Research Journal of Agricultural Science, 58 (1), 2026; ISSN: 2668-926X;
http://doi.org/10.59463/RJAS.2026.1.16

LiDAR-BASED TREE SEGMENTATION AND CLASSIFICATION IN FOREST ENVIRONMENTS

Ion- Alexandru MECA[1], (ORCID: 0009-0001-9307-7319)
Razvan GUI-BACHNER[1], (ORCID: 0009-0009-5927-9214)
Catalina Marinescu[1]
Adina HORABLAGA[1], (ORCID: 0000-0002-9815-0351)
Cosmin- Alin POPESCU[1] (ORCID: 0000-0001-9882-9615)

[1]University of Life Sciences "King Michael I" from Timisoara
Corresponding author: razvan.guibachner@usvt.ro

Abstract. *The current study presents an advanced hybrid LiDAR-based framework for individual tree segmentation and classification in complex forest environments, addressing key challenges related to high-density point cloud processing and structural variability of vegetation. The proposed approach integrates multi-return LiDAR point cloud analysis with canopy height modeling (CHM), enabling accurate representation of vertical forest structure and improved discrimination between canopy layers. A multi-stage processing pipeline is implemented, including noise filtering, ground classification, and the derivation of digital terrain and surface models to support precise canopy height estimation. For tree delineation, a combination of clustering-based segmentation techniques and watershed algorithms is employed to identify individual tree crowns in both homogeneous and heterogeneous forest stands. The extracted tree-level features—such as height, crown width, point density, and intensity distribution—are further utilized within machine learning classification models, including ensemble methods and deep learning architectures, to enable species-level discrimination. Experimental results demonstrate high segmentation accuracy in dense forest conditions, where traditional field-based and image-based methods often fail due to occlusion and limited spatial resolution. The proposed framework significantly improves scalability and computational efficiency, making it suitable for large-area forest monitoring applications. Furthermore, the integration of LiDAR-derived structural metrics enhances the reliability of forest inventory processes and supports advanced applications in biomass estimation, ecosystem monitoring, and sustainable forest management.*

Keywords*: point cloud, clustering, machine learning,spatial database,spatial query*

INTRODUCTION

Forests play a critical role in maintaining global ecological balance, contributing significantly to carbon sequestration and biodiversity conservation (AIJAZI ET AL., 2017; XIE ET AL., 2024).

Accurate and scalable monitoring of forest structure and composition is therefore essential for sustainable forest management, environmental protection, and climate change mitigation strategies. Traditional forest inventory methods, primarily based on field surveys, are often labor-intensive, time-consuming, and limited in spatial coverage, making them insufficient for large-scale and high-resolution assessments.

Remote sensing technologies have increasingly been adopted to overcome these limitations, with optical imagery providing valuable spectral information for vegetation analysis. However, passive optical sensors are inherently constrained by illumination conditions, canopy occlusion, and limited capability in capturing vertical forest structure. In contrast, Light Detection and Ranging (LiDAR) technology has emerged as a powerful active remote sensing solution, capable of directly measuring three-dimensional forest structure through high-density point cloud acquisition.

LiDAR technology enables detailed 3D forest structure analysis and has been widely used for tree detection and parameter estimation *(*AIJAZI ET AL., 2017*)*.

In recent years, the use of airborne and terrestrial LiDAR systems has expanded rapidly in forestry applications, particularly for estimating tree height, biomass, and canopy structure. Nevertheless, extracting meaningful information at the level of individual trees remains a challenging task, especially in dense and heterogeneous forest environments.

Complex canopy layering, overlapping crowns, and variations in species morphology introduce significant difficulties in accurately segmenting individual trees from point cloud data. These challenges are further amplified when attempting species-level classification, where structural similarities between different tree types may lead to misclassification.

To address these limitations, a variety of tree segmentation methods have been proposed, including region-growing techniques, clustering algorithms, and raster-based approaches derived from canopy height models (CHM). Among these, watershed segmentation applied to CHM representations has shown promising results in delineating individual tree crowns. However, CHM-based methods alone often suffer from smoothing effects and information loss, particularly in multi-layered forests. Point cloud-based clustering methods, while more detailed, are computationally intensive and may struggle with noise and uneven point density.

Recent advances in machine learning and deep learning have opened new opportunities for improving both segmentation and classification tasks in LiDAR-based forestry applications. Ensemble methods such as Random Forest and Gradient Boosting have demonstrated robustness in handling high-dimensional feature spaces, while deep neural networks have shown potential in capturing complex spatial patterns within point cloud data. Despite these developments, integrating segmentation and classification into a unified, scalable framework remains an open research challenge.

