-3, pp. 2629-2634, 2018.

[30] M. Dalponte and D. A. Coomes, Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data, Methods in Ecology and Evolution, vol. 7, no. 10, pp. 1236-1245, Oct. 2016.

[31] G. Goldbergs, S. W. Maier, S. R. Levick, and A. Edwards, Efficiency of Individual Tree Detection Approaches Based on Light-Weight and Low-Cost UAS Imagery in Australian Savannas, Remote Sensing, vol. 10, no. 2, 2018.

[32] M. Herrero-Huerta et al., Dense Canopy Height Model from a low-cost photogrammetric platform and LiDAR data, Trees, vol. 30, no. 4, pp. 1287-1301, 2016.

[33] C. A. Silva et al., Imputation of Individual Longleaf Pine (Pinus palustris Mill.) Tree Attributes from Field and LiDAR Data, Canadian Journal of Remote Sensing, vol. 42, no. 5, pp. 554-573, Sept. 2016.

[34] P. Dong, and Q. Chen, LiDAR Remote Sensing and Applications. CRC Press, 2017.

[35] Y. Lin et al., Individual Tree Crown Delineation Using Airborne LiDAR Data and Aerial Imagery in the Taiga-Tundra Ecotone, Remote Sensing, vol. 16, no. 21, 2024

[36] B. Xiang et al., ForestFormer3D: A Unified Framework for End-to-End Segmentation of Forest LiDAR 3D Point Clouds, in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2025.

[37] J. Taher et al., Multispectral airborne laser scanning for tree species classification: a benchmark of machine learning and deep learning algorithms, 2025.

[38] Liu and P. Dong, A new method for generating canopy height models from discrete-return LiDAR point clouds, Remote Sensing Letters, vol. 5, June 2014.

[39] E. A. Aston et al., A Protocol for Extracting Structural Metrics From 3D Reconstructions of Corals, (in English), Methods vol. 9, Apr. 2022.

[40] P. Dong, Characterization of individual tree crowns using three-dimensional shape signatures derived from LiDAR data, International Journal of Remote Sensing, vol. 30, pp. 6621-6628, Nov. 2009.

[41] P. Dong, Sensitivity of LiDAR-derived three-dimensional shape signatures for individual tree crowns: A simulation study, Remote Sensing Letters, vol. 1, pp. 159-167, Sept. 2010.