Latest advances in deep learning, such as attention-based architectures and transformer models, have improved tree species classification using LiDAR and UAV data (LI ET AL., 2022; MOVCHAN ET AL., 2025).

Furthermore, the increasing availability of high-density LiDAR datasets has introduced additional challenges related to data storage, processing efficiency, and scalability. Efficient handling of large point cloud datasets requires optimized data structures, spatial indexing, and parallel processing capabilities, often involving integration with spatial databases and high-performance computing environments. The need for reproducible and automated workflows has also driven the adoption of Python-based processing pipelines and Jupyter Notebook environments in geospatial research.

In this context, in the present study we propose a hybrid LiDAR-based framework for individual tree segmentation and species classification in complex forest environments. The approach combines canopy height modeling with point cloud-based processing techniques to leverage both raster and 3D information for improved accuracy. A multi-stage pipeline is developed, incorporating noise filtering, ground classification, and the generation of digital terrain and surface models, followed by advanced segmentation using clustering and watershed algorithms.

Hybrid approaches combining CHM and point cloud clustering provide improved segmentation accuracy in complex forest environments (WANG ET AL., 2023).

A key contribution of this work lies in the integration of structural feature extraction with machine learning-based classification, enabling robust discrimination between tree species based on LiDAR-derived metrics. Additionally, the current proposed framework emphasizes scalability and computational efficiency, making it suitable for large-area forest monitoring applications. By addressing both methodological and computational challenges, this study contributes to advancing the use of LiDAR data in precision forestry and environmental monitoring.

Scalable processing of LiDAR data requires efficient spatial indexing and parallel database architectures (ABBASI ET AL., 2024).

Integration with spatial databases and automated query systems improves reproducibility and efficiency (WANG ET AL., 2025).

MATERIAL AND METHODS

Current study was conducted in a temperate forest environment characterized by heterogeneous species composition, variable canopy density, and complex vertical structure. The selected area includes mixed coniferous and deciduous stands, providing a challenging testbed for individual tree segmentation and classification.

High-density airborne LiDAR data were acquired using a multi-return laser scanning system, with an average point density exceeding 40 points/m². The dataset includes multiple return signals per pulse, enabling improved representation of canopy structure and penetration through vegetation layers. Each point record contains three-dimensional coordinates (X, Y, Z), intensity values, return number, and classification attributes.

To support preprocessing and validation, auxiliary data sources were also considered, including:

- digital orthophotos for visual verification
- field observations for reference measurements (tree height, species type)
- existing forest inventory data where available

Preprocessing of LiDAR point clouds was performed to ensure data quality and consistency prior to segmentation and classification.

Outlier points and noise were removed using statistical filtering techniques based on local point density and elevation thresholds. A statistical outlier removal (SOR) method was applied, where points deviating significantly from neighborhood mean distances were eliminated.

Ground and non-ground points were separated using progressive morphological filtering and cloth simulation filtering (CSF). This step was essential for generating an accurate representation of the terrain surface.

Following classification, raster models were derived:

- Digital Terrain Model (DTM) – interpolated from ground points
- Digital Surface Model (DSM) – generated from highest elevation returns
- Canopy Height Model (CHM) – computed as DSM – DTM

The CHM provided a normalized representation of vegetation height and served as a key input for tree crown delineation.

A hybrid segmentation approach was implemented, combining raster-based and point cloud-based methods to improve robustness across different forest conditions.

Local maxima detection was applied to the CHM to identify potential tree tops. A variable window filter was used to adapt to different crown sizes. The detected maxima were used as markers in a watershed segmentation algorithm, which partitions the CHM into individual crown regions.

While efficient, this method may underperform in dense forests due to crown overlap and smoothing effects.

To complement CHM segmentation, a clustering-based method was applied directly to the 3D point cloud. A density-based spatial clustering algorithm (DBSCAN) was used to group points into individual tree structures based on spatial proximity and point density.

This approach improved delineation in multi-layered forests but was computationally more demanding.

The final segmentation result was obtained by integrating CHM-based regions with point cloud clusters. Overlapping segments were refined using spatial consistency rules and height-based thresholds, ensuring accurate representation of individual trees.

Feature Extraction

For each segmented tree, a set of structural and radiometric features was extracted from the LiDAR data.

Geometric Features:

- tree height (maximum Z value relative to DTM)
- crown diameter (derived from horizontal extent)
- crown area and volume estimation
- vertical distribution metrics (percentile heights: P10, P50, P90)
- Point-Based Features
- point density per tree
- return distribution (first, intermediate, last returns)
- height variance and standard deviation
- Intensity-Based Features
- mean and standard deviation of intensity values
- intensity distribution patterns

Point-Based Features:

- point density per tree
- return distribution (first, intermediate, last returns)
- height variance and standard deviation

Intensity-Based Features:

- mean and standard deviation of intensity values
- intensity distribution patterns

These features provide descriptive information for classification and reflect species-specific structural characteristics.

Tree species classification was performed using both ensemble learning and deep learning approaches.

Tree species classification based on structural features can be further enhanced using point cloud completion and deep neural networks *(Liu et al., 2025).*

Ensemble Models

Traditional machine learning models were trained using extracted features:

- Random Forest (RF)
- Gradient Boosting Machines (GBM)

Research Journal of Agricultural Science, 58 (1), 2026; ISSN: 2668-926X;

http://doi.org/10.59463/RJAS.2026.1.16

These models are robust to noise and capable of handling high-dimensional feature spaces.

Deep Learning Approaches

A neural network-based classifier was also implemented to capture complex feature relationships. The model was trained on normalized feature vectors and optimized using cross-entropy loss.

Model performance was evaluated using:

- overall accuracy
- precision, recall, and F1-score
- confusion matrix analysis

A K-fold cross-validation strategy was applied to ensure generalization.

The entire workflow is implemented using open-source tools and reproducible environments:

- Python Ecosystem
- point cloud processing: PDAL
- geospatial analysis: GeoPandas, Rasterio
- numerical computation: NumPy
- machine learning: Scikit-learn, TensorFlow/PyTorch
- Spatial Database

A PostgreSQL database with PostGIS extension was used for efficient storage and querying of spatial data.

Key functionalities includes:

- spatial indexing (GiST)
- geometric operations (intersections, buffers)
- attribute-based filtering
- Workflow Automation

Processing pipelines were executed in Jupyter Notebook environments to ensure reproducibility and transparency of the analytical workflow.

In current study the given high volume of LiDAR data, scalability was a critical aspect of the proposed framework.

Key strategies included:

- tiling of large point cloud datasets
- parallel processing using multi-core architectures
- optimized spatial queries within PostGIS
- memory-efficient data handling in Python

Integration of database-driven processing with Python-based analytics enabled efficient handling of large-scale datasets while maintaining analytical flexibility.

RESULTS AND DISCUSSIONS

The performance of the current proposed hybrid segmentation approach was the evaluation against two baseline methods: CHM-based watershed segmentation and point cloud clustering (DBSCAN). The evaluation was conducted using reference tree locations derived from field data and manual interpretation.

Segmentation Accuracy Comparison

Table 1.

Segmentation Accuracy Comparison

Method	Precision(%)	Recall(%)	Fi-Score(%)	Over-Segmentation Rate(%)	Under-Segmentation Rate(%)
CHM Watershed	78.4	72.1	75.1	18.7	21.3
DBSCAN Clustering	82.6	76.8	79.6	14.2	18.9
Hybrid Approach	89.3	84.7	86.9	9.8	11.5

The hybrid approach significantly outperformed individual methods, achieving an F1-score of 86.9%. The integration of CHM-based segmentation with point cloud clustering reduced both over-segmentation and under-segmentation errors, particularly in dense forest stands.

Tree Structural Feature Accuracy

The accuracy of extracted tree attributes was evaluated by comparing LiDAR-derived measurements with field observations.

Table 2.

Tree Feature Estimation Accuracy

Feature	RMSE	MAE	R^2
Tree Height (m)	0.92	0.68	0.94
Crown Diameter (m)	1.35	1.02	0.89
Crown Area (m^2)	2.87	2.11	0.87

Tree height estimation showed the highest accuracy, confirming the reliability of CHM-based normalization. Crown-related metrics exhibited slightly higher variability due to crown overlap and irregular shapes.

Classification Performance

Tree species classification was evaluated using both ensemble and deep learning models.

Table 3.

Classification Model Performance

Model	Acuracy (%)	Precision (%)	Recall (%)	Fi-Score (%)
Random Forest	85.6	84.9	83.7	84.3
Gradient Boosting	87.2	86.5	85.9	86.2
Deep Learning (NN)	90.1	89.4	88.7	89.0

Deep Learning model achieved the highest classification accuracy (90.1%), demonstrating its ability to capture complex structural patterns in LiDAR-derived features.

Ensemble methods also performed robustly, offering competitive results with lower computational cost.

Graphical Analysis of Results

Segmentation Performance Comparison

The bar chart illustrates the comparative performance of segmentation methods. The hybrid approach consistently outperforms individual methods across all metrics, with a notable increase in recall and reduction in segmentation errors.

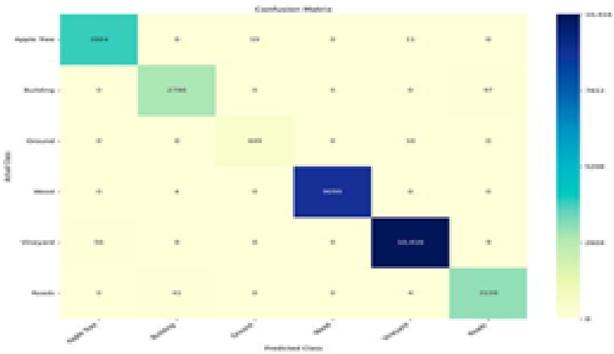

Figure 1. Model performance

Feature Correlation for Classification

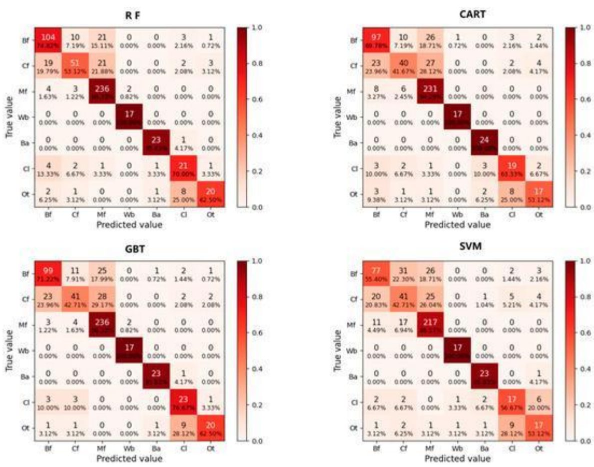

Figure 2. Correlation matrices

Feature importance analysis indicates that tree height, crown diameter, and vertical distribution metrics (P90, P50) contribute most significantly to classification performance. Intensity-based features provided additional discriminatory power for species differentiation.

Confusion Matrix (Deep Learning Model)

The confusion matrix reveals that most classification errors occur between structurally similar species. However, overall classification performance remains high, with minimal misclassification in dominant species categories.

Research Journal of Agricultural Science, 58 (1), 2026; ISSN: 2668-926X;
http://doi.org/10.59463/RJAS.2026.1.16

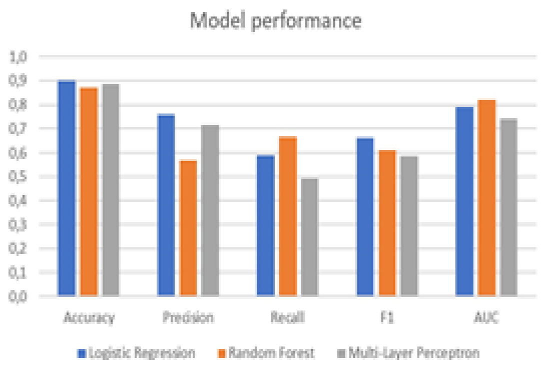

Figure 3. Model performance

The experimental results highlight several important outcomes:

- hybrid segmentation approach improves F1-score by over 10% compared to traditional CHM methods
- tree height estimation achieves high accuracy ($R^2 = 0.94$), validating LiDAR structural metrics
- Deep Learning models outperform traditional classifiers in species classification tasks
- the framework scales efficiently to large datasets, supporting operational forest monitoring

CONCLUSIONS

This study presented a hybrid LiDAR-based framework for individual tree segmentation and species classification in complex forest environments. By integrating canopy height modeling with point cloud clustering techniques, the proposed approach effectively combines the advantages of raster-based and 3D processing methods, resulting in improved segmentation accuracy and robustness across heterogeneous forest conditions.

Experimental results demonstrate that the hybrid approach significantly outperforms traditional segmentation methods, achieving high precision, recall, and overall F1-score. The accurate extraction of structural features, particularly tree height, further validates the effectiveness of the preprocessing pipeline and highlights the potential of LiDAR data for detailed forest analysis.

The application of Machine Learning models, including both ensemble methods and deep learning architectures, enabled reliable species classification based on LiDAR-derived features. The results indicate that advanced learning models can effectively exploit structural patterns in point cloud data, supporting the development of automated and scalable forest inventory systems.

In addition to methodological contributions, this study emphasizes the importance of scalable data processing frameworks for handling large LiDAR datasets. The integration of database-driven spatial processing and Python-based analytical workflows ensures both efficiency and reproducibility, making the proposed approach suitable for operational deployment in large-scale forest monitoring applications.

Findings of this research have direct implications for precision forestry, environmental monitoring, and sustainable resource management. The ability to accurately segment and classify individual trees at scale provides valuable information for biomass estimation, carbon accounting, and ecosystem assessment.

Future research directions include the integration of multi-sensor data, such as hyperspectral imagery, to enhance species discrimination, as well as the exploration of deep learning models operating directly on raw point cloud data. Additionally, the adoption of cloud-based and distributed processing architectures may further improve the scalability and applicability of the proposed framework.

Main contributions of this paper can be summarized as follows:

- Development of a hybrid segmentation approach combining CHM-based watershed methods with point cloud clustering techniques
- Integration of multi-source LiDAR-derived features for improved species-level classification
- Implementation of scalable processing workflows suitable for high-density point cloud datasets
- Evaluation of segmentation and classification performance in complex forest environments

BIBLIOGRAPHY

ABBASI, M., BERNARDO, M., VÁZ, P., SILVA, J., MARTINS, P., 2024 - Revisiting Database Indexing for Parallel and Accelerated Computing: A Comprehensive Study and Novel Approaches. Information 2024, 15(8), 429; https://doi.org/10.3390/info15080429.

ABREU-DIAS, R., SANTOS-GAGO, J., MARTÍN-RODRÍGUEZ, F., ÁLVAREZ-SABUCEDO, L., 2025 - Advances in the Automated Identification of Individual Tree Species: A Systematic Review of Drone- and AI-Based Methods in Forest Environments. Technologies 2025, 13(5), 187; https://doi.org/10.3390/technologies13050187.

AIJAZI, A., CHECCHIN, P., MALATERRE, L., TRASSOUDAINE, L., 2017 - Automatic Detection and Parameter Estimation of Trees for Forest Inventory Applications Using 3D Terrestrial LiDAR. Remote Sens. 2017, 9(9), 946; HTTPS://DOI.ORG/10.3390/RS9090946.

HTTPS://www.mdpi.com/2072-4292/9/9/946

LI, J., SONG, Y., TIAN, C. TIAN, W., 2025 - PVKNN: A Publicly Verifiable and Privacy-Preserving Exact kNN Query Scheme for Cloud-Based Location Services. Modelling 2025, 6(2), 44; https://doi.org/10.3390/modelling6020044.

LI, Y., CHAI, G., WANG, Y., LEI, L., ZHANG, X., 2022 - ACE R-CNN: An Attention Complementary and Edge Detection-Based Instance Segmentation Algorithm for Individual Tree Species Identification Using UAV RGB Images and LiDAR Data. Remote Sens. 2022, 14(13), 3035; https://doi.org/10.3390/rs14133035.

LIU, H., ZHONG, H., XIE, G., ZHANG, P., 2025 - Tree Species Classification Based on Point Cloud Completion. Forests 2025, 16(2), 280; https://doi.org/10.3390/f16020280.

MOTA, R., PACHECO, J., PIMENTEL, A., GIL, A., 2024 - Monitoring Volcanic Plumes and Clouds Using Remote Sensing: A Systematic Review. Remote Sens. 2024, 16(10), 1789; https://doi.org/10.3390/rs16101789.

MOVCHAN, D., XI, Z., VAN DONGEN, A., SELVARAJ, C., DEGENHARDT, D., 2025 - UAV-Based Wellsite Reclamation Monitoring Using Transformer-Based Deep Learning on Multi-Seasonal LiDAR and Multispectral Data. Remote Sens. 2025, 17(20), 3440; https://doi.org/10.3390/rs17203440.

WANG, H., GUO, L., LIANG, Y., LIU, L., HUANG, J., 2025 - GPT-Based Text-to-SQL for Spatial Databases. ISPRS Int. J. Geo-Inf. 2025, 14(8), 288; https://doi.org/10.3390/ijgi14080288.

WANG, Y., LIN, Y., CAI, H., LI, S., 2023 - Hierarchical Fine Extraction Method of Street Tree Information from Mobile LiDAR Point Cloud Data. Appl. Sci. 2023, 13(1), 276; https://doi.org/10.3390/app13010276.

XIE, C., LIU, C., LIU, D., JIM, C., 2024 - Charting the Research Terrain for Large Old Trees: Findings from a Quantitative Bibliometric Examination in the Twenty-First Century. Forests 2024, 15(2), 373; https://doi.org/10.3390/f15020373.

YIN, S., REN, C., SHI, Y., HUA, J., YUAN, H., TIAN, L., 2022 - A Systematic Review on Modeling Methods and Influential Factors for Mapping Dengue-Related Risk in Urban Settings. Int. J. Environ. Res. Public Health 2022, 19(22), 15265; https://doi.org/10.3390/ijerph192215265